TEACHER EVALUATION OF READING UNDERACHIEVEMENT OF INTERMEDIATE GRADE

STUDENTS AS COMPARED TO EVALUATION OF READING UNDERACHIEVEMENT

BY DISCRIMINATE TESTING PROCEDURES

By

TWILLA AVARD JORDAN

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1958

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1961

Submitted to the Faculty of the Graduate School of
the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF EDUCATION
August, 1965

TEACHER EVALUATION OF READING UNDERACHIEVEMENT OF INTERMEDIATE GRADE

STUDENTS AS COMPARED TO EVALUATION OF READING UNDERACHIEVEMENT

BY DISCRIMINATE TESTING PROCEDURES

Thesis Approved:

_____
Thesis Adviser

_____

_____

_____

_____
Dean of the Graduate School

593508

ii

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

PRESENTATION OF THE PROBLEM

Introduction

Much emphasis is being placed on remedial reading programs for those children who, for some reason, are failing to read as well as teachers, parents, or others feel they should be reading. Remedial reading programs are being offered in the public schools, universities, and by private individuals and clinics. Classroom teachers make efforts to provide individual remedial reading instruction to children in their classrooms.

The classroom teacher is, in most cases, the one person who is responsible for the referral of children to these remedial reading programs. The classroom teacher, who works with the children throughout the school year, should be in a position to make the best judgment of a child's achievement in school work. Are classroom teachers adequately identifying reading underachievers or should some method of discriminate testing be used to better identify those children who are underachieving in reading and who would possibly benefit most from a remedial reading program? Do classroom teachers tend to recommend children for remedial reading programs who are in fact only slow learners rather than underachievers? Are teachers overlooking a significant number of children with average or above average intellectual ability who should be achieving at a level higher than they are achieving in reading? When teachers

are asked to select students for special attention in reading or to identify those who are underachievers in reading do they tend to select those who are the poorest readers regardless of whether they have the mental ability to read better?

Statement of the Problem

The purpose of this study was to determine whether there was a significant difference in the identification of children who were under-achieving in reading when diagnosed by their own classroom teacher and when diagnosed by the use of objective evaluation using the data obtained from the Wechsler Intelligence Scale for Children and the reading test of the California Achievement Test. To facilitate the analysis of this problem, consideration has been given to four main questions, the answers to which should offer suggestions for improving the methods for evaluation of reading achievement.

A. Are there any significant differences in agreement of teacher evaluation and test evaluation of reading achievement of the subjects of this study? This question was answered by comparing various categories of the total population of the study.

B. Are there any significant differences between the number of subjects of the study identified as achievers and underachievers by the teacher evaluation and those identified on the basis of test data? This question was answered by comparing various categories of the total population of the study.

C. Are there any significant differences between the number of subjects identified as achievers by teachers while being identified as underachievers on the basis of test data and the number of subjects

identified as underachievers by teachers while being identified as
achievers on the basis of test data? Various categories of the total
population of the study were compared in order to answer this question.

D. Are there any significant differences in the incidence of
reading underachievement between the various intellectual levels and
between the subjects of different sex?

Need for the Study

There is an evergrowing interest and concern for providing remedial
reading programs for those children who are underachievers in reading.
With the passage of new federal aid to education bills, additional
finances have been made available to public school systems with which
they will be able to provide increased services in remedial reading
programs. It is most important that the effort, time, and money used in
these programs be used to the best advantage for the children of the
schools. One of the main factors upon which the success of any remedial
reading program depends is the correct identification of those pupils
who will be participating in the program. Since most of the referrals
to a remedial reading program are made by the classroom teacher, it is
important to find out how adequately she is able to identify those from
her classroom who are actually reading underachievers. If the class-
room teacher is not doing as adequate a job of identifying these under-
achievers as could be done with more objective testing procedures, it
is important that this be brought to light so that teacher training
programs, reading supervisors of public and private schools, and the
teachers themselves be made to realize the need for improvement, and in
order that corrective measures may be sought to improve this situation.

Definition of Terms

<u>Underachiever</u> refers to a student who is not achieving at the level which his mental ability combined with the criteria of grade placement and chronological age shows him to be capable of achieving in reading.

<u>Mental Ability</u> or <u>intellectual ability</u> refers to the mental age as obtained on the <u>Wechsler Intelligence Scale for Children</u>.

<u>Remedial Reading Instruction</u> is the process of reading instruction, many times on an individual basis, whereby the teacher begins with the underachieving student's present level of reading, working toward the goal of bringing his reading achievement up to his mental ability.

<u>Reading Achievement Grade Placement</u> refers to the grade placement score obtained by the subjects on the <u>California Achievement Test</u> and will sometimes be referred to by the initials AGP.

<u>Anticipated Achievement Grade Placement</u> refers to the grade equivalent score that a given subject might be expected to achieve on the <u>California Achievement Test</u> as determined by the Anticipated Achievement Calculator and will sometimes be referred to by the initials AAGP.

Delimitations

<u>Scope of the Study</u>: This investigation includes a comparison of: (1) Evaluation of reading achievement as determined on the basis of test data; and (2) Evaluation of reading underachievement as determined on the basis of teacher judgment. In order to arrive at the test data evaluation, the subjects were administered the <u>Wechsler Intelligence Scale for Children</u> and the <u>California Reading Test</u>. To obtain the teacher evaluation, a rating instrument, on which each teacher rated the

pupils of her classroom as either achievers or underachievers in reading, was used.

This investigation is not concerned with the methods of teaching reading or with the underlying causes of reading underachievement or with the correction of reading problems.

Assumptions: 1. The reading test (California Achievement Test Form W) used in this study is a reliable and valid measurement of the reading ability of the fourth grade students of this study.

2. The intelligence test (Wechsler Intelligence Scale for Children) used in this study is a reliable and valid measure of the intellectual or mental ability of these fourth grade children.

3. The Anticipated Achievement Calculator is an accurate predictor of expected achievement on the California Achievement Test when used with mental ages derived from the Wechsler Intelligence Scale for Children.

4. Each subject making a score on the reading test of a grade equivalent equal to or greater than his anticipated achievement grade placement score will be classified as an achiever.

5. Each subject making a score on the reading test of a grade equivalent less than his anticipated achievement grade placement score minus one standard error of measurement will be classified as an underachiever.

6. It is not to be assumed that each child classified as an underachiever should be referred to a remedial reading program. To ascertain whether referral should be made would necessitate additional evaluation.

CHAPTER II

REVIEW OF RELATED LITERATURE

Introduction

Much of the literature found which is related to the topic of
reading underachievement is concerned with the causes and prevention
of reading underachievement. The review of the literature for this
study is restricted to methods and procedures used in the identifi-
cation of reading underachievers and will be discussed under the
following areas of interest: (1) importance of identification of
reading underachievers; (2) methods used to identify reading under-
achievers; (3) strengths and weaknesses of instruments of measurement
used to identify reading underachievers; and (4) teacher evaluation
as compared to evaluation as determined by test data.

Importance of Identification of Reading Underachievers

Reading is one of the most important skills which is taught in
our schools. Underachievement in reading is not limited to children
of sub-normal mental ability but is found at all levels of intelligence.
McCullough (1962), has aptly stated the concern which is felt for the
reading underachiever:

> The school's concern for the underachiever in reading is society's
> concern for the worker who fails to master his most valuable tool;
> it is concern for the worker who, in a sense, rejects the most
> refined of modern devices for a primitive approach to the tasks
> which confront him. Multiply him by millions, and he can spell

6

doom of that society.

Reading underachievement or retardation in reading contributes to many problems in our society. Witham (1962), states: "Reading retardation remains the greatest single factor among school dropouts, according to a recent report from the Director of the N.E.A. Project on School Dropouts!"

One of the concerns in relation to reading underachievement is that of the bright underachiever in reading. One of the greatest hindrances in dealing with a bright underachiever is the fact that there is a tendency to underestimate his intelligence and to overestimate his reading achievement. Regardless of the level of intelligence which an underachiever possesses, unless he has special help, he is likely to remain an underachiever. This special help cannot be given to a child unless methods of evaluation which result in diagnosis and treatment of the problem are employed.

The major objective and purpose of any good reading program is that of teaching each child to read to the full extent of his capacity. In order to fulfill this objective, identification of those pupils who are not reading to the full extent of their capacity must be made. Austin (1962), states, concerning poor readers:

> Their identification is the responsibility of every teacher, regardless of his field of specialization or the academic level at which he is teaching. Once identified, diagnostic procedures should be initiated for each disabled reader--and immediate attention should be given to him.

Betts (1957), advances two major reasons for emphasizing the expectancy of reading achievement: ". . . first, to identify pupils who are not achieving in terms of their capacities; second, to avoid insistence upon participation in certain types of reading activities by

pupils who do not stand to profit from them."

Methods Used to Identify Reading Underachievers

Strengths and weaknesses of a number of different methods used for the identification of those who are underachieving in reading have been reported. All of these different methods take into account the factor of mental age.

Betts (1957), discusses different formulas that can be used to determine reading achievement. He states:

> The accomplishment ratio or achievement quotient (A.Q.) has been used for expressing relative educational development to mental development. The A.Q. is found by dividing the educational age by the mental age, thus:

$$(A.Q. = \frac{E.A.}{M.A.})$$

> The derived reading age of the reading test can be substituted for E.A. in order to determine the reading quotient (R.Q.). A reading age equivalent to a mental age would produce an R.Q. of 1.00.

Monroe (1932), has also devised a reading index or formula to use in the determination of reading achievement. This reading index is obtained by making a comparison of the child's composite reading age with his average chronological, mental, and arithmetic age. This formula may be stated:

$$R.I. = \frac{R.Q.}{(C.A. + M.A. + A.A.) \div 3}$$

If a child has a reading index of less than 1.00 he would be considered to be a retarded reader. A reading index of 1.00 or more would be made by those children reading at or above expectancy level.

Another formula which takes into account other factors in addition to mental age is the one reported by Betts (1957), which gives some weight to chronological age. It is as follows:

$$R.Q. = \frac{R.A.}{(2M.A. + C.A.) \div 3}$$

Betts also states: "In general, any means of estimating capacity of achievement should emphasize mental age as one factor."

Johnson (1956), reports on findings of a study which he conducted to determine which of three methods of identification of underachievers was the most significant. He analyzed these three methods separately and also collectively: (1) identification of subjects reading significantly below grade level; (2) identification of subjects with a reading age below their mental age; and (3) identification by the use of a reading index developed by Monroe with .80 being used to divide the severe from marginal disabilities. On the basis of facts revealed from this study, Johnson feels that the easiest and most effective criterion to use is a combination of reading level significantly below grade placement and reading level significantly below mental age level.

The California Test Bureau (1957), has developed an Anticipated Achievement Calculator which was designed ". . . to provide realistic expectancies for an individual pupil by accounting for his mental maturity, his chronological age, and his actual school experience." An Intellectual Status Index (I.S.I.) is assigned for each pupil and his achievement is evaluated on the basis of this rather than on mental age or I.Q. alone. These anticipated grade placements are interpreted as "the norm performance of a nationwide sample of pupils

in the same grade having comparable chronological age and mental age characteristics."

When an over-age pupil and an under-age pupil are compared, the effect of the adjustment through use of the I.S.I. can be demonstrated. Pupils A and B were alike in the following respects:

Actual Grade Placement	5.1
Mental Age	139 months
Grade Chronological Age	127 months
Total Raw Score on CTMM	72
Intellectual Status Index	107

Pupils A and B were different in respect to two factors:

	Pupil A	Pupil B
Chronological Age	118 months	135 months
Intelligence Quotient	114	101

The I.S.I. of 107 for these two pupils represents an adjustment downward from an I.Q. of 114 for the younger pupil and upward from an I.Q. of 101 for the older pupil. Therefore, the same achievement in the basic skills may be expected from both.

Woodbury (1963), in his report on "The Identification of Under-achieving Readers," states:

> When the performances of pupils on standardized reading and scholastic aptitude tests are compared, the units of comparison are likely to be reading ages and mental ages, which do not take into account the many pervasive conditions which contribute to erroneous results and, subsequently, to erroneous judgments.

The purpose of Woodbury's study was to identify the comparative frequencies of underachievement by pupils whose performances on a reading achievement test and a scholastic aptitude test were compared: (1) by the traditional age-ratio procedure, and (2) by a differential index technique which endeavored to control some of the errors which are present in age-to-age comparisons. Tests administered were the Pintner General Ability Tests and the Verbal Series and the Paragraph Meaning

Subtest of the Stanford Achievement Battery. Results showed that only 18 of the 56 pupils identified by the traditional age-ratio method were identified as underachievers according to their differential indices. Woodbury states:

> The findings of this study suggest, then, that many of the differences between reading ages and mental ages of pupils are really differences which are attributable to error and not to the phenomenon of reading achievement varying significantly below expected achievement.

Shanner (1956), reports on the relationships between norms for mental maturity and achievement tests, in which norming procedures similar to those reported to have been used in the development of the California Test Bureau's Anticipated Achievement Calculator are used. The composites for achievement and mental maturity were plotted as histograms and from these norms a formula concerning the expected achievement (Ex) in terms of intellectual grade placement (I.G.P.) was devised as follows:

$$Ex = N + .715 \ (I.G.P. - N)$$

The symbol N indicates the norm for the grade in grade placement units at the time of testing. Shanner states that the system used by the California Test Bureau " . . . is not unique to the California Achievement Test and the California Test of Mental Maturity but is applicable to all achievement and mental maturity tests normed in grade placement and age units respectively.

Larson and Selland (1958), conducted a study to determine the comparison of reading ages with mental ages. The findings of this study showed 36.7 per cent of the subjects having reading ages which were one or more years less than their respective mental ages. Of the 175 subjects reading below, 125 had intelligence quotients of 100 or more on the Kuhlmann-Anderson I.Q. Test. This study took into consideration

only the mental age versus the reading age factor.

One drawback in using only the mental age criterion in making evaluations of reading underachievement is shown by Bliesmer (1954), in his study of the comparison of the reading abilities of bright and dull children of comparable mental ages. The purpose of this research was to determine the extent to which children of equal mental age but markedly different in chronological age and I.Q. tend to be alike with respect to achievement in reading. Subjects who obtained an I.Q. score on the Stanford-Binet of 84 or below composed the dull group while subjects who obtained an I.Q. of 116 or above composed the bright group. From this study Bliesmer made the following conclusions:

(1) Bright children are significantly superior to dull children of comparable mental ages with respect to comprehension, memory for factual details, perception of relationships among definitely stated ideas, and listening comprehension.
(2) Bright children and dull children were the same in reading rate, ability in word recognition and word meaning.
(3) Levels of expectation should not be as high for dull children as for bright children of comparable mental ages.

When using the criteria of age and I.Q. in relation to the improvement of reading, Chansky (1963), found:

There is no empirical support for the belief that children with high I.Q.'s make the greatest progress in remedial reading. It appears to the writer that there is need to seek a criterion other than mental ability in making selections for classes in remedial reading.

Turner (1961), reports findings which differ from those reported by Chansky. In an evaluation of junior high pupils done by Turner to determine which pupils show the most gains in reading during a year in special reading classes of fifteen pupils it was found that:

The greatest improvement in reading ability was made by pupils with average I.Q.'s who were below their reading potential when they entered the class and by those with average and above average I.Q.'s who entered the class in order to improve one phase of their

reading ability. The least gains were made by pupils with I.Q.'s below 80, by those who entered the class as very retarded readers, and by those with visual defects.

Winkley (1962), reports on seven of the common methods used for the identification of reading underachievers. These seven methods are as follows:

1. Mental Age Discrepancy

2. Expected Achievement Grade Placement Tables

3. Anticipated Achievement Calculator (California Test Bureau)

4. Bond and Tinker Formula

5. Stanine Comparison

6. Deviation from Regression Line

7. Monroe Index

In a study of 500 fifth grade children to determine the percentage of the students that would be identified as underachievers, Winkley applied each of these seven techniques. Test data on which the evaluations were based were obtained from each subject's record. These tests included the California Test of Mental Maturity (primary) which had been administered in the third grade and the California Test of Mental Maturity (elementary) which had been administered to the subjects when they were in the fourth grade. The Iowa Tests of Basic Skills and Gates Reading Survey had been administered in the fourth grade also.

A total of 141 of the 500 subjects were identified as underachievers by one or more methods. The Bond and Tinker formula identified 17 per cent of the total 500 as underachievers, this being the highest percentage of any of the methods used. The Anticipated Achievement Calculator identified 6 per cent of the total 500 as underachievers, this being the lowest percentage of any of the methods used. The Bond and

Tinker formula identified the highest percentage of the total 141 as underachievers, that being 60 per cent while the Anticipated Achievement Calculator identified the lowest percentage which was 21 per cent. The Anticipated Achievement Calculator identified 74 per cent of those which were identified by four or more indices while the Bond and Tinker formula identified 95 per cent of those which were identified by four or more indices. Only 1 per cent of those identified by the Anticipated Achievement Calculator were identified by that method alone while 6 per cent of those identified by the Bond and Tinker formula were identified as underachievers by that method only. The Mental Age Discrepancy index identified 48 per cent of the total 141 underachievers as underachievers while this method identified 91 per cent of those which were identified by four or more indices. Of the seven indices or methods used, the Bond and Tinker Formula identified the most underachievers while the Anticipated Achievement Calculator identified the smallest number. From this we can conclude that for this particular situation the Anticipated Achievement Calculator yields a rather conservative evaluation and is not likely to identify as underachievers those who are achievers, while it can be said of the Bond and Tinker Formula that it would be more likely to identify as underachievers some subjects who possibly are not underachievers.

<center>Strengths and Weaknesses of Instruments of Measurement

Used to Identify Reading Underachievers</center>

In order to make an adequate evaluation of reading underachievement, it is very essential that the instruments used in the measurement of mental ability and reading achievement be valid instruments and that

the person using these instruments be aware of the strengths and weaknesses of the instruments.

Strang and Bracken (1957), state:

Any group intelligence test is partly an achievement test; it requires a good deal of reading ability. If the teacher is fortunate enough to have the results of an individual test such as the Stanford Revision of the Binet, the Wechsler Adult Intelligence Scale or the Wechsler Intelligence Scale for Children, he will gain still more information about the individual's mental ability.

Roswell and Natchez (1964), express a very similar opinion concerning intelligence tests: "Those administered above third grade level usually require the ability to read. Obviously the results of such tests cannot be relied upon when given to children with reading deficiencies."

Altus (1952), reports on the relationship between the Wechsler Intelligence Scale for Children and two widely-used group tests administered to fifty-five junior high school students. The correlation between the Wechsler Intelligence Scale for Children full-scale score and the California Test of Mental Maturity total I.Q. score was .81. The correlation of the Wechsler Intelligence Scale for Children expectancy grades in reading (as derived from the mental age of the WISC) gave a Pearsonian coefficient of .84. The correlation between the California Test of Mental Maturity expectancy and reading test was .74. The results show that the WISC had a higher correlation with the reading test than did the CTMM.

Altus (1955), also reports on the relationships between the verbal and non-verbal parts of the California Test of Mental Maturity and the Wechsler Intelligence Scale for Children. Both of these tests yield a language and a non-language I.Q. One hundred children, most of whom had been referred to the guidance clinic for classes of retarded children were the subjects of this study. The results are shown in Table I.

TABLE I

COMPARISON OF MEAN I.Q. SCORES OF WISC AND CTMM

	Verbal	Non-verbal	Total
WISC	82.7	89.8	84.5
CTMM	81.8	88.9	84.8

The difference between total I.Q.'s on the two tests was less than ten points in 65 per cent of the cases.

Cooper (1958), reports findings in regard to predicting school achievement for bilingual pupils. Six tests of intelligence which were either partially or wholly performance or nonverbal were administered to the subjects of this study. Three of the tests were group and three were individual intelligence tests. The intelligence test scores were correlated with scores obtained on the California Achievement Test. Teacher ratings for each child, regarding his school success, were also obtained. The Wechsler Intelligence Scale for Children had the highest positive correlation with the achievement test of any of the six intelligence tests used. The correlation between the Wechsler Intelligence Scale for Children and the California Achievement Test was .77. The next highest correlation was between the Leiter International Performance Scale which is also an individual intelligence test. The correlation between it and the California Achievement Test was .66. The correlation between the California Test of Mental Maturity and the California Achievement Test was .64. The correlation for the verbal portion of the Wechsler Intelligence Scale for Children was greater

than that reported for the full-scale WISC. The findings of this study show that, of the intelligence tests used, the WISC was the best predictor of school achievement on the California Achievement Test.

Barratt and Baumgarten (1957), in a study of the relationships of the Wechsler Intelligence Scale for Children and the Stanford-Binet to school achievement, report the findings given in Table II concerning thirty achievers and thirty non-achievers in grades four to six as defined by teachers' ratings of school performance. Table III shows the correlation between the I.Q. scores and the reading scores as obtained from the California Achievement Test.

TABLE II

COMPARISON OF WISC AND STANFORD-BINET I.Q. SCORES
OF ACHIEVERS AND NON-ACHIEVERS

| | WISC Mean I.Q. | | | Stanford-Binet |
	Verbal	Performance	Full-scale	Mean I.Q.
Achievers	121.17	110.10	117.47	126.47
Non-achievers	82.23	91.50	86.90	88.27

TABLE III

CORRELATION BETWEEN WISC AND STANFORD-BINET I.Q. SCORES
AND CALIFORNIA READING TEST SCORES

| | WISC | | | |
	Verbal	Performance	Full-scale	Stanford-Binet
Achievers	.61	.29	.56	.62
Non-achievers	.51	.30	.63	.46

Smith and Fillmore (1954), report a correlation between the Wechsler Intelligence Scale for Children and the Ammons Full Range Picture Vocabulary Test of .75. The mean I.Q. reported for the Ammons was 107.89 and the mean I.Q. for the WISC was 100.85.

In view of the findings as reported on the basis of these research studies, it can be concluded that the Wechsler Intelligence Scale for Children has a high positive correlation with the California Achievement Test, and with other intelligence tests. It should be pointed out that even though the Wechsler Intelligence Scale for Children does show a positive correlation with other intelligence tests it tends to yield a lower mean I.Q. score. The mean I.Q. of the WISC is reported to have been found lower than that of the Stanford-Binet, the Ammons, and is slightly lower than that of the California Test of Mental Maturity. In view of these findings it should be realized that any statement of the anticipated achievement based on the scores of the WISC as the mental predictor should be considered as a conservative prediction and that if other mental ability predictors were used the anticipated achievement would possibly be higher than that predicted by the WISC.

Tait (1955), reports on findings in regard to the comparison of the California Achievement Test Form AA and four other achievement Tests. When the California Achievement Test is compared to the Metropolitan Achievement Test Form R, the Stanford Achievement Test Form D, the Iowa Every-Pupil Test Form O, and the Coordinated Scales of Attainment Form A, it is found that the California Achievement Test tends to be somewhat less difficult than the norms of the other tests considered and it also tends to be more closely related to actual grade placement.

Finley (1963), reports findings which are in agreement with those reported by Tait. California has approved six tests as acceptable in meeting the state requirements for achievement tests. Finley's study was a comparison of three of the tests, namely the California Achievement Test, the Metropolitan Achievement Tests, and the Iowa Test of Basic Skills. Six classes for each of the three subtest areas (Reading Comprehension, Arithmetic Reasoning, and Language) at grade levels three and five were selected to participate in this study.

At both of the grade levels, the California Achievement Test mean grade placement score for the Reading Comprehension subtest was higher than that of either of the other achievement tests. The results are given in Table IV.

TABLE IV

COMPARISON OF READING COMPREHENSION SUBTEST SCORES ON THREE
READING ACHIEVEMENT TESTS

Test	Norm for Grade	Mean	N
California	3.5	4.24	159
Metropolitan	3.5	3.49	159
ITBS	3.5	3.74	159
California	5.6	6.01	159
Metropolitan	5.6	5.02	159
ITBS	5.6	5.26	159

Stake (1961), in a comparison of the California Achievement Test, the Iowa Test of Basic Skills, the Metropolitan Achievement Test, the Science Research Associates Achievement Test and the Stanford Achieve-

ment Test, concluded that the California Achievement Test over estimated the achievement of the 570 third grade children by more than half a year. Differences reported among the other four tests were non-significant.

Taylor and Crandall (1962), in a comparison of the California Achievement Test, the Iowa Test of Basic Skills, the Metropolitan Achievement Test, the Stanford Achievement Test, and the Science Research Associates Achievement Test, found that children taking the California Achievement Test consistently received higher scores and that children taking the Metropolitan and the Stanford Achievement Tests fairly consistently received lower scores. This comparison was conducted with matched groups of fifth and eighth grade students.

It might be concluded on the basis of the results of the reported studies that the California Achievement Test yields a higher achievement grade placement than several other achievement tests. When making an evaluation of reading underachievement, the possibility exists that some children who are rated as achievers on the basis of this test are in fact underachievers. The use of achievement grade placement scores from the California Achievement Test will be considered to yield a conservative evaluation of reading underachievement. Some children who are actually underachievers will not be so classified on the basis of this test while probably no children will be classified as underachievers who are not underachievers.

Teacher Evaluation as Compared to Evaluation as Determined

by Test Data

Lytton (1961), reports on an experiment conducted to compare

students selected for a special reading program by their teachers with

those selected on the basis of test results. Half of the children

selected for each of three adjustment centers were selected by the

discrepancy between mental age and reading age and half by teachers'

judgment. Pre- and post-tests in reading were administered and average

gains were derived. Some of the children were selected by both the

teacher and on the basis of test results and were considered in both

categories when gains were being evaluated. On the basis of this study,

Lytoon concludes:

> The use of selection by teachers without the use of tests is
> evidently quite a practicable proposition. The teachers'
> choices did quite as well in reading as those selected by tests.
> The average I.Q. of teacher-selected was lower and the average
> A.Q. higher than that of test-selected.

Summary

In this chapter summarized evidence has been cited concerning

the various methods, formulas, and indices which are often used in

evaluation of reading achievement. It has been brought out that even

though the factor of mental age has often been used as the only criteri-

on for an evaluation, it is probably not adequate when used alone.

It has been shown that a combination of factors provides for greater

dependability when making judgments of reading achievement.

Other evidence cited has shown that the achievement tests which

are available for the measurement of reading ability do not all possess

the same characteristics for any level and when using an achievement

test for the purpose of identifying reading achievers and underachievers,

it is important to know the strengths and weakness of the measurement

used. Other evidence, related to testing devices, points out the im-

portance of using a measurement of intelligence which does not depend on reading ability. When assessing mental ability, an individual intelligence test is preferable.

Evidence from one source has brought about the conclusion that evaluation of underachievement based on teacher judgment has been quite as satisfactory as that based on objective test evaluation for the purpose of making recommendations for special reading classes.

The concern of the current study is to determine how comparable teacher judgment is to evaluation based on objective test data. Chapter III includes a discussion of the subjects, procedures, and instruments employed in the study.

CHAPTER III

METHOD AND PROCEDURE

Introduction

This chapter will deal with the population selected for the study,
the tests used to measure intellectual ability and reading achievement,
the method used to determine reading underachievement, and the statis-
tical processes used to test the significance of the comparison
of teacher evaluation of underachievement and the evaluation of under-
achievement as based on data obtained from discriminate testing proce-
dures.

Population Used

The sample used for this study was all the students of eight fourth
grade classrooms in the Portales, New Mexico, Public School system.
This sample is assumed to be a representative sample of children found
at this level.

Portales is a farming community located in eastern New Mexico,
and has a population of 10,000. Eastern New Mexico University is
located in the city of Portales.

Procedure

Before any tests were administered, each teacher was given a rating

sheet on which she was requested to evaluate each of her students, on the

basis of her own judgment, as to whether the student was an achiever or

an underachiever in reading. These rating sheets were distributed to the

teachers on February 9, 1965, and all sheets were checked and returned.

The teachers were asked to make their evaluations at the specific time

chosen as it was felt that they would have had sufficient time by this

phase of the school year to have become thoroughly acquainted with

the progress each child was making in his reading program, and would

perhaps be able to make a more adequate judgment of the child's reading

achievement than had they been asked to evaluate the children earlier

in the school year. The teachers were also asked to check whether

or not they felt each child would profit from remedial reading instruc-

tion. The teachers were not told that a comparison of their evalua-

tions of underachievement was to be made with the evaluations of under-

achievement as determined by testing as it was felt that the rating

would reflect a more true picture of the teachers' judgment if they

were unaware of this fact.

A total of 235 children participated in part or all of this study.

Some children were administered the Wechsler Intelligence Scale for

Children but had moved or were absent on the date the California Achieve-

ment Test was administered. Both tests were administered to 225 children.

Instruments Used in Study

The Wechsler Intelligence Scale for Children was used in this

study for the following reasons: (1) it is an individual intelligence

test. Since any group test is also partly an achievement test, an

individual intelligence test is a better indicator of learning aptitude,

especially when being used with children who may have reading inade-
quacies; (2) the test is divided into major parts, verbal and per-
formance; (3) the test was standardized using a large sample; (4) the
total correlation between the different tests of the Scale is high;
(5) the reliability for the two parts of the test shows .88 for the
total Verbal score and .86 for the total Performance score. The relia-
bility for the Full-scale score is .92. This signifies a rather high
reliability.

The Wechsler Intelligence Scale for Children, generally referred
to as the WISC, was developed by David Wechsler and has grown logically
out of the Wechsler-Bellevue Intelligence Scales which are used with
adolescents and adults. It is published by the Psychological Corporation.
Most of the items in the WISC are from Form II of the earlier scales,
with the main additions being new items at the easier end of each test
to permit examination of children as young as five years of age. The
maximum age for which norms of the WISC are given are through the age
of fifteen.

The WISC has been standardized over a five-year period of experi-
mental tryouts, field testing, and statistical analysis. It was
standardized on a sample of 100 boys and 100 girls at each age from five
through fifteen years. The sample was selected to meet certain sampling
requirements based on U.S. Census Bureau data for 1940, with some adjust-
ment for the recent shift of population toward the west. Basic consider-
ations were: (1) areas of the United States; (2) urban-rural samplings;
and (3) parental occupation. A total of 2200 children were used in the
standardization process.

The WISC differs from most other intelligence tests in that it

does not place emphasis upon the factor of mental age. However, it

does give scales for deriving the mental age, and since one part of this

study requires the use of mental age in determining statistical signifi-

cance, the mental ages were derived for the tests given.

Reliability coefficients have been computed by the split-half

technique, with appropriate correction for full length of the test by

the Spearman-Brown formula. The reliability coefficient is: for the

Verbal score .88; for the Performance score .86; and for the Full-Scale

score .92 for age seven and one-half. For age ten and one-half, the

reliability coefficient is: for the Verbal score .96; for the Perfor-

mance score .89; and for the Full-Scale score .95. The Standard Error

of Measurement for age ten and one-half is: for the Verbal score 3.00;

for the Performance score 4.98; and for the Full-Scale score 3.36.

Wechsler (1949), lists the following seven intelligence Classifi-

cations for the WISC.

I.Q.	Classification	Per Cent Included
130 and above	Very Superior	2.2
120-129	Superior	6.7
110-119	Bright Normal	16.1
90-109	Average	50.0
80-89	Dull Normal	16.1
70-79	Borderline	6.7
69 and below	Mental Defective	2.2

This test was administered by examiners, who had been trained in

the administration of the Wechsler Intelligence Scale for Children,

between the dates of February 10, 1965, and April 8, 1965.

The reading test of the California Achievement Test Elementary

Form W was used in this study for the following reasons: (1) it is a

standardized group test of reading achievement; (2) it gives a reading

vocabulary, reading comprehension, and a total reading score with norms

for each subtest; (3) results can be interpreted in terms of reading age as expressed in months, which is an important factor when evaluating reading underachievement; (4) the elementary form was designed and standardized to be used to test grades four, five, and six; and (5) the Anticipated Achievement Calculator was designed and normed to be used with scores obtained on this test.

The Reading Vocabulary subtest of the California Achievement Test yields a reliability coefficient of .91, the reliability coefficient of the Reading Comprehension subtest is .92, and the reliability coefficient for Total Reading is .95. The Standard Error of Measurement in Grade Placement norms is for Reading Vocabulary 0.4, for Reading Comprehension 0.4, and for Total Reading 0.3.

The Anticipated Achievement Calculator was used to derive the expected level of achievement for the subjects of this study. From the mental age scores obtained on the WISC, an Intellectual Status Index (I.S.I.) was assigned for each subject. This was used as a reference in deriving the Anticipated Achievement Grade Placement from the Anticipated Achievement Calculator. If a subject's Achievement Grade Placement (AGP) was equal to or greater than his Anticipated Achievement Grade Placement (AAGP), he was rated as an achiever. If the subject's AGP was less than his AAGP score minus one standard error of measurement, he was rated as an underachiever. The standard error of measure is .3 of a grade placement unit. A subject's AGP had to be .4 of a grade placement unit below his AAGP before he was classified as an underachiever. The Anticipated Achievement Calculator was used because it uses the criteria of chronological age and actual school experience, in addition to the criterion of mental age, as factors in the evaluation of

underachievement.

Statistical Design

The statistical technique selected for testing the significance of the differences between teacher evaluation and test evaluation of reading underachievement was the chi-square.

The chi-square technique known as the goodness-of-fit technique was used to determine whether there were significant differences between teacher evaluation of underachievement and the evaluation which might be expected on the basis of test data. The chi-square values were calculated using the following formula:

$$X^2_c = \frac{(f_o - f_e - .5)^2}{f_e}$$

in which f_o is the frequency observed or the teacher evaluation and f_e refers to the frequency expected which is the evaluation based on test data. The correction for continuity has been made for this formula. Garrett (1958), states, concerning the table used for calculating the chi-square: "When the table is 2 X 2 fold (when df = 1), X^2 is subject to considerable error unless a correction for continuity is made."

In testing for significant differences in the incidence of under-achievement between certain categories of the sample population and to test significant differences in agreement between teacher evaluation and test evaluation, the following chi-square formula was used:

$$X^2_c = \frac{N(|AD - BC| - N/2)^2}{(A+B)(C+D)(A+C)(B+D)}$$

Garrett (1958), also states in regard to this formula: "When entries in a fourfold table are quite small (for example, 5 or less) Yate's

correction for continuity should be applied to the formula." The
correction of continuity has been included in the formula given.

In all tests of significance, a confidence level beyond the .05
level of confidence will be required before a null hypothesis will be
rejected.

CHAPTER IV

TREATMENT OF DATA AND ANALYSIS OF RESULTS

Introduction

This chapter will be limited to a presentation of the statistical
procedures used in this study and analyses of the results of the data
obtained from these instruments and procedures.

Two chi-square techniques have been used to determine whether
there were any significant differences as stated by null hypotheses.
The chi-square goodness-of-fit test has been used to compare teacher
evaluation of reading achievement with evaluation as determined on
the basis of objective test data. Chi-square, as computed in a four-
fold contingency table, has been used to determine whether agreement
between teacher evaluations and test evaluations differed more signif-
icantly for some groups or categories of the population than it did for
other groups. The total sample population has been analyzed in catego-
ries by sex and intelligence levels in order to determine whether there
were any significant differences.

The sample population has been divided into three groups on the
basis of intelligence levels. These divisions were made as a result
of the I.Q. scores obtained on the Wechsler Intelligence Scale for
Children and based on the Intelligence Classification table as set
forth by Wechsler (1949). The below average level includes those
subjects who made a full-scale intelligence quotient of 89 or below.

The average category is made up of those subjects with full-scale in-
telligence quotients of 90 through 109. The above average category
consists of all subjects with full-range intelligence quotients of 110
and above. According to Wechsler (1949), the normal population distri-
bution of I.Q.'s would show 50 per cent of the subjects falling in the
average range with 25 per cent in the above average range and 25 per cent
in the below average range. Table V shows the distribution of the
sample population of this study by intelligence levels.

TABLE V

DISTRIBUTION OF SUBJECTS BY INTELLIGENCE LEVELS

Intelligence Level	Boys	Girls	Total	Per cent
Below Average	27	18	45	20
Average	56	59	115	51
Above Average	32	33	65	29
n =	115	110	225	100%

In order to make objective evaluations of the reading achieve-
ment of each of the subjects, the Wechsler Intelligence Scale for
Children, the California Reading Test, and the Anticipated Achieve-
ment Calculator were used. A summary of the data obtained from the
use of these instruments is presented in Table VI. These data are
presented by intelligence levels and include the mean full-scale
I.Q. score, the Anticipated Achievement Grade Placement score in

years and months and the Achievement Grade Placement as obtained on the

California Reading Test which is also reported in years and months, and

the difference between the Anticipated Achievement and the actual Achievement.

TABLE VI

MEAN INTELLIGENCE QUOTIENTS, MEAN ANTICIPATED ACHIEVEMENT
SCORES, MEAN ACHIEVEMENT SCORES, AND MEAN DIFFERENCE

Intelligence Level	Number	WISC I.Q.	AAGP Score	AGP Score	Difference between AAGP and AGP
Below Average	45	80	3.77	4.31	+ .54
Average	115	101	4.95	5.19	+ .34
Above Average	65	119	5.94	5.92	- .02
n = 225	Mean 102		4.95	5.23	+ .28

Analysis of Agreement Between Teacher Evaluation and

Test Evaluation

The purpose of this part of the study was to determine whether

there were significant differences between teacher evaluation and

test evaluation of reading achievement when comparing different cate-

gories of the sample population. Tests were made to determine whether

relationships between teacher and test evaluations were the same for

various categories within the total population. Tests of signifi-

cance have been applied to the differences between teacher evaluation

and test evaluation of the above average intelligence group, the average

intelligence group, and the below average intelligence group. These

groups have been further divided into the boys and girls of each group. Table VII shows the distribution of agreement and disagreement between test evaluation and teacher evaluation by categories.

TABLE VII

AGREEMENT BETWEEN TEACHER EVALUATION AND TEST EVALUATION

Category	Agree	Disagree	N
Below Average Total	27	18	45
Below Average Boys	19	8	27
Below Average Girls	8	10	18
Average Total	87	28	115
Average Boys	37	19	56
Average Girls	50	9	59
Above Average Total	57	8	65
Above Average Boys	28	4	32
Above Average Girls	29	4	33
Total Population	171	54	225
Total Girls	84	31	115
Total Boys	87	23	110

On the basis of agreement and disagreement between teacher and test evaluations, the following null hypotheses were tested:

(1) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the total above average category and subjects of the total below average category.

(2) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the total above average category and subjects of the total average category.

(3) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the total below average category and subjects of the total average category.

(4) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the above average boys category and subjects of the above average girls category.

(5) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the average boys category and subjects of the average girls category.

(7) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the below average boys category and subjects of the above average boys category.

(8) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the above average boys category and subjects of the average boys category.

(9) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the below average boys category and subjects of the average boys category.

(10) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the above average girls category and subjects of the below average girls category.

(11) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the above average girls category and subjects of the average girls category.

(12) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the below average girls category and subjects of the average girls category.

Table VIII presents the chi-square value for each of the hypotheses tested.

TABLE VIII

DIFFERENCES IN AGREEMENT OF TEACHER EVALUATION AND TEST
EVALUATION FOR VARIOUS CATEGORIES OF THE POPULATION

Categories Compared	Chi-square
Above Average Total and Below Average Total	10.07 **
Above Average Total and Average Total	3.03
Below Average Total and Average Total	3.13
Above Average Boys and Above Average Girls	.17
Average Boys and Average Girls	4.48 *
Below Average Boys and Below Average Girls	2.0
Above Average Boys and Below Average Boys	1.71
Above Average Boys and Average Boys	3.79
Below Average Boys and Average Boys	.02
Above Average Girls and Below Average Girls	8.9 **
Above Average Girls and Average Girls	.01
Below Average Girls and Average Girls	10.0 **

* With 1 df significant beyond the .05 level of confidence
** With 1 df significant beyond the .01 level of confidence

There are significant differences at the .01 level of confidence

between the agreement of teacher evaluation and test evaluation for
the following categories: (1) above average total and below average
total; (2) above average girls and below average girls; and (3) below
average girls and average girls. There is a significant difference at
the .05 level of confidence between the agreement of teacher evaluation
and test evaluation for the average boys and average girls. Agreement
between teacher evaluation and test evaluation did not differ signifi-
cantly between the other categories compared. On the basis of the
significance found, four of the null hypotheses were rejected. These
were hypothesis number one, hypothesis number five, hypothesis number
ten, and hypothesis number twelve. The alternative hypotheses accepted
can be stated thus:

(1) There is a significant difference in agreement of teacher
evaluation and test evaluation between subjects of the total above
average category and subjects of the total below average category.

(2) There is a significant difference in agreement of teacher
evaluation and test evaluation between subjects of the average
boys category and subjects of the average girls category.

(3) There is a significant difference in agreement of teacher
evaluation and test evaluation between subjects of the above average
girls category and subjects of the below average girls category.

(4) There is a significant difference in agreement of teacher
evaluation and test evaluation between subjects of the below average
girls category and subjects of the average girls category.

An analysis of the differences between teacher evaluation and
test evaluation for the above average total group and the below average
total group taken from Table VII shows that out of a total 65 evalua-

tions for the above average total group, the teacher and the test evaluations agreed in 57 and disagreed in 8 of the cases, while for the below average total of 45 evaluations, the teacher and the test evaluations agreed in 27 and disagreed in 18 of the cases. The significant difference reported is due to the disagreement of test and teacher evaluation of the below average total group. From this it can be concluded that teacher evaluations differ more significantly from test evaluations for the subjects of the below average group than they do for the subjects of the above average total group.

From Table VII, it is found that of the total 18 below average girls, the teacher and test evaluations agreed in 8 and disagreed in 10 of the cases. For the total 59 average girls, the teacher and test evaluations agreed 50 and disagreed 9 times, and for the above average girls it is found that teacher and test evaluations agreed in 29 and disagreed in 4 of the total 33 cases. In the comparison of teacher evaluation and test evaluation of girls of different intelligence levels, it is found that significant differences occur when the below average girls are compared with the average girls and also when the below average are compared with the above average girls. From this it can be concluded that teacher evaluations differ more significantly from test evaluations for below average girl subjects than for either the average girl subjects or the above average girl subjects.

At the .05 level of confidence it is found that agreement between teacher and test evaluations differ significantly for average boys and for average girls. Out of the total 56 average boys, teacher and test evaluations agreed in 37 and disagreed in 19 of the cases. From a total of 59 average girls, teacher and test evaluations agreed in 50

and disagreed in 9 of the cases. From this it can be concluded that teacher evaluations differ more significantly from test evaluations for the boys of the average intelligence category than for the girls of the average intelligence group of this study.

Teacher Evaluation of Reading Achievement Compared to Evaluation Which Might be Expected on the Basis of Test Data

The purpose of this part of the study was to determine whether significant relationships occur between teacher evaluation of reading achievement of a specified category of the sample when compared to test evaluation of reading achievement of that same category or group. To test for significance, the following hypotheses were tested:

(1) There is no significant difference between the number of subjects of the total population of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(2) There is no significant difference between the number of subjects of the above average group of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(3) There is no significant difference between the number of subjects of the boys of the above average group of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(4) There is no significant difference between the number of subjects of the girls of the above average group of this study identified as achievers and underachievers by teacher judgment and the number which

might be expected on the basis of test data.

(5) There is no significant difference between the number of subjects of the average group of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(6) There is no significant difference between the number of subjects of the boys of the average group of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(7) There is no significant difference between the number of subjects of the girls of the average group of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(8) There is no significant difference between the number of subjects of the total below average group of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(9) There is no significant difference between the number of subjects of the boys of the below average group of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(10) There is no significant difference between the number of subjects of the girls of the below average group of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(11) There is no significant difference between the number of subjects of the total girl population of this study identified as

achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

(12) There is no significant difference between the number of subjects of the total boy population of this study identified as achievers and underachievers by teacher judgment and the number which might be expected on the basis of test data.

Table IX shows the distribution of the evaluation of reading achievement on the basis of teacher judgment.

TABLE IX

EVALUATION OF READING ACHIEVEMENT BASED ON TEACHER JUDGMENT

Category	Achievers	Underachievers	Total
Below Average Boys	13	14	27
Below Average Girls	8	10	18
Below Average Total	21	24	45
Average Boys	36	21	56
Average Girls	46	13	59
Average Total	82	33	115
Above Average Boys	25	7	32
Above Average Girls	29	4	33
Above Average Total	54	11	65
Total Boys	74	41	115
Total Girls	83	27	110
Total	157	68	225

Table X is the other of the two tables which are given to show the distribution of the evaluation of reading achievers and under-achievers. Table X shows the distribution of the evaluation of reading achievers and underachievers on the basis of test data.

TABLE X

EVALUATION OF READING ACHIEVEMENT BASED ON TEST RESULTS

Category	Achievers	Underachievers	Total
Below Average Boys	21	6	27
Below Average Girls	18	0	18
Below Average Total	39	6	45
Average Boys	39	17	56
Average Girls	53	6	59
Average Total	92	23	115
Above Average Boys	21	11	32
Above Average Girls	25	8	33
Above Average Total	46	19	65
Total Boys	81	34	115
Total Girls	96	14	110
Total	177	48	225

It is found by a comparison of Table IX and Table X that of the total population of 225 subjects, the teachers evaluated 157 subjects as achievers and 68 as underachievers. Evaluation based on test data shows 177 classified as achievers and 48 classified as underachievers.

In order to test the twelve null hypotheses stated, the following chi-square formula has been used.

$$X^2_c = \frac{(f_o - f_e - .5)^2}{f_e}$$

The observed frequencies are taken from Table IX and the expected frequencies are taken from Table X.

Table XI gives the chi-square values for the twelve hypotheses tested.

TABLE XI

VALUES FOR DIFFERENCES BETWEEN TEACHER EVALUATION
AND TEST EVALUATION

Category compared	Chi-square
Below Average Boys	12.04 **
Below Average Total	58.89 **
Average Boys	.52
Average Girls	7.83 **
Average Total	4.90 *
Above Average Boys	1.64
Above Average Girls	2.00
Above Average Total	4.18 *
Total Boys	2.03
Total Girls	12.89 **
Total	10.06 **

* With 1 df significant beyond the .05 level of confidence
** With 1 df significant beyond the .01 level of confidence

In using the goodness-of-fit chi-square technique, it was found that there are significant differences beyond the .01 level of confidence in the number of subjects identified as achievers and underachievers by teachers and the number identified on the basis of test data for the following groups: (1) below average boys; (2) below average total; (3) average girls; (4) total girls; and (5) total population. Significant differences beyond the .05 level of confidence were found for the following groups: (1) average total; and (2) above average total. Since test evaluation has identified no underachievers in the group of below average girls, the chi-square formula could not be applied to this category. It is recommended that this formula not be used where the frequency expected is less than five.

The null hypotheses for the seven categories which show significant differences were rejected and the following alternative hypotheses accepted:

(1) There is a significant difference in the number of subjects of the below average boys group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(2) There is a significant difference in the number of subjects of the below average total group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(3) There is a significant difference in the number of subjects of the average girls group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(4) There is a significant difference in the number of subjects of the average total group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(5) There is a significant difference in the number of subjects of the total girls group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(6) There is a significant difference in the number of subjects of the total population of this study identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(7) There is a significant difference in the number of subjects of the above average total group identified as achievers and under-achievers by teachers as compared to the number which might be expected on the basis of test data.

Hypothesis number ten cannot be tested by this technique and the other four null hypotheses were accepted as stated.

It should be pointed out that the statistical technique used to test for differences between teacher evaluation and test evaluation takes into consideration only the number of subjects identified by each method and does not take into account the fact that the number identified by teacher evaluation and by test evaluation are not necessarily the same subjects. Some subjects identified as achievers by test data were classified as underachievers by teacher evaluation, while some subjects identified as underachievers by test data were classified as achievers by teacher evaluations.

Analysis of Distribution of Agreement and Disagreement Between

Teacher Evaluation and Test Evaluation

Table XII shows the distribution of individual subjects according

to those rated by both teacher and test data as achievers, those rated

by teacher as achievers and rated by test data as underachievers, those

rated as underachievers by teachers while being rated as achievers

by test evaluation, and those rated as underachievers by both teacher

and test data.

TABLE XII

DISTRIBUTION OF AGREEMENT AND DISAGREEMENT BETWEEN TEACHER
EVALUATION AND TEST EVALUATION

Category	AA	AU	UA	UU
Below Average Total	21	0	18	6
Below Average Boys	13	0	8	6
Below Average Girls	8	0	10	0
Average Total	73	9	19	14
Average Boys	28	8	11	9
Average Girls	45	1	8	5
Above Average Total	46	8	0	11
Above Average Boys	21	4	0	7
Above Average Girls	25	4	0	4
Total Population	140	17	37	31
Total Boys	62	12	19	22
Total Girls	78	5	18	9

In Table XII, the symbol A refers to achiever and the symbol U refers to underachievers. The first symbol of a pair refers to teacher evaluation while the second refers to the evaluation based on test data. For example, AU means that the subjects were classified as achievers by teacher evaluation and as underachievers on the basis of test evaluation.

It is found from an examination of Table XII, that for the total below average category, no subjects were identified as achievers by teacher judgment who were identified as underachievers on the basis of test data. For the above average total category, no subjects were identified as underachievers by teacher judgment who were identified as achievers on the basis of test data.

Out of a total 54 evaluations on which teacher judgment and test data evaluations did not agree, it is found that in 37 of these cases the teacher had identified the subjects as underachievers while they were identified as achievers by the test data. In 17 of the total 54 disagreements, the teacher had identified the subjects as achievers while on the basis of test data they were identified as underachievers. To check for significant differences in regard to these data, the following null hypotheses were made:

(1) There is no significant difference, greater than what would be expected on the basis of chance alone, in the number of disagreements of teacher evaluation and test evaluation between the total population of AU subjects and the total population of UA subjects.

(2) There is no significant difference, greater than what would be expected on the basis of chance alone, in the number of disagreements of teacher evaluation and test evaluation between the total girl

population of AU subjects and the total girl population of UA subjects.

(3) There is no significant difference, greater than what would be expected on the basis of chance alone, in the number of disagreements of teacher evaluation and test evaluation between the total boy population of AU subjects and the total boy population of UA subjects.

These hypotheses were tested using the chi-square goodness-of-fit formula which was used to obtain the values reported in Table XI. The f_e for this set of data was 50 per cent of the total frequency being tested, based on the assumption that in a homogeneous population, 50 per cent of the disagreements between teacher evaluation and test evaluation would fall in the AU category and 50 per cent would fall in the UA category. Table XIII shows the chi-square values obtained when testing the null hypotheses.

TABLE XIII

CHI-SQUARE VALUES OBTAINED FROM AN ANALYSIS OF AGREEMENT AND
DISAGREEMENT BETWEEN TEACHER EVALUATION AND TEST
EVALUATION

Category	Chi-square
Total Population	6.68 **
Total Boy Population	1.16
Total Girl Population	6.26 **

** With 1 df significant beyond the .01 level of confidence

On the basis of the significance reported, two of the null hypotheses were rejected and the following alternative hypotheses

were accepted:

(1) There is a significant difference, greater than what would be expected on the basis of chance alone, in the number of disagreements of teacher evaluation and test evaluation between the total population of AU subjects and the total population of UA subjects.

(2) There is a significant difference, greater than what would be expected on the basis of chance alone, in the number of disagreements of teacher evaluation and test evaluation between the total girl population of AU subjects and the total girl population of UA subjects.

From Table XII, it can be concluded that the significance is a result of a greater number of subjects being evaluated as underachievers by teacher judgment while being evaluated as achievers on the basis of test data.

Incidence of Reading Underachievement at
Various Intellectual Levels

The final procedure to be presented in this chapter is to give consideration to the three intellectual levels and make a determination of whether there are significant differences in the incidence of reading underachievement at any of these levels.

The following null hypotheses were tested:

(1) There is no significant difference in the incidence of underachievement between the below average boys and the below average girls of this study.

(2) There is no significant difference in the incidence of underachievement between the average boys and the average girls of this study.

(3) There is no significant difference in the incidence of under-achievement between the above average boys and the above average girls of this study.

(4) There is no significant difference in the incidence of under-achievement between the below average total and the average total population of this study.

(5) There is no significant difference in the incidence of under-achievement between the below average total and the above average total population of this study.

(6) There is no significant difference in the incidence of under-achievement between the average total and the above average total population of this study.

(7) There is no significant difference in the incidence of under-achievement between the total boy population of this study and the total girl population of this study.

To test these hypotheses, the chi-square formula, given below, was used:

$$X^2_c = \frac{N(|AD - BC| - N/2)^2}{(A+B)\ (C+D)\ (A+C)\ (B+D)}$$

Data were obtained from Table IX. In Table XIV, the chi-square values which were obtained from the application of this formula are presented.

On the basis of the significance reported in Table XIV, two of the seven null hypotheses were rejected while five were accepted. The two alternative hypotheses accepted are:

(1) There is a significanct difference in the incidence of under-achievement between the average boy population and the average girl

population of this study.

(2) There is a significant difference in the incidence of under-achievement between the total boy population and the total girl population of this study.

TABLE XIV

CHI-SQUARE VALUES OBTAINED FROM AN ANALYSIS OF THE DIFFERENCES
IN READING UNDERACHIEVEMENT OBSERVED AT THE DIFFERENT
INTELLECTUAL LEVELS

Intelligence Levels Compared	Chi-square
Below Average Boys and Below Average Girls	.99
Average Boys and Average Girls	3.93 *
Above Average Boys and Above Average Girls	.15
Below Average Total and Average Total	.36
Below Average Total and Above Average Total	1.80
Average Total and Above Average Total	.86
Total Boys and Total Girls	5.43 *

* With 1 df significant beyond the .05 level of confidence

In an analysis of the cause of the significant differences of the two hypotheses rejected, it is found, from an examination of Table IX, that the significant difference between the average boy population and the average girl population is a result of more underachieving boys than girls. The significant difference reported between the total boy population and the total girl population is a

result of significantly more boy underachievers than girls.

Summary

This chapter has presented a detailed analysis of the statistical treatment of the data, with discussion of those hypotheses which were rejected and those which were accepted. An analysis has been made concerning the evaluation of reading underachievement by teachers and by test data. Comparisons between the two methods of evaluation have been made. Differences in the incidence of reading underachievement at different intellectual levels and for the different sexes have been compared statistically.

CHAPTER V

SUMMARY AND CONCLUSIONS

General Summary of the Investigation

This investigation was designed to examine the comparison of

evaluation of underachievement based on teacher judgment with that

based on objective test results. Consideration was also given to

the determination of whether teacher and test evaluation agreed more

closely for some groups of the sample population than for other groups

and whether the incidence of underachievement was greater for some

groups than for others. To determine this, the total population of

the study was divided into categories according to intellectual level

and these categories were further divided according to the sex of the

subjects.

The subjects of the study consisted of the total population of 225

fourth grade students of the Portales, New Mexico, Public Schools. The

Wechsler Intelligence Scale for Children, the California Reading Test

Form W, and the Anticipated Achievement Calculator were used to obtain

data for this study.

Summary of Data

On the basis of the test data, 177 subjects were identified as

achievers and 48 were identified as underachievers. The underachievers

made up 21 per cent of the total sample population. On the basis of

teacher judgment, 157 subjects were identified as achievers while 68 were classified as underachievers. The teachers identified 30 per cent of the total number of subjects as underachievers.

The teachers identified 37 subjects as underachievers who were classified as achievers on the basis of test data and they identified 17 as achievers who were classified as underachievers on the basis of test data.

The total above average group was the only group in which the Achievement Grade Placement mean score was lower than the Anticipated Achievement Grade Placement mean score.

Summary of Results

The chi-square test for determining significant differences was used to test the hypotheses presented in this study. The .05 level of confidence was considered the necessary level of confidence for the rejection of a null hypothesis. The following hypotheses were rejected:

(1) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the total above average category and subjects of the total below average category.

(2) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the average boys category and subjects of the average girls category.

(3) There is no significanct difference in agreement of teacher evaluation and test evaluation between subjects of the above average girls category and subjects of the below average girls category.

(4) There is no significant difference in agreement of teacher evaluation and test evaluation between subjects of the below average

girls category and subjects of the average girls category.

(5) There is no significant difference in the number of subjects of the below average boys group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(6) There is no significant difference in the number of subjects of the below average total group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(7) There is no significant difference in the number of subjects of the average girls group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(8) There is no significant difference in the number of subjects of the average total group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(9) There is no significant difference in the number of subjects of the total girls group identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(10) There is no significant difference in the number of subjects of the total population of this study identified as achievers and underachievers by teachers as compared to the number which might be expected on the basis of test data.

(11) There is no significant difference in the number of subjects of the above average total group identified as achievers and under-

achievers by teachers as compared to the number which might be expected on the basis of test data.

(12) There is no significant difference, greater than what would be expected on the basis of chance alone, in the number of disagreements of teacher evaluation and test evaluation between the total population of subjects classified as achievers by teachers while being classified as underachievers by test data and the total population classified as underachievers by teachers while being classified as achievers by test data.

(13) There is no significant difference, greater than what would be expected on the basis of chance alone, in the number of disagreements of teacher evaluation and test evaluation between the total girl population of subjects classified as achievers by teachers while being classified as underachievers by test data and the total girl population classified as underachievers by teachers while being classified as achievers by test data.

(14) There is no significant difference in the incidence of underachievement between the average boy population and the average girl population of this study.

(15) There is no significant difference in the incidence of underachievement between the total boy population and the total girl population of this study.

General Conclusions

In general, the teachers tended to identify students of below average intellectual ability as underachievers when many of those subjects are achieving at or above their expected achievement level.

This points out the need for adequate procedures for the evaluation of children's mental ability. For the program to do what the name implies, a remedial reading program must provide for children who are reading below their achievement capacity. If children are just slow learners and are reading up to their capacity, they do not belong in a remedial reading class.

Eight children in the above average level of intelligence category who were classified as achievers by their teachers were identified as underachievers by testing. Although this number cannot be proven to be statistically significant, it is still very important that these children not be overlooked when provisions are made for reading under-achievers. They would possibly benefit from special help much more than the 37 subjects identified as underachievers by teachers but found to be achievers by the results of test data.

Many reading underachievers will never be in a remedial reading program. Many do not need to be. It would be impossible to place every child who is reading below his capacity level in a remedial reading program.

In order to refer children to special reading services which will result in individual diagnosis for the child, screening techniques must be used. As has been stated previously in this paper, the purpose of this study has not been to identify children who should be placed in remedial reading programs but to identify all subjects who are reading below their anticipated achievement level. Any child who is underachieving, probably needs help, even though a small amount is all that is necessary in many of these cases; while any child who is reading up to his capacity, regardless of how low that may be,

can be harmed, often considerably, by being pressured to do better and by constant threat of failure.

Probably the most important conclusion which can be based on the findings of this study is that teachers seem to have a misconception of what the term underachievement means and that emphasis should be placed on the fact that achievement or underachievement is an individual matter and cannot be determined by how well the child does in comparison to the rest of the class or according to the norms of any test, but should be determined on the basis of how well he achieves in comparison to <u>his</u> capacity for achievement.

SELECTED BIBLIOGRAPHY

Altus, Grace T. "A Note on the Validity of the Wechsler Intelligence
    Scale for Children," Journal of Consulting Psychology, XVI (June,
    1952), 231.

_____. "Relationships Between Verbal and Non-Verbal Parts of the
    CTMM and WISC." Journal of Consulting Psychology, XIX (April,
    1955), 143-144.

_____. "WISC Patterns of a Selective Sample of Bilingual School
    Children." The Journal of Genetic Psychology, LXXX (December,
    1953), 241-48.

Austin, Martha Lou. "Testing and Teaching the Retarded Reader in Grades
    Two and Three." The Underachiever in Reading. Ed. H. Alan
    Robinson. Chicago: University of Chicago Press, 1962, pp. 33-41.

Barratt, Ernest, and Doris L. Baumgarten. "The Relationship of the
    WISC and the Stanford-Binet to School Achievement." Journal
    of Consulting Psychology, XXI (April, 1957), 144.

Betts, Emmett Albert. Foundations of Reading Instruction. New York:
    American Book Company, 1957.

Bliesmer, Emery P. "Reading Abilities of Bright and Dull Children of
    Comparable Mental Ages." Journal of Educational Psychology,
    XLV (October, 1954), 321-33.

Bond, Guy L., and Miles A. Tinker. Reading Difficulties, Their Diagno-
    sis and Correction. New York: Appleton-Century-Crofts, Inc., 1957.

Burks, Harold. "The Characteristics of Poor and Good Readers as
    Disclosed by the Wechsler Intelligence Scale for Children."
    Journal of Educational Psychology, XLVI (December, 1955), 488-93.

Carey, Helen B. "The Bright Underachiever in Reading." The Under-
    achiever in Reading. Ed. H. Alan Robinson. Chicago: University
    of Chicago Press, 1962, pp. 70-76.

Chansky, Norman M. "Age, I.Q. and Improvement of Reading." Journal
    of Educational Research, LVI (April, 1963), 439.

Clark, Willis W. "Evaluating School Achievement in Basic Skills
    in Relation to Mental Ability." Journal of Educational Research,
    XLVI (November, 1952), 179-91.

Cooper, James G. "Predicting School Achievement for Bilingual Pupils." Journal of Educational Psychology, LXIX (February, 1958), 31-5.

Coull, William H. "A Normative Survey of Reading Achievement of Alberta Children in Relation to I.Q., Sex, Bilingualism and Grade Placement," Alberta Journal of Educational Research, II (March, 1956), 18-29.

Dockrell, W. B. "The Use of WISC in the Diagnosis of Retarded Readers," The Alberta Journal of Educational Research, VI (June, 1960), 86-91.

Durrell, Donald D. Improving Reading Instruction. Yonkers (N.Y.): World Book Company, 1956.

Finley, Carmen J. "A Comparison of the California Achievement Test, Metropolitan Achievement Test and Iowa Test of Basic Skills," California Journal of Educational Research, XIV (March, 1963), 77-88.

Garrett, Henry E. Statistics in Psychology and Education. New York: Longmans, Green and Co., 1958.

Hoyt, Jeanne S. and Dorothy S. Blackmore. "Fifty Seventh Graders: A Comparison of Their Reading Achievement and Expected Achievement in Grades One Through Seven," Journal of Educational Research, LIII (January, 1960), 163-71.

Johnson, G. Orville. "A Critical Evaluation of the Problem of Remedial Reading," Elementary School Journal, LVII (January, 1957), 217-20.

Josephine C.J.S., Sister. "Survey of the Research Related to the Reading Ability of the Gifted," Journal of Educational Research, LIII (February, 1960), 237-39.

Larson, Robert E. and Cynthia Selland. "A Comparison of Reading Ages with Mental Ages," Journal of Educational Research, LII (October, 1958), 55-59.

Lauriana C.S.S.F, Sister Mary. "Actual and Expected Reading Achievement in Detroit," The Catholic Educational Review, LIX (May, 1961), 305-12.

Lytton, H. "An Experiment in Selection for Remedial Education," British Journal of Educational Psychology, XXXI (February, 1961), 79-94.

McCullough, Constance M. "The School's Concern with the Underachiever," The Underachiever in Reading. Ed. H. Alan Robinson. Chicago: University of Chicago Press, 1962, pp. 4-9.

Moe, Iver L. and Frank Nanice. "Reading Deficiencies Among Able Pupils," Journal of Developmental Reading, III (Autumn, 1959), 11-26.

Monroe, Marion. *Children Who Cannot Read*. Chicago: University of
    Chicago Press, 1932.

Plattor, Emma, et. al. "Relationship Between Reading Retardation and
    the Measurement of Intelligence," *Personnel and Guidance Journal*,
    XXXVIII (September, 1959), 149-51.

Robinson, Helen M. "Characteristics of the Underachiever," *The
    Underachiever in Reading*. Ed. H. Alan Robinson. Chicago:
    University of Chicago Press, 1962, pp. 9-18.

Roswell, Florence, and Gladys Natchez. *Reading Disability, Diagnosis and
    Treatment*. New York: Basic Books, Inc., 1964.

Shanner, William M. "Relationships Between Norms for Mental Maturity
    and Achievement Tests," *California Journal of Educational Research*,
    VII (January, 1956), 15-21.

Smith, Louis M. and Arline R. Fillmore. "The Ammons FRPV Test and
    the WISC for Remedial Reading Cases," *Journal of Consulting
    Psychology*, XVIII (October, 1954), 332.

Stake, Robert E. "Overestimation of Achievement with the California
    Achievement Test," *Educational and Psychological Measurement*,
    XXI (Spring, 1961), 59-62.

Strang, Ruth, and Dorothy Kendall Bracken. *Making Better Readers*.
    Boston: D. C. Heath and Company, 1957.

Tait, Arthur T. "A Comparative Study of Five Major Achievement
    Tests," *California Journal of Educational Research*, VI
    (May, 1955), 99-100.

Taylor, Edward A. and James H. Crandall. "A Study of the Norm Equiva-
    lence of Certain Tests Approved for the California State Testing
    Program," *California Journal of Educational Research*, XIII
    (September, 1962), 186-92.

Tiegs, Ernest W. and Willis W. Clark. *Manual, California Achievement
    Tests (Elementary Form W)*. Monterey, California: California
    Test Bureau, 1957.

Turner, Carla S. "Improving Selection of Pupils for Remedial Reading:
    A Report of Research," *The English Journal*, L (January, 1961), 23-33.

Wechsler, David. *Wechsler Intelligence Scale for Children*. New York:
    Psychological Corporation, 1949.

Winkley, Carol K. "Building Staff Competence in Identifying Under-
    achievers," *The Underachiever in Reading*. Ed. H. Alan Robinson.
    Chicago: University of Chicago Press, 1962, pp. 155-61.

Witham, Anthony P. "Techniques for Identifying the Underachiever in Grades Four Through Eight," The Underachiever in Reading. Ed. H. Alan Robinson. Chicago: University of Chicago Press, 1962, pp. 23-27.

Woodbury, Charles. "The Identifying of Underachieving Readers," Reading Teacher, XVI (January, 1963), 218-23.

APPENDIX A

EVALUATION INSTRUMENT FOR TEACHER CLASSIFICATION OF FOURTH

GRADE SUBJECTS AS EITHER ACHIEVERS OR UNDER-ACHIEVERS

TEACHER EVALUATION OF PUPIL ACHIEVEMENT IN READING

1. On the basis of your own judgment please check in the appropriate column whether the following students are achievers or underachievers in reading.

   For the purpose of this evaluation underachiever is defined as a student who is achieving below the level which his mental ability indicates he is capable of achieving. An achiever is defined as a student who is achieving at or above the level which his mental ability indicates him to be capable of achieving.

2. Please put a check mark in the third column for those students whom you feel would profit most from remedial reading instruction.

Name	Achiever	Under-achiever	Would Profit from Remedial Instruction
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			
16.			

APPENDIX B

DATA AND TEST SCORES USED TO CLASSIFY

THE SUBJECTS OF THE STUDY

TABLE B-I

DISTRIBUTION OF TEST DATA AND TEACHER EVALUATION FOR SUBJECTS
WITH BELOW AVERAGE INTELLIGENCE QUOTIENTS

| WISC | | | California Reading Test | | | Teacher | Test | |
I.Q. F-S	M.A. F-S	ISI	AAGP	AGP	Difference	Rating	Rating	Sex
75	106	90	4.0	4.2	+ .2	A	A	M
84	107	91	4.0	3.8	- .2	U	A	F
85	109	92	4.1	2.8	-1.3	U	U	M
89	111	94	4.2	4.5	+ .3	A	A	M
82	99	84	3.6	5.7	+2.1	U	A	F
86	106	90	4.0	4.5	+ .5	U	A	M
74	98	83	3.5	2.7	- .8	U	U	M
73	102	86	3.7	2.8	- .9	U	U	M
87	111	94	4.2	5.1	+ .9	A	A	M
77	85	81	3.4	4.3	+ .9	U	A	F
84	102	86	3.7	5.5	+1.8	U	A	M
87	111	94	4.2	4.3	+ .1	U	A	M
84	114	97	4.4	4.7	+ .3	U	A	F
77	97	82	3.5	3.3	- .2	U	A	F
89	115	97	4.4	3.4	-1.0	U	U	M
84	101	86	3.7	4.9	+1.2	U	A	F
78	95	81	3.4	4.2	+ .8	A	A	M
88	102	86	3.7	5.1	+1.4	A	A	F
75	96	81	3.4	3.0	- .4	U	U	M
79	92	78	3.2	4.9	+1.7	U	A	M
85	107	91	4.0	5.3	+1.3	U	A	M
85	108	92	4.1	4.2	+ .1	U	A	M
83	102	86	3.7	4.2	+ .5	A	A	F
71	90	76	3.1	3.3	+ .2	A	A	M
72	88	74	2.9	4.2	+1.3	U	A	F
73	90	76	3.1	4.0	+ .9	U	A	F
88	100	85	3.6	4.6	+1.0	U	A	F
89	112	95	4.3	4.6	+ .3	A	A	M
79	98	84	3.6	4.0	+ .4	A	A	M
75	102	87	3.8	3.9	+ .1	A	A	F
78	109	93	4.2	4.1	- .1	U	A	M
74	100	85	3.6	4.7	+1.1	A	A	M
65	88	75	3.0	5.1	+2.1	A	A	F
75	93	79	3.3	4.6	+1.3	A	A	F
83	105	90	4.0	4.1	+ .1	A	A	M
72	99	85	3.6	3.2	- .4	U	U	M
72	100	85	3.6	3.9	+ .3	A	A	F
77	103	88	3.8	5.3	+1.5	A	A	M
74	93	79	3.3	4.2	+ .9	U	A	F
72	96	82	3.5	4.6	+1.1	A	A	M

TABLE B-I (Continued)

WISC I.Q. F-S	M.A. F-S	ISI	California Reading Test			Teacher Rating	Test Rating	Sex
			AAGP	AGP	Difference			
85	118	101	4.7	4.9	+ .2	A	A	F
72	105	89	3.9	4.7	+ .8	A	A	M
83	115	97	4.4	4.9	+ .5	A	A	M
82	102	86	3.7	4.6	+ .9	U	A	M
83	116	98	4.5	5.2	+ .7	A	A	F

TABLE B-II

DISTRIBUTION OF TEST DATA AND TEACHER EVALUATION FOR SUBJECTS
WITH AVERAGE INTELLIGENCE QUOTIENTS

WISC I.Q. F-S	M.A. F-S	ISI	California AAGP	AGP	Reading Test Difference	Teacher Rating	Test Rating	Sex
107	133	113	5.5	5.0	- .5	A	U	M
99	123	104	4.9	5.3	+ .4	U	A	F
103	122	103	4.8	5.9	+1.1	A	A	F
92	127	109	5.2	4.3	- .9	U	U	M
91	118	100	4.6	4.9	+ .3	A	A	F
109	122	103	4.8	3.8	-1.0	U	U	M
90	116	98	4.5	4.7	+ .2	U	A	M
109	130	110	5.3	6.5	+1.2	A	A	F
109	141	120	6.0	5.7	- .3	A	A	M
96	112	95	4.3	2.0	-2.3	U	U	M
98	113	97	4.4	5.2	+ .8	A	A	F
107	138	118	5.8	5.3	- .5	U	U	M
101	117	100	4.6	5.6	+1.0	A	A	F
103	132	113	5.5	4.9	- .6	A	U	F
98	116	99	4.6	5.2	+ .6	A	A	M
108	133	113	5.5	5.4	- .1	A	A	M
101	106	90	4.0	5.6	+1.6	A	A	F
104	96	81	3.4	5.7	+2.3	A	A	F
104	127	108	5.2	4.6	- .6	U	U	F
101	111	94	4.2	5.4	+1.4	U	A	M
91	114	97	4.4	5.6	+1.2	A	A	F
109	129	109	5.2	6.5	+1.3	A	A	F
96	112	95	4.3	4.8	+ .5	A	A	M
103	135	111	5.6	5.3	- .3	U	A	M
99	115	97	4.4	5.8	+1.4	A	A	M
100	123	104	4.9	5.6	+ .7	A	A	F
109	126	107	5.1	6.9	+1.8	A	A	F
103	118	100	4.6	5.9	+1.3	A	A	F
109	135	115	5.6	5.8	+ .2	A	A	F
106	122	104	4.9	5.7	+ .8	A	A	F
99	131	112	5.4	6.8	+1.4	A	A	M
101	134	114	5.6	5.1	- .5	A	U	M
101	122	103	4.8	4.8	0	U	A	F
108	129	109	5.2	6.4	+1.2	A	A	F
107	133	113	5.5	5.4	- .1	A	A	F
104	121	103	4.8	5.0	+ .2	A	A	F
101	128	109	5.2	4.2	-1.0	U	U	F

TABLE B-II (Continued)

WISC I.Q. F-S	M.A. F-S	ISI	California Reading Test AAGP	AGP	Difference	Teacher Rating	Test Rating	Sex
109	124	105	5.0	6.6	+1.6	A	A	F
98	126	107	5.1	3.2	-1.9	U	U	F
107	130	110	5.3	5.8	+ .5	A	A	F
107	128	109	5.2	4.5	- .7	U	U	F
99	114	97	4.4	4.9	+ .5	A	A	F
105	144	122	6.1	5.0	-1.1	A	U	M
97	122	104	4.9	5.0	+ .1	U	A	M
105	126	108	5.2	5.5	+ .3	A	A	M
102	125	107	5.1	5.7	+ .6	A	A	M
101	115	98	4.5	4.8	+ .3	A	A	F
92	110	94	4.2	5.5	+1.3	A	A	F
101	131	112	5.4	5.4	0	A	A	F
98	113	97	4.4	5.5	+1.1	A	A	F
109	130	111	5.4	4.5	- .9	A	U	M
90	125	107	5.1	5.1	0	A	A	M
96	110	94	4.2	4.3	+ .1	U	A	M
107	130	111	5.4	6.4	+1.0	A	A	M
100	111	94	4.2	4.1	- .1	U	A	F
101	116	98	4.5	5.1	+ .6	A	A	M
92	112	95	4.3	4.8	+ .3	A	A	M
103	118	100	4.6	4.9	+ .3	A	A	F
103	135	114	5.6	5.8	+ .2	A	A	F
109	130	110	5.3	5.9	+ .6	A	A	F
94	120	102	4.8	5.4	+ .6	A	A	F
101	122	103	4.8	3.5	-1.3	U	U	M
106	122	103	4.8	5.4	+ .6	U	A	F
96	116	98	4.5	4.6	+ .1	A	A	M
99	122	107	5.1	5.6	+ .5	A	A	F
99	124	105	5.0	5.9	+ .9	A	A	F
94	116	98	4.5	5.7	+1.2	A	A	M
103	116	98	4.5	4.8	+ .3	U	A	M
104	117	98	4.6	5.8	+1.2	U	A	M
107	132	112	5.4	6.0	+ .6	A	A	M
99	119	101	4.7	5.2	+ .5	U	A	M
95	116	98	4.5	4.6	+ .1	U	A	M
101	114	97	4.4	6.3	+1.9	A	A	M
96	110	93	4.2	5.2	+1.0	A	A	M
96	121	103	4.8	5.0	+ .2	A	A	M
104	130	110	5.3	6.5	+1.2	A	A	M
109	128	109	5.2	6.9	+1.7	A	A	F

69

TABLE B-II (Continued)

WISC I.Q. F-S	M.A. F-S	ISI	California Reading Test AAGP	AGP	Difference	Teacher Rating	Test Rating	Sex
99	116	98	4.5	5.3	+ .9	U	A	F
93	112	95	4.3	6.1	+1.8	A	A	F
107	117	99	4.6	6.0	+1.4	A	A	M
105	118	101	4.7	5.4	+ .7	A	A	F
107	126	108	5.2	5.2	0	A	A	M
96	106	91	4.0	4.8	+ .8	A	A	F
107	156	133	6.8	3.9	-2.8	U	U	M
99	126	108	5.2	4.5	- .7	A	U	M
109	125	107	5.1	5.7	+ .6	A	A	F
98	135	115	5.6	5.3	- .3	A	A	F
95	110	94	4.2	4.7	+ .5	A	A	M
93	108	92	4.1	4.6	+ .5	A	A	F
93	119	102	4.8	5.9	+1.1	A	A	M
94	118	101	4.7	4.3	- .4	A	U	M
96	118	101	4.7	4.7	0	A	A	F
96	116	99	4.6	4.7	+ .1	A	A	F
97	110	94	4.2	5.1	+ .9	A	A	F
96	121	103	4.8	5.1	+ .3	A	A	M
94	114	97	4.4	5.3	+ .9	A	A	F
101	140	119	5.4	4.9	-1.0	A	U	M
106	119	102	4.8	4.8	0	A	A	M
93	107	91	4.0	4.7	+ .7	U	A	F
107	126	108	5.2	3.7	-1.5	U	U	M
99	121	103	4.8	6.2	+1.4	A	A	F
98	129	110	5.3	4.9	- .4	A	U	M
92	110	93	4.2	4.8	+ .6	U	A	F
91	115	98	4.5	5.1	+ .6	A	A	F
104	118	101	4.7	5.6	+ .9	A	A	F
93	116	98	4.5	5.4	+ .9	A	A	M
90	116	99	4.6	4.0	- .6	U	U	M
104	119	101	4.7	4.8	+ .1	A	A	M
91	110	93	4.2	4.2	0	A	A	M
97	115	98	4.5	4.5	0	U	A	M
104	123	104	4.9	5.3	+ .4	U	A	M
107	126	108	5.2	4.1	-1.1	U	U	M
101	124	105	5.0	5.5	+ .5	U	A	F
104	121	103	4.8	6.3	+1.5	A	A	F
98	125	107	5.1	4.5	- .6	U	U	F

TABLE B-III

DISTRIBUTION OF TEST DATA AND TEACHER EVALUATION FOR SUBJECTS

WITH ABOVE AVERAGE INTELLIGENCE QUOTIENTS

WISC I.Q. F-S	M.A. F-S	ISI	California Reading Test AAGP	AGP	Difference	Teacher Rating	Test Rating	Sex
143	173	146	8.2	6.5	-1.7	A	U	F
122	148	125	6.3	6.6	+ .3	A	A	F
123	150	127	6.4	4.5	-1.9	U	U	F
117	132	112	5.4	5.1	- .3	A	A	M
128	148	125	6.3	6.9	+ .6	A	A	F
117	152	129	6.6	5.9	- .7	U	U	M
112	134	114	5.6	5.8	+ .2	A	A	F
114	142	120	6.0	4.6	-1.4	U	U	M
116	138	117	5.8	5.2	- .6	A	U	F
134	158	134	6.9	4.9	-2.0	A	U	M
125	151	128	6.5	6.9	+ .3	A	A	M
111	139	118	5.8	6.2	+ .4	A	A	F
131	161	136	7.0	6.8	- .2	A	A	F
115	126	107	5.1	6.3	+1.2	A	A	F
114	130	110	5.3	3.3	-2.0	U	U	M
117	137	116	5.7	5.9	+ .2	A	A	F
122	151	129	6.6	7.7	+1.1	A	A	M
111	129	110	5.3	6.3	+1.0	A	A	F
116	144	123	6.2	6.6	+ .4	A	A	F
110	130	110	5.3	5.4	+ .1	A	A	F
125	147	125	6.3	7.1	+ .8	A	A	M
115	136	115	5.6	6.2	+ .6	A	A	F
113	145	123	6.2	5.8	- .4	A	U	M
112	133	113	5.5	4.5	-1.0	U	U	M
115	135	114	5.6	6.1	+ .5	A	A	F
117	134	114	5.6	5.8	+ .2	A	A	F
121	147	126	6.4	6.9	+ .5	A	A	M
120	141	120	6.0	6.4	+ .4	A	A	M
125	147	125	6.3	6.6	+ .3	A	A	F
136	155	131	6.7	7.6	+ .9	A	A	M
117	146	124	6.2	5.5	- .7	A	U	M
136	159	135	7.0	5.0	-2.0	A	U	F
120	138	117	5.8	5.7	- .1	A	A	F
117	136	115	5.6	6.6	+1.0	A	A	F
117	136	115	5.6	5.7	+ .1	A	A	M
121	141	120	6.0	5.9	- .1	A	A	F
135	159	135	7.0	7.9	+ .9	A	A	F

TABLE B-III (Continued)

WISC I.Q. F-S	M.A. F-S	ISI	California Reading Test AAGP	AGP	Difference	Teacher Rating	Test Rating	Sex
130	157	133	6.8	4.9	-1.9	U	U	F
115	142	121	6.0	5.1	- .9	U	U	F
123	137	117	5.8	5.8	0	A	A	M
113	136	116	5.7	6.5	+ .8	A	A	F
110	130	111	5.4	5.7	+ .3	A	A	F
111	123	105	5.0	6.0	+1.0	A	A	M
115	138	118	5.8	5.7	- .1	A	A	M
112	146	125	6.3	6.3	0	A	A	M
111	124	106	5.0	5.3	+ .3	A	A	F
113	133	114	5.6	6.7	+1.1	A	A	M
118	145	124	6.2	6.5	+ .3	A	A	M
126	145	124	6.2	6.9	+ .7	A	A	M
112	134	114	5.6	4.4	-1.2	U	U	M
110	122	103	4.8	5.8	+1.0	A	A	M
120	148	125	6.3	5.6	- .7	A	U	M
115	143	121	6.0	7.1	+1.1	A	A	F
120	144	122	6.1	6.0	- .1	A	A	M
111	122	103	4.8	5.6	+ .8	A	A	F
121	133	113	5.5	7.0	+1.5	A	A	M
117	143	121	6.0	5.7	- .3	A	A	M
122	153	130	6.6	5.6	-1.0	U	U	M
117	137	116	5.7	5.9	+ .2	A	A	M
125	145	123	6.2	6.2	0	A	A	F
113	131	111	5.4	5.4	- 0	A	A	M
117	131	111	5.4	5.0	- .4	A	U	F
113	129	110	5.3	5.9	+ .6	A	A	F
112	133	114	5.6	5.2	- .4	U	U	F
115	138	118	5.8	4.9	- .9	U	U	M

VITA

Twilla Lurline Avard Jordan

Candidate for the Degree of

Doctor of Education

Dissertation: TEACHER EVALUATION OF READING UNDERACHIEVEMENT OF
                INTERMEDIATE GRADE STUDENTS AS COMPARED TO EVALUATION
                OF READING UNDERACHIEVEMENT BY DISCRIMINATE TESTING
                PROCEDURES

Major Field: Elementary Education - Reading

Biographical:

    Personal Data: Born at Carleton, Oklahoma, August 24, 1934,
        the daughter of Joseph S. and Irene Foley Avard.

    Education: Attended grade school at Manitou, Oklahoma, and
        Mountain Park, Oklahoma; graduated from Mountain Park
        High School in 1952; attended Altus Junior College
        1955-1956; received Bachelor of Science degree from the
        Oklahoma State University, Stillwater, Oklahoma, with
        a major in Home Economics Education in May, 1958;
        attended Southwestern State College, Weatherford, Okla-
        homa during the 1960 summer term; received the Master
        of Science degree from the Oklahoma State University,
        with a major in Elementary Education, in May, 1961;
        completed requirements for the Doctor of Education de-
        gree in August, 1965.

    Professional Experience: Taught third grade at the Indiahoma
        Elementary School, Indiahoma, Oklahoma, 1958-59; taught
        chemistry and general science at the Butler High School,
        Butler, Oklahoma, 1960-61; served as graduate assistant
        at the Oklahoma State University, 1961-63; taught fourth
        grade at the R. M. James Elementary School, Portales,
        New Mexico, 1963-65.

    Professional Organizations: National Education Association;
        New Mexico Education Association; Omicron Nu; Phi Upsilon
        Omicron; Kappa Delta Pi; International Reading Association.