1956

An Investigation of the New Scott-Foresman Basic Reading Tests, the Lee-Clark Reading Readiness Test, the Lee-Clark Reading Tests, and the Gates Reading Readiness Test in Primary Grades

Milton A. Dallman
Central Washington University

AN INVESTIGATION OF THE NEW SCOTT-FORESMAN BASIC READING TESTS, THE LEE-CLARK READING READINESS TEST, THE LEE-CLARK READING TESTS, AND THE GATES READING READINESS TEST IN PRIMARY GRADES

A Thesis

Presented to

The Graduate Faculty

Central Washington College of Education

In Partial Fulfillment

of the Requirements for the Degree

Master of Education

---

by

Milton A. Dallman

August 1956

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Education, in the Graduate School of the Central Washington College of Education.

Approved by:

_____

Dr. Eldon E. Jacobsen

_____

Dr. Maurice L. Pettit

_____

Dr. Ernest L. Muzzall

_____

Mr. Edward K. Erickson

_____

Mr. Marvin J. Schroeder

ACKNOWLEDGMENTS

Grateful acknowledgment is extended to Dr. Eldon E. Jacobsen for his assistance and encouragement in directing the writing of this paper and to Dr. Maurice L. Pettit and Dr. Ernest L. Muzzall for serving on the thesis committee.

Further acknowledgment is extended to Mr. Marvin J. Schroeder and Mr. Edward K. Erickson of the Ellensburg Public Schools for their guidance and advice offered during the work toward the Master's degree and for also serving on the thesis committee.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

CHAPTER I

INTRODUCTION

A variety of instruments and techniques are employed in measuring school achievement. The use of standardized tests is usually recommended, however, as the most objective and best single means of measuring achievement and of identifying the retarded.[1]

Since reading is taught in all schools, tests have been most numerous in this field. Although most reading tests are good, many of them are poorly conceived and inadequately standardized. Many of them do not provide the usual and necessary data on reliability and validity.[2] Among the many tests on which additional validational data is desired are the Lee-Clark Reading Tests and the New Scott-Foresman Basic Reading Tests.

A. PURPOSES

There were four purposes for the study:

1. To determine the interrelationships existing between the primary grade tests within the New Scott-Foresman Basic Reading Test series.

---

[1]Walter Monroe (ed.), Encyclopedia of Educational Research (New York: The Macmillan Company, 1950), p.1461.

[2]Frank Freeman, Theory and Practice of Psychological Testing (revised; New York: Henry Holt and Company, 1955), p. 392.

2. To determine the relative efficiency of two reading readiness tests, the Lee-Clark test and the Gates test, in predicting children's later success in reading.

3. To determine the relative efficiency of the Lee-Clark Reading Test: Primer and the Lee-Clark Reading Test: First Reader in predicting subsequent success in reading.

4. To establish norms for the New Scott-Foresman Basic Reading Tests that may be used in the community in which the study was conducted.

B. IMPORTANCE OF THE STUDY

Tests of educational achievement are particularly valuable in the primary and intermediate grades where they are used to measure pupils' basic skills. For these purposes standardized tests, with their norms of performance and their diagnostic methods, provide teachers and others with superior instruments for the measurement of pupil progress.[3]

More than a hundred standardized silent-reading tests have been developed for use in the elementary and secondary schools.[4] The results of these reading tests

---

[3] Ibid, p. 400

[4] Paul Witty and David Kopel, Reading and the Educative Process (Boston: Ginn and Company, 1939), p. 237.

are commonly used by teachers and administrators as a means for measuring achievement, diagnosis of particular difficulties, and as prediction of future reading success and success in other subjects. After a test has been constructed there remains the evaluation of the instrument with respect to validity and reliability. Despite optimum testing conditions, no test is a perfect instrument. Hence, every test should be accompanied by statements of reliability and validated against other accepted criteria. The Scott-Foresman tests, although apparently validated on their standardization sample, have formulated no information as to the reliability of their instruments nor have they provided validational data by any other means.

The most important question that needs to be raised regarding any psychological test concerns its validity, that is, the degree to which the test actually measures what it purports to measure. The determination of validity usually requires an independent external criterion of that which the test is designed to predict.[5]

Results of educational achievement tests are helpful in forecasting the subject's probable future level and quality of learning in the several school areas.[6] It is

---

[5]Anne Anastasi, Psychological Testing (New York: The Macmillan Company, 1954), p. 29.

[6]Freeman, op. cit., p. 8.

worthwhile to consider all tests as behavior samples from which predictions regarding other behavior may be made.[7] If tests are used for predictive purposes it probably is the responsibility of local schools to determine for themselves the effectiveness of instruments that do not have complete and informative data in that respect.

All first grade teachers need to know which pupils are ready to read when they enter the grade. The first grade has often been said to be the crucial point in the child's education.[8] Readiness may be determined by the use of tests, teacher observation, or a combination of these and other means but tests probably provide the simplest method of selection.

A comprehensive and continuing testing program is a powerful educational instrument. The nature of the test influences greatly how and what pupils study and how and what teachers teach. Good tests clarify new objectives and give added meaning to old. They stimulate good teaching and learning procedure.[9] This study endeavors to determine, within limits, the value of some of the standardized tests

---

[7]Anastasi, op. cit., p. 23.

[8]J. Murray Lee, Willis Clark, and Dorris Lee, "Measuring Reading Readiness," Elementary School Journal, XXXIV (May, 1934), p. 656.

[9]Walter W. Cook, "Achievement Tests," Encyclopedia of Educational Research, (ed. Walter Monroe, New York: The Macmillan Company, 1950), p. 1463.

used in the past in the locale of the study and, as far as possible, the value of the New Scott-Foresman Basic Reading Tests being used at present in two of the Ellensburg, Washington elementary schools. Such determinations would, it is believed, hold implications for other users of the instruments under surveyance.

Anastasi says, "Any norm, however expressed, is restricted to the particular normative population from which it was derived."[10] This implies, and others agree, that the establishment of local norms for any test in use is desirable. Since no two communities are exactly alike with respect to economic status, ethnic background, social standards, etc. and since these factors influence testing results to a considerable extent, the establishment of local norms is a necessary procedure in conducting the most efficient testing program. This particular study is especially important for teachers of the Lincoln School for it provides them with additional normative information and standards for comparisons of their own children.

In summary, the study is important because it attempts to provide additional validational data on several standardized reading tests, to provide information with regard to the prediction of future success in reading, and to establish

---

[10]Anastasi, op. cit., p. 87.

local norms for the Scott-Foresman tests which will enable teachers using them to better understand the relative reading accomplishments of children under their guidance.

C. PROCEDURE

Instruments. The New Scott-Foresman Basic Reading Tests consist of separate tests to accompany the corresponding New Scott-Foresman Basic Readers. The tests at the primary level are:

1. The New Pre-Primers, 1_1
2. The New Fun With Dick and Jane, 1_2
3. The New Our New Friends, 1_3
4. The New Friends and Neighbors, 2_1
5. The New More Friends and Neighbors, 2_2
6. The New Streets and Roads, 3_1
7. The New More Streets and Roads, 3_2

The Scott-Foresman tests are designed as tests of reading achievement and diagnosis to be administered upon completion of the work in the corresponding Scott-Foresman reader.[11] Thus, for example, the test for the second grade reader, The New Our New Friends, is to be given to the

---

[11]Marion Monroe, General Manual for the New Scott-Foresman Basic Reading Tests (Chicago: Scott-Foresman Company, 1954), p. 1.

group of children who have been reading that particular book under the direction and guidance of the room teacher.

It is readily apparent that such a testing program does not follow the ordinary procedures used in standardized testing since there is no recommended time schedule for their administration. Ordinary standardized achievement tests are usually recommended for use in the spring or fall or both. The Scott-Foresman tests are administered upon completion of the reader and, depending upon the progress of the reading group, may be given at anytime during the school year.

One might consider this type of testing as similar to many of the teacher-made objective tests used in evaluation of many school subjects. The Scott-Foresman tests, however, are more standardized because they are published and have established norms. This type of testing might be also comparable to the type of evaluation accomplished in the non-graded type class where the reader and the test are given to children supposedly "ready" for them, having advanced to one reader from another through the various exercises and experiences advocated and practiced by the individual teachers.

The Lee-Clark Reading Tests are designed as instruments of reading achievement for the first and second grades.

They are comparable to many of the well-known standardized reading achievement tests and advocated for use in a spring or fall testing program. The Lee-Clark Primer Test was administered to Lincoln School first graders in the 1954-55 school year. This group constitutes the second grade class for the present study. The Lee-Clark First Reader Test was given to Lincoln School second graders during the 1954-55 school year and they constitute the third grade group of the present study.

The Lee-Clark Reading Readiness Test and the Gates Reading Readiness Test are well-known and commonly used reading readiness tests. Reading readiness tests are commonly advocated for use in the latter part of a child's kindergarten year or during the first part of first grade. They are designed "to appraise the abilities measured, to provide guidance for various reading readiness activities, and to forecast roughly the status of pupils when they enter first grade."[12] Both tests were given to children during their kindergarten year. The Lee-Clark Readiness Test was given to children who were first graders during the study and the Gates test was given to children who were second graders during the study.

---

[12]Arthur Gates, Manual of Directions for Gates Reading Readiness Tests (New York: Bureau of Publications, Teachers College, Columbia University, 1939), p. 7.

Locale. The study was conducted in the community of Ellensburg, Washington. Ellensburg is largely an agricultural community that is located in the center of the Kittitas Valley in Central Washington. The valley is flanked on the west and north by the Cascade Mountains and on the south and east by a series of dry, rolling hills. The city of Ellensburg has a population of approximately 8500 people, including some 1400 students who attend the local college. The combined population of Ellensburg and its adjacent rural areas totals approximately 14,000 people. Many of the rural children of the valley attend the Ellensburg schools. Transportation by school bus is provided. Ellensburg is a fairly stable community as evidenced by the fact that the population has not greatly increased in the past two decades. Ellensburg is somewhat above average regarding wealth per capita.

The educational facilities of the community consist of three public elementary schools, one private parochial elementary school, a public junior high, public senior high, and a college largely devoted to the training of teachers. The public schools maintain close contact with the college through a teacher-training program and consultation with educational specialists and staff members on particular problems.

Each of the public elementary schools is relatively heterogenous, drawing their pupils from families having a wide variety of economic backgrounds in the city and country. The elementary schools make no attempt to group children on the basis of ability or other factor in the rooms at each grade level and, consequently, they are similar in pupil composition. Teachers use the group method in the teaching of reading and some of the other subjects. Most teachers maintain three distinct reading groups during the year and, although fluctuation is allowed in this grouping, they usually remain relatively stable.

Two of the three public elementary schools, Lincoln and Washington, were used for the accumulation of data pertaining to the study. The majority of the information, however, came from the results of the testing program at the Lincoln School. All primary teachers at the Lincoln School used the New Scott-Foresman Basic Reading Tests as one of the bases for their evaluation of reading progress during the 1955-56 school year. The number of children from Lincoln tested in this program totaled 308. The number of children from the Washington School was considerably reduced since the Scott-Foresman testing was not initiated until after mid-year.

Statistical Method. The study attempted to establish relationships between various reading tests and for this reason the Pearson product-moment correlation coefficient was used. Essentially, a correlation coefficient expresses the degree of correspondence, or relationship, between two sets of scores and may range anywhere from -1.0, a perfect negative relationship, to +1.0, a perfect positive relationship. Scattergrams were prepared for each of the correlations. The basic formula, the formula for the standard error, and an example of the development of one of the coefficients is included in Appendix A.

Since the major focus of the study centered around the New Scott-Foresman tests, all correlations use these tests as one of the comparative instruments. The correlations computed were as follows:

1. Lee-Clark Reading Readiness Test scores with Scott-Foresman test scores.

2. Gates Reading Readiness Test scores with Scott-Foresman test scores.

3. Scott-Foresman test scores with other Scott-Foresman test scores.

4. Lee-Clark Primer Test scores with Scott-Foresman test scores.

5. Lee-Clark First Reader Test scores with Scott-Foresman test scores.

A total of thirty correlations were computed for the study. Average correlations, using the Fisher z formula, were also computed in a number of cases bringing the total number of correlations to thirty-five.

D. LIMITATIONS OF THE STUDY

There were two major limitations of this study. First, the samples of population used in the various correlations were limited to two schools and the test scores available on the primary children in those schools. Not all primary children have attended kindergarten. Those not attending were not given reading readiness tests. This resulted in the small number of cases for some of the correlations. Second, the time intervals between the administrations of the Scott-Foresman tests varied with each reading group and homeroom. As expressed earlier, tests were given to reading groups on the basis of rate of progress. Some children spent more time in a particular reader than others.

To clarify the limitations further, it should be kept in mind that the use of sampling statistics rests on the assumption that the sampling has been random. Random sampling implies that "every individual in the population

has an equal chance of being chosen."[13] This study has included the samples that were available. Guilford describes this as an "incidental sample" and states, "Results thus obtained can be generalized beyond such groups with considerable risk."[14] This incidental sample approach was true of the correlations between the reading readiness tests and the Scott-Foresman tests. The Gates readiness test scores were available on only a small percentage of the children who were second graders at the time of the study since these tests were given during their kindergarten year. The Lee-Clark readiness test scores were available on a larger percentage of children who were first graders during the time of the study but, again, was limited because all the children did not attend kindergarten.

The incidental sample approach was not true in the correlations of the tests within the Scott-Foresman series since the individuals were tested in groups according to their ability to read. Guilford defines this as stratification in sampling and says:

> . . . is a common procedure used to help prevent biases. It is a step in the direction of experimental control since it operates with subgroups of more

---

[13] J. P. Guilford, Fundamental Statistics in Psychology and Education (New York: McGraw-Hill Book Company, 1950), p. 177.

[14] Ibid, p. 180.

homogenous composition within the larger population.[15]

The first limitation also involved the number of cases used in several of the correlations. If one asks, "How small is N before we have a small sample?" the answers from different sources will vary. There is general agreement, however, that the division is in the range of 25-30. Sampling distributions depart from the normal form more and more as N decreases. For this reason the correlations indicated in the chapter of results should be considered in view of the number of cases involved and the standard error for each correlation. Since the standard error of measurement of 'r' is inversely proportional to the size of the sample, the fewer the number of cases in the sample the greater will be the standard error and, hence, the less significant the coefficient.

The second limitation involved the time interval between testings. A definite time schedule was not required for completion of the Scott-Foresman readers. Rather, individual teachers took as much time as they believed necessary for completion of the book. This is believed proper in view of the individual differences existing within subgroups of any particular grade and room. Freeman

---

[15] Ibid, p. 178.

states:

> The effects of practice and learning during the interval will depend upon the content of the test being used and upon the examinee's experiences during the interval. For example, if some months have elapsed between two administrations of an educational achievement test, different pupils may have had different amounts of and qualities of instruction during the period.[16]

The test scores would, then, reflect the time and instructional differences.

Two other questions regarding the administration of the Scott-Foresman tests may be asked. First, do teachers have a tendency to teach "toward the test" in situations such as this? It is assumed that they did not but when it is seen that the various sections of each succeeding test are measuring many of the same factors one might wonder if there would be a tendency for some teachers to stress improvement in those particular abilities and neglect others that are omitted in the tests and less easily measured. Second, since each succeeding test measures practically the same skills and abilities, one would wonder if this could be termed "practice effects" and if such effects were tending to cause each succeeding test score to raise in a significant manner. This would have to be checked by methods beyond the scope of those proposed for this study.

---

[16]Freeman, op. cit., p. 25.

Numerous other questions concerning the use of the Scott-Foresman tests in the primary grades arise but this study has been limited to an initial survey of the tests currently in use in the schools cited and the scores available on school records.

CHAPTER II

REVIEW OF THE LITERATURE

Much has been written in regard to tests and testing programs. A survey of the test literature reveals information of varying amounts and qualities that pertains to the four standardized tests on which this study is focused. The four tests present a logical breakdown for examination of the literature and the reviews are presented in that manner.

A. THE NEW SCOTT-FORESMAN BASIC READING TESTS

The New Scott-Foresman Basic Reading Tests were copywrited in 1954. The manual accompanying the tests does not clearly specify the population used for standardization. The manual states:

> The first tentative standardization of the New Basic Reading Tests is based on a population of approximately 20,000 first-, second-, and third-grade children from many representative American areas. The children were drawn from typical town and city school systems, as well as from rural schools, in all sections of the United States. The majority of the teachers rated their communities as average. A small group of teachers rated their communities superior, but these were balanced by an almost equal number rated low. As nearly as could be determined, the population appeared to be a typical one.[1]

---

[1]Marion Monroe, General Manual for The New Basic Reading Tests (Chicago: Scott-Foresman Company, 1954), p. 3.

Information as to the reliability and validity of the New Scott-Foresman Tests has not been published. The reason for this is explained, in part, by the philosophy of the originators regarding the purposes of their tests. A letter from Mary Steuteville of the Scott-Foresman Company in response to a question from the writer regarding the reliability and validity of their tests is quoted in part.

> You'll find it a little hard to compare the New Basic Reading Tests with standardized tests for grade placement. Our tests are tests of mastery of skills that have been taught in the New Basic Reading Program, and aren't designed for indicating grade placement, although, of course, they will help a teacher determine how to group pupils.[2]

It appears that the Scott-Foresman tests attempt to measure many of the reading skills found in other standardized reading achievement tests and should, then, include as much information as possible on the reliability and validity of their instruments. It seems reasonable to ask what any achievement test measures if not mastery of certain skills. It is important that teachers know how well any achievement test meets the criteria of a good measuring instrument if the teacher is to depend on the test for evaluation of progress, diagnosis, or for any other major classroom purpose.

---

[2]Letter from Mary Steuteville, Scott-Foresman Company, Chicago, Illinois, dated February 24, 1956.

B. THE LEE-CLARK READING READINESS TEST

The Lee-Clark Reading Readiness Test has become one of the most widely used instruments for determining: (1) whether entering first grade pupils are ready for reading instruction, and (2) the probable length of time before formal reading activities are advisable for immature children. The test booklet provides interesting and easily administered exercises or test items, which are designed primarily to predict probability of success in first-grade reading.[3]

Experimental work and research have shown the Lee-Clark Reading Readiness Test to predict the first grade reading success or failure of pupils somewhat better than some of the commonly used group intelligence tests. It also correlates well with success in reading as measured years later in the sixth grade.[4]

The reliability coefficient for the Lee-Clark Reading Readiness Test, obtained by the split-half method and corrected by the Spearman-Brown formula for 170 entering first grade pupils is .93. Similar reliability coefficients for the sub-tests vary from .83 to .94.[5]

---

[3] Catalog of the California Test Bureau (Los Angeles: California Test Bureau, 1955), p. 46.

[4] Ibid.

[5] Ibid.

Review of the test literature revealed three studies that used the Lee-Clark Reading Readiness Test as a measure of predicting reading achievement in later grades. Two of these used the old (1934) edition of the test for their study. A fourth study compared teachers prognosis against the results of the Lee-Clark test as a prediction of future reading success.

Moreau studied 275 pupils in ten San Francisco schools to determine the degree to which first grade intelligence and reading readiness tests predicted reading achievement over a period of the first five years of elementary school. Reading achievement, as measured by the California Basic Skills Test, of the pupils in the low sixth grade was correlated with scores on the Pintner-Cunningham Intelligence Test and the Lee-Clark Reading Readiness Test which they had taken in the first month of the first grade. The Lee-Clark test correlated .654 for boys, .456 for girls, and .462 for the total sample with the reading achievement test. Mental age, as measured by the Pintner-Cunningham test correlated .490 for boys, .447 for girls, and .532 for the total sample with the reading achievement test. I.Q., measured by the same intelligence test, correlated .416 for boys, .320 for girls, and .392 for the total sample with

the reading achievement test.[6]

It was concluded from Moreau's study that:

> (1) Scores on the Lee-Clark Reading Readiness Test and the Pintner-Cunningham Test of Intelligence given in the first grade predict reading achievement in the first five grades almost as reliably as in the first grade. (2) The intercorrelation between these two tests indicate slight over-lapping, but considerable difference between the functions which they measure. (3) Mental age seems to be a better predictor of reading achievement than I.Q. but it exceeds only slightly the Lee-Clark reading readiness score for the total sample. (4) Although these tests overlap somewhat, they measure different enough functions to justify continued use of both tests.[7]

If, as Moreau states, the tests are measuring different functions it would seem reasonable to combine the results of the two measures using multiple correlations for better prediction.

Robinson made a study of the relative values of two standardized tests in predicting first-grade achievement in reading. One of the primary purposes of the study was to determine the relative value of a group intelligence and a group reading readiness test in predicting reading ability in the first grade. Ninety-four children entering the low first grade (chronological ages ranged from 66 months to 83 months) in two schools of the Sacramento, California schools

---

[6]Margaret Moreau, "Long Term Prediction of Elementary Reading Achievement," California Journal of Educational Research, Volume 1, Number 4, (September, 1950), pp. 175-178.

[7]Ibid, p. 178.

were administered the Pintner-Cunningham Primary Test, the Lee-Clark Reading Readiness Test, and a vocabulary test constructed from the pupils' reading workbooks. Mental age, with two administrations of the test, correlated .559 and .559 with the Lee-Clark Reading Readiness score. Mental Age, with two administrations of the test, correlated .432 and .582 with the vocabulary test. The Lee-Clark scores correlated .671 with the vocabulary test. The Lee-Clark test thus predicted with greater accuracy than the other instruments the potential ability of the 94 children to succeed in first grade reading.[8]

Lee, Clark, and Lee correlated the Lee-Clark Reading Readiness Test with the Lee-Clark Primer Test for 100 children and found a correlation of .49. The correlation of their reading readiness test with success in reading as measured by the Gates Silent Reading Tests, Types 1, 2, and 3 was .54 for 100 cases. The correlation of the reading readiness test and the Lee-Clark Reading Test: Primer with a group of pupils all of whom had kindergarten experience was .68 for 92 children.[9] The Lee-Clark Reading Readiness

---

[8]Agnes Robinson, "A Study of the Relative Values of Two Standardized Tests in Predicting First-Grade Achievein Reading", (Unpublished Master's Thesis, Sacramento State College, Sacramento, California, 1952). Taken from abstract provided by the California Test Bureau.

[9]J. Murray Lee, Willis Clark, and Dorris Lee, "Measuring Reading Readiness," Elementary School Journal, Volume 34, (May, 1934), pp. 656-666.

Test thus predicted reading success with a significant degree of accuracy when the Gates Silent Reading Tests and the Lee-Clark Primer Test were used as the criteria of comparison.

Henig performed a study aimed to determine the comparative forecasting value of the Lee-Clark Reading Readiness Test and of teachers estimates of their pupils likelihood of success in learning to read. Tests were given to 98 beginning first graders about three weeks after the opening of school. The teachers ranked their pupils prior to administration of the test using the commonly accepted indications of reading readiness as ability to talk in sentences, retell a story, follow directions, discriminate between sounds, etc. as the basis for their estimates. The contingency coefficient between the test results and teachers prognosis of pupil success in reading equaled .60. This, of course, indicates a substantial degree of agreement between the two measures. Henig suggests that teachers were just as successful in predicting the degree of success of children as was the standardized test.[10] It would seem that, if tests overlap in measuring the same functions, as

---

[10]Max Henig, "Predictive Value of a Reading-Readiness Test and of Teachers Forecasts," Elementary School Journal, Volume 50, (September, 1950), pp. 41-46.

long as they are not measuring identical functions it would be advisable to use them both. Also, if they correlated only to the point of .60 with each other, it is quite possible for one to be superior to the other for predictive purposes.

C. THE LEE-CLARK PRIMER AND FIRST READER TESTS

The two Lee-Clark Reading Tests are designed to measure the reading ability of first and second grade pupils and to aid in the analysis of achievement in silent reading skills. The reliability coefficients for the First Reader Test, calculated on the results of 139 pupils at the end of the second half of the first grade, was found to be .91 using the split-half technique and corrected by the Spearman-Brown formula. The coefficients of reliability for the Primer Test, calculated from the test results of 232 first grade pupils given at mid-year resulted in a coefficient of equivalence of .91 for Form A with Form B.[11]

Review of the test literature revealed only two studies in which the Lee-Clark Reading Tests were discussed. The California Test Bureau, publishers of the tests, explain this lack of information in a letter to the writer. It is

---

[11]Catalog of the California Test Bureau, op. cit., p. 47.

quoted in part.

> While we have not collected data systematically on the Lee-Clark Reading and Reading Readiness Tests, you will probably find the information you are looking for in the enclosed abstracts.[12]

One of the studies pertaining to the Lee-Clark Reading Tests was previously cited and discussed in connection with the literature on the Lee-Clark Reading Readiness Test. The second study, and the only one in which the Lee-Clark Reading Tests provided information of correlations with other reading achievement tests, was a doctoral dissertation by Doris E. Nason of Boston University. Her study attempted to determine whether or not pupils in grades one and two would achieve higher scores on those standardized reading tests which contained the greatest percent of words in common with basal readers. The basic readers used for the study were the Scott-Foresman series, the Row-Peterson series, and the Macmillan series.[13]

The conclusions of the study were:

> (1) It appears that the commonness of vocabulary between test and text has slight effect upon reading achievement as measured by standard tests. (2) The

---

[12]Letter from Wallace High of the California Test Bureau, Los Angeles, California, dated April 5, 1956.

[13]Doris Nason, "The Influence of Vocabulary Common to Test and Textbook on Primary Reading Scores," (unpublished Doctor's Dissertation, Boston University, Boston, Massachusetts, 1951). Taken from an abstract provided by the California Test Bureau.

> sensitivity to common test-text vocabulary does not diminish after Grade one. The evidence of the effect of vocabulary upon test scores is more clear-cut in Grade two than it is in Grade one. (3) Skill in phonics does lessen the effect of common test-text vocabulary upon test scores. (4) Pupils in the Scott-Foresman system in both grades one and two are the most effected by common test-text vocabulary, and pupils in the MacMillan system are least affected.[14]

In this study the Lee-Clark Primer Test correlated .45 with Mental Age on the Otis Quick Scoring Mental Ability Test and .70 with a special phonics test also given. Correlations of Mental Age and Phonics Test with other reading tests ranged from .42 to .49 for Mental Age and from .61 to .70 for the Phonics Test. The Lee-Clark First Reader test was correlated with Mental Age and the Phonics Test in Grade two and coefficients of .33 and .68, respectively, were found. Coefficients for other reading tests ranged from .24 to .40 with Mental Age and from .21 to .68 for the Phonics Test.[15]

The conclusion pertaining to the Scott-Foresman Readers is significant for this study since if any advantages are to accrue in terms of effecting standardized achievement test scores in reading that advantage appears to be in favor of students using the Scott-Foresman Readers.

---

[14] Ibid.

[15] Ibid.

D. THE GATES READING READINESS TEST

A review of the literature suggests that the Gates Reading Readiness Test is a sound instrument in all respects. Marion Monroe, in a written review in Oscar Buros' Third Mental Measurements Yearbook, lists the reliability coefficients of the separate sub-tests from .78 to .96 and the coefficient for the entire test as .97. Validity was measured by correlating the test with Gates Primary Reading Tests, Types I and II, and the correlation coefficients ranged from .57 to .89 for various schools with an average coefficient of .706.[16]

Several important studies by the author of the test and others have been made. The most important will be mentioned briefly here.

Gates conducted a series of studies in the late 1930's designed to appraise the predictive values of tests shown to be most useful as members of a battery for predicting reading progress during the first year of school. The study indicated that correlations yielded by similar reading tests are not as high where reading attainments are

---

[16]Marion Monroe, Review of the Gates Reading Readiness Test in the Third Mental Measurements Yearbook, ed. Oscar Buros (Highland Park, New Jersey: The Gryphon Press, 1949).

measured at the end of the first term as when measured at the end of a full year, probably because reading abilities are more stabilized and also more reliably measured by standardized tests at the end of the year than at the end of the term. Thus, had a full year transpired between the readiness and achievement testing, higher correlations would have been expected. Multiple correlations of the five separate sub-tests were computed with reading ability and, as reported previously in the Third Mental Measurements Yearbook, they ranged from .57 to .89 with a mean multiple r of .706. The figures indicate a high predictive value when it is known that the achievement tests were given at the end of the first semester.[17]

The sub-tests of the Gates tests were selected after a series of investigations extending over several years. In one extensive study, nearly one hundred different tests, examinations, ratings, etc. were given and tested to determine which were the most useful for diagnosing reading readiness. The most promising abilities shown by this and other studies were then embodied in tests of the types most likely to work well, and tried out on the entire population of children entering school in a typical small city in

---

[17]Arthur Gates, "A Further Evaluation of Reading Readiness Tests," Elementary School Journal, Volume 40, (April 1940), pp. 577-591.

which various methods of teaching were employed. After careful study of these results, a revised test was developed and tried out in another group of schools during the 1938-39 school year. The present test is based on all that could be learned from these and other studies of reading readiness.[18]

B. SUMMARY OF THE REVIEW OF THE LITERATURE

This chapter has attempted to describe the most important relative facts arising from past studies of the standardized instruments utilized in this study. A review of the literature has revealed information that is briefly summarized below.

Currently, there is inadequate statistical data regarding the reliability and validity of the New Scott-Foresman Basic Reading Tests. This is despite the statement of the publishers that they believe the Scott-Foresman tests are not comparable to other standardized reading achievement tests. Although there is an inference that the Scott-Foresman tests are somewhat analagous to the teacher-made tests this should not excuse the need for further

---

[18]Arthur Gates, Manual of Directions for the Gates Reading Readiness Tests (New York: Bureau of Publications, Teachers College, Columbia University, 1939), pp. 2-3.

information on reliability and validity in view of the fact that the tests are published and have been standardized.

There is a substantial amount of information on the reliability and validity of the Lee-Clark Reading Readiness Test, most of which is quite favorable to the test.

There is considerable information on the reliability and validity of the Gates Reading Readiness Test, most of which is also favorable to the test. However, the majority of the information compiled was gathered in studies made approximately twenty years ago.

There is substantial information on the reliability of the Lee-Clark Reading Tests but, admittedly by the test authors, no correlations are available with other reading achievement tests as a measure of future reading success.

CHAPTER III

RESULTS OF THE STUDY

Three of the four purposes of the study sought to establish interrelationships among two well-known reading readiness tests, the Scott-Foresman tests, and two popular reading achievement tests. These interrelationships were established by computing Pearson product-moment coefficients and are presented in this chapter. The relationships fall into three categories, identical with the three purposes of the study, and are presented in the order listed below:

(1) Interrelationships between the Lee-Clark Reading Readiness Test, the Gates Reading Readiness Test, and the New Scott-Foresman Basic Reading Tests.

(2) Interrelationships between the various individual tests of the New Scott-Foresman Series.

(3) Relationships between the Lee-Clark Reading Tests and the New Scott-Foresman Tests.

A. INTERRELATIONSHIPS BETWEEN THE LEE-CLARK READING READINESS TEST, THE GATES READING READINESS TEST AND THE NEW SCOTT-FORESMAN TESTS

One of the purposes of the study was to determine the effectiveness of the Lee-Clark Reading Readiness Test and

the Gates Reading Readiness Test as instruments for predicting future reading success by correlating the reading readiness test scores with Scott-Foresman test scores. In this study the Lee-Clark readiness scores of children in kindergarten during the 1954-55 school year were correlated with test scores of the same children given the Scott-Foresman Tests in 1955-56. The Gates readiness scores, given to kindergarten children during the 1953-54 school year, were correlated with scores of the same children, now second graders, given the Scott-Foresman Tests in the 1955-56 school year.

Past studies of the Lee-Clark Reading Readiness Test have indicated that it predicts reading success reasonably well as high as the sixth grade. Other studies have shown it to be a good predicting instrument for first grade reading success. These studies used other well-known standardized tests as instruments for the comparisons. The Scott-Foresman tests are different from the ordinary standardized tests in that they are administered to reading groups upon completion of the corresponding reader. In other words, they are given to students at different times determined by the rate of progress of various reading groups within individual rooms. When this procedure is necessitated by reading programs adapting to individual

differences in rate of progress, it is expected that the correlations would be attenuated.

The results of the relationships between the Lee-Clark Reading Readiness Test and the grade one Scott-Foresman tests are listed in Table I. Two of the three correlations with the individual reading tests were found to be insignificant at this grade level. The test for The New Fun With Dick and Jane correlated .38±.14 with the Lee-Clark test and is significant at the five percent level of confidence.

Two other statements of relationships were made using the same instruments. The first was a correlation of the average score of those children who had completed two or more of the Scott-Foresman tests with the Lee-Clark results. A correlation coefficient of .35±.13 was found, also significant at the five percent level. An average correlation coefficient, using Fisher's z, was insignificant.

The reason for the difference in outcome using these two approaches is attributable to their basic difference in meaning. The average score on the Scott-Foresman tests involves a process which, in effect, makes one long test out of the various ones in the series. It is analagous to putting together various teacher-made achievement tests

given during a semester and finding the child's average in order to arrive at a semester grade. The average r using Fisher's z, on the other hand, is a means of indicating the general trend in relationship for the various single tests in the series avoiding any combination of test performances.

A final correlation involved a comparison between the Lee-Clark test and the number of readers completed during the first grade. We could say that if teachers were to rate their first grade pupils as to success in reading at that level, a good indication would be given by the number of reading books that a child has completed. The better readers would complete more books than the slower readers since they are being allowed to advance at their own rate of speed. The correlation coefficient for this relationship was $.68 \pm .07$, significant at the one percent level. Thus, even though the relationships between the Lee-Clark test and the individual Scott-Foresman tests were largely insignificant, the Lee-Clark test is a good predictor of reading success in the first grade when it is correlated with the actual number of basic readers that each child has completed. The number of readers completed for each grade level would appear to be a good criteria of reading success since in schools assigning grades to children in reading, this is often a major determinant of the

TABLE I

RELATIONSHIPS BETWEEN THE LEE-CLARK READING READINESS TEST AND THE SCOTT-FORESMAN TESTS IN GRADE ONE, LINCOLN SCHOOL

Measure	Correlation Coefficient and Standard Error	Number of Cases
The New Pre-Primers	.25 ± .13	54
The New Fun With Dick and Jane	.38 ± .14	40
The New Our New Friends	.13 ± .20	19
Average r (using Fisher's z)	.25 ± .15	38
Scott-Foresman Average*	.35 ± .13	41
Number of Readers Completed	.69 ± .07	56

*Average Scores included only those children given two or more of the basic reading tests.

grade. On the other hand, the criteria of success is less justifiable, for correlational studies, where children are compared on a test covering a basic reader regardless of the time it has taken him to complete the reader. The latter is comparable to assigning an achievement grade to each child in terms of his ability to succeed.

Table II shows the relationships between the Gates Reading Readiness Test, given to 1955-56 second graders during their kindergarten year, and the Scott-Foresman tests. Two of the correlations with individual Scott-Foresman tests are insignificant. The correlation between the Gates test and the test for The New Friends and Neighbors is significant at the five percent level of confidence. Using Fisher's z, an average r for the three individual tests was found to be insignificant.

A somewhat surprising but favorable occurrence, in view of the above low correlations, is that the Scott-Foresman average test score, using the average score for those children who completed two or more of the tests, correlated .42±.15 with the Gates test scores and is significant at the one percent level.

A correlation coefficient of .44±.13 was found when the Gates test was compared with the number of second grade Scott-Foresman readers completed by second grade children.

TABLE II

RELATIONSHIPS BETWEEN THE GATES READING READINESS TEST AND THE SCOTT-FORESMAN TESTS IN GRADE TWO, LINCOLN SCHOOL

Measure	Correlation Coefficient and Standard Error	Number of Cases
The New Our New Friends	.32 ± .20	21
The New Friends and Neighbors	.33 ± .16	33
The New More Friends and Neighbors	.08 ± .23	17
Average r (using Fisher's z)	.25 ± .08	23
Scott-Foresman Average*	.42 ± .15	32
Number of Second Grade Readers Completed	.44 ± .13	37

*Average Scores included only those children given two or more of the basic reading tests.

This was significant at the one percent level and indicated that the Gates test was also a moderately good predictor of reading success when the rate of progress is used as the comparative standard.

It is seen that neither the Lee-Clark test nor the Gates test predicted reading success in the individual Scott-Foresman readers. Both tests predicted quite well, however, when they were correlated with the number of readers completed at each grade level and the latter approach appears to use a much more valid criterion of reading success.

B. INTERRELATIONSHIPS OF THE NEW SCOTT-FORESMAN BASIC READING TESTS

The second purpose of the study was to determine if any significant relationships existed between the various tests of the New Scott-Foresman series in the primary grades. Does the score of one test predict the score of the succeeding tests even though they are administered at varying rates of progression? The results of these interrelationships are indicated in Tables III through VII.

Table III summarizes the results of the interrelationships existing between the Scott-Foresman tests given to first graders at the Lincoln Elementary School.

Standard errors are presented with each correlation coefficient. The three coefficients show a significantly high relationship. The test for The New Pre-Primers correlated .78 ± .05 with the test for The New Fun With Dick and Jane and was significant at the one percent level of confidence. The test for The New Pre-Primers correlated .34±.16 with the test for The New Our New Friends but was still significant at the five percent level. The test for The New Fun With Dick and Jane correlated .54±.13 with the test for The New Our New Friends and was significant at the one percent level. Scores on the tests mentioned give a fairly accurate prediction of scores on succeeding tests. That is, children who score high on the first test, which apparently reflects how well they learn to read the first book, tend to score high on the next test. The converse is also obviously true. Those children scoring lower on the first test and apparently comprehending the book less well, tend to score low on the subsequent tests. This occurs despite the fact that some of the children scoring lower may take a much longer time studying on the one book and still tend to know the words and comprehend the ideas less well than some children studying the book a much shorter period of time.

TABLE III

INTERRELATIONSHIPS OF THE NEW SCOTT-FORESMAN BASIC READING TESTS IN GRADE ONE, LINCOLN SCHOOL

Measures	Correlation Coefficient and Standard Error	Number of Cases
The New Pre-Primers, 1_1 with The New Fun With Dick and Jane, 1_2	.78 ± .05	69
The New Pre-Primers, 1_1 with The New Our New Friends, 1_3	.34 ± .16	30
The New Fun With Dick and Jane, 1_2 with The New Our New Friends, 1_3	.54 ± .13	30

Table IV illustrates the interrelationships between the Scott-Foresman tests given to the second graders at the Lincoln School. Since many of the second graders are still not capable of reading from the second grade books some of the correlations involved include first grade books.

The three correlation coefficients shown are all significant at the one percent level, indicating a very substantial relationship between tests given to the second graders. The test for The New Fun With Dick and Jane correlated .57±.12 with the test for The New Our New Friends. The test for The New Our New Friends correlated .64±.09 with the test for The New Friends and Neighbors. The test for The New Friends and Neighbors correlated .64±.10 with the test for The New More Friends and Neighbors. The moderately high correlations suggest that children scoring low on the earlier tests tend to stay low and the relatively better readers still maintain the more favorable positions.

Table V is similar to Table IV in that it illustrates the results of the correlations of tests at the second grade but it includes scores of children given the same tests at the Washington School. The scores from the two schools were combined for the correlation because of the advantage of using an additional number of cases.

TABLE IV

INTERRELATIONSHIPS OF THE NEW SCOTT-FORESMAN BASIC READING TESTS FOR GRADE TWO, LINCOLN SCHOOL

Measure	Correlation Coefficient and Standard Error	Number of Cases
The New Fun With Dick and Jane, 1_2 with The New Our New Friends, 1_3	.57 ± .12	27
The New Our New Friends, 1_3 with The New Friends and Neighbors, 2_1	.64 ± .09	46
The New Friends and Neighbors, 2_1, with The New More Friends and Neighbors, 2_2	.64 ± .10	36

TABLE V

INTERRELATIONSHIPS BETWEEN THE NEW SCOTT-FORESMAN BASIC READING TESTS IN GRADE TWO, LINCOLN AND WASHINGTON SCHOOLS

Measures	Correlation Coefficient and Standard Error	Number of Cases
The New Fun With Dick and Jane, 1_2 with The New Our New Friends, 1_3	.76 ± .07	35
The New Our New Friends, 1_3 with The New Friends and Neighbors, 2_1	.77 ± .05	57
The New Friends and Neighbors, 2_1 with The New More Friends and Neighbors, 2_2	.72 ± .07	46

The coefficients show high interrelationships and all are significant at the one percent level. The scores for The New Fun With Dick and Jane correlated .76±.07 with scores for The New Our New Friends. Scores for the New Our New Friends correlated .77±.05 with scores for The New Friends and Neighbors and scores for The New Friends and Neighbors correlated .72 ± .07 with scores for The New More Friends and Neighbors. The coefficients are slightly higher in all cases than the coefficients for the same tests at the Lincoln School. This may be due to the sample becoming broader in range with the addition of more cases.

Table VI shows the interrelationships between the test for The New Fun With Dick and Jane and test for The New Our New Friends in grades one and two at the Lincoln Elementary School. The correlation coefficient here was .26±.05 and was significant at the five percent level. Two other correlations of the same tests were previously mentioned. One correlated the tests given to second graders at the Lincoln School and the other correlated the tests given to second graders at both Lincoln and Washington schools. The coefficients were .57±.12 and .76±.07 respectively. The coefficients using second graders alone are considerably higher than the coefficient for a combined grade one and two group. This may be due in part first,

TABLE VI

INTERRELATIONSHIPS BETWEEN TWO OF THE NEW SCOTT-FORESMAN BASIC READING TESTS USED IN GRADES ONE AND TWO, LINCOLN SCHOOL

Measures	Correlation Coefficient and Standard Error	Number of Cases
The New Fun With Dick and Jane, 1_2 with The New Our New Friends, 1_3	.26 ± .05	56

to tests being taken at a relatively closer time interval when in one given grade or second, to greater changes occurring between two school grades than within one school year. Maturation might also be a factor that has influenced the differences. Children might possibly make more relative changes in reading growth in the first grade than other grades.

The interrelationships between the Scott-Foresman tests at the third grade level are shown in Table VII. The number of cases in each school, by itself, was insufficient to base any conclusions on the resulting coefficient. The correlation coefficient between the test for The New Streets and Roads and the test for The New More Streets and Roads was .56±.23 at the Lincoln School, significant at the five percent level of confidence, and .78±.10 at the Washington School, significant at the one percent level of confidence. The coefficient for the Lincoln and Washington schools combined was .73±.09 and was significant at the one percent level. Using Fisher's z, the average of the three above correlations was .71±.12, significant at the one percent level. Although the number of cases in each of the correlations was small, it was sufficient to give some idea of the relationship between the two tests at the third grade level.

TABLE VII

INTERRELATIONSHIPS BETWEEN THE NEW SCOTT-FORESMAN TESTS, THE NEW STREETS AND ROADS AND THE NEW MORE STREETS AND ROADS, GRADE THREE, LINCOLN AND WASHINGTON SCHOOLS

School	Correlation Coefficient and Standard Error	Number of Cases
Lincoln School	.56 ± .23	10
Washington School	.78 ± .10	16
Lincoln and Washington Schools Combined	.73 ± .09	26
Average r (using Fisher's z)	.71 ± .12	17

C. RELATIONSHIPS BETWEEN THE LEE-CLARK READING TESTS AND THE NEW SCOTT-FORESMAN BASIC READING TESTS

The third purpose of the study was to determine the existing relationships between two reading achievement tests, administered in the Lincoln School during the year previous to the Scott-Foresman tests, and the individual tests within the Scott-Foresman series. The Lee-Clark Reading Test: Primer was given to 1955-56 second grade children the year prior to the Scott-Foresman tests and the Lee-Clark Reading Test: First Reader was given to the 1955-56 third grade children the year prior to the study. Neither of the Lee-Clark Reading Tests had apparently been validated against outside criteria. Since the Scott-Foresman tests have also not been validated in terms of empirical studies with other tests, the correlations presented here are simply to show the relationship between reading achievement tests at this grade level. Each purports to measure reading achievement. Consequently, if correlations show them to be measuring the same factors to some considerable extent it would be reasonable to suggest that the validity of each might be enhanced.

These correlations are also an attempt to determine the effectiveness of the Lee-Clark Reading Tests as

predictors of future reading success by correlating them with Scott-Foresman test scores. The reader is reminded again that the Scott-Foresman tests were administered on the basis of rate of group progress and that children were given a particular test upon completion of the accompanying reading book. In this way they are somewhat different than the usual standardized reading achievement test.

The results of the correlations are presented in Table VIII. The majority of the coefficients are moderately high to high and all but two are significant at the one percent level. The coefficient of correlation for The New Fun With Dick and Jane and the Lee-Clark Primer Test shows almost no relationship. This may be due to a chance relationship from the fact that the number of cases was low but more probably to the fact that this test was administered to low reading groups of the second grade which in previous correlations have not been as consistent as the middle and upper groups.

The coefficient of correlation for the Lee-Clark Primer Test and The New Our New Friends is .54±.10 and significant at the one percent level. The coefficient of correlation for the Lee-Clark Primer Test and the test for The New Friends and Neighbors is .91±.02 and significant at the one percent level. This is one of the highest

TABLE VIII

RELATIONSHIPS BETWEEN THE LEE-CLARK READING TEST: PRIMER AND THE NEW SCOTT-FORESMAN BASIC READING TESTS AT LINCOLN SCHOOL

Measure	Correlation Coefficient and Standard Error	Number of Cases
The New Fun With Dick and Jane	.02 ± .23	23
The New Our New Friends	.54 ± .10	53
The New Friends and Neighbors	.91 ± .02	64
The New More Friends and Neighbors	.34 ± .17	30
Average r (using Fisher's z)	.56 ± .10	42
Scott-Foresman average*	.64 ± .07	73

*Average scores included only those children given two or more of the basic reading tests.

coefficients obtained in the entire study and indicates a very high degree of relationship between the two tests.

The coefficient of correlation for the Primer Test and the test for The New More Friends and Neighbors is .34±.17 and significant at the five percent level. This may be surprising in view of the previous correlation of .91 but it may be partly due to the fact that the thirty cases represented the high reading groups which would restrict the range in one dimension and tend to attenuate the correlation.

An average coefficient was computed using Fisher's z and found to be .56±.10 which is significant at the one percent level. An average Scott-Foresman score was also correlated with the Lee-Clark test and a coefficient of .64±.07 was found. This is also significant at the one percent level.

In general the Lee-Clark Primer Test seems to be a fairly good indicator of future reading success as measured by the Scott-Foresman tests. Although the coefficients fluctuate from .02 to .91, the majority of them are quite significant. There is considerable doubt as to the predictive ability of the Lee-Clark Test as measured against the test for The New Fun With Dick and Jane when given to second graders. It would seem that this is more likely to be due

to possible low reliability of any single Scott-Foresman test since reliabilities that are quite adequate are reported for the Lee-Clark Test but are not reported for the Scott-Foresman tests. Also, the number of children taking The New Fun With Dick and Jane in the second grade is small and the ability range is restricted which would attenuate the relationship.

The Lee-Clark First Reader Test was compared with the Scott-Foresman third grade tests as a final phase of the relationships. The number of cases involved in the correlations was quite small and any interpretation of the relationships should be made with that in mind. The Lee-Clark First Reader Test was administered to second grade pupils in 1954-55 and who were third graders during the time of the study.

The coefficient of correlation for the test for The New Streets and Roads and the Lee-Clark First Reader Test is .50±.13 and is significant at the one percent level. This is a substantial relationship. The coefficient of correlation for the test for The New More Streets and Roads and the Lee-Clark test is .44±.18, also significant at the one percent level. Using Fisher's z, an average correlation was also computed and found to be .47 and significant at the one percent level. It appears that the Lee-Clark

TABLE IX

RELATIONSHIPS BETWEEN THE LEE-CLARK READING TEST: FIRST READER AND THE NEW SCOTT-FORESMAN BASIC READING TESTS IN GRADE THREE, LINCOLN SCHOOL

Measure	Correlation Coefficient and Standard Error	Number of Cases
The New Streets and Roads	.50 ± .13	32
The New More Streets and Roads	.44 ± .18	20
Average r (using Fisher's z)	.47 ± .15	26

First Reader Test is a good predictor of reading achievement as measured by third grade Scott-Foresman tests.

In summary, this chapter reported relationships between two reading reading tests and the Scott-Foresman tests for first and second grade children in Lincoln School; it presented interrelationships among many of the tests within the Scott-Foresman test series; it determined relationships between two other standardized reading achievement tests and some of the Scott-Foresman tests.

The relationships found between the Lee-Clark Reading Readiness Test and the Scott-Foresman tests given to first grade children were largely insignificant. However, when the number of readers completed by first grade children was correlated with the Lee-Clark Readiness Test scores a high relationship was established. This latter approach would appear to be a better criterion of reading success than comparisons with test results covering varying time intervals.

Correlations of the Gates Reading Readiness Test and the Scott-Foresman tests given to second graders also resulted in insignificant or low relationships. Again, however, the Gates test predicted success reasonably well when correlated with the number of second grade readers completed by children of the second grade. The results of the

correlations of the two reading readiness tests and the individual Scott-Foresman tests should in no way reflect on the validity of the reading readiness instruments since they were being compared with instruments on which additional information as to reliability and validity was desired.

The interrelationships reported between the various tests within the Scott-Foresman series are all moderately high or high indicating that children who score high on one test are likely to score high on succeeding tests and children scoring low will continue to score relatively low. This occurs despite the fact that some children may spend considerably longer in a book than other children.

The relationships established between the Lee-Clark Reading Tests and the Scott-Foresman tests were also, in general, moderately high to high. This would indicate that the Lee-Clark Reading Tests are fairly good predictors of reading success when compared to the Scott-Foresman tests.

CHAPTER IV

IMPLICATIONS OF THE STUDY

Of major importance is how the information that has been gathered will be interpreted by teachers and administrators using the standardized tests in question and what implications may exist for their continued usage. The information presented in the previous chapter appears significant to educators in many respects. It also has presented many separate but interrelated questions and areas for further thought and consideration.

First, the Lee-Clark Reading Readiness Test scores do not predict very well how children will perform on single Scott-Foresman reading achievement tests given to first graders. This should not reflect on the validity of the Lee-Clark test since the comparisons are being made with an instrument that is itself of questionable validity. The low correlation may be due to questionable reliability of the Scott-Foresman tests and to the fact that children took the achievement tests at varying intervals of time. A child spending the entire year or major portion of a school year on one reader may come to learn the material well enough to score quite high on that one test. Meanwhile, another child covering three readers in the same period of

time and having taken the first test much earlier would not be expected to score significantly higher on that one test.

The Gates Reading Readiness Test, when correlated with second grade Scott-Foresman test scores, revealed the same magnitude of relationships as the Lee-Clark test did with first grade Scott-Foresman scores. They do not appear to be a good predictor of test scores for that particular grade level. Again, this should not necessarily reflect on the validity of the Gates test.

Second, the Lee-Clark Reading Tests: Primer and First Reader appear to be good predictors of subsequent reading success when measured with the Scott-Foresman tests given to the same children a year later. Teachers and administrators may place confidence in the fact that the Lee-Clark Reading Tests provide effective evaluation of reading progress at least in terms of indicating their future reading success.

Third, a reliability (stability) study of the Scott-Foresman tests would allow one to place greater confidence in decisions resulting from analysis of the results. For example, one could then be more confident in concluding that a child scoring high on a given test really knows the vocabulary and comprehends the ideas presented in the accompanying book well enough to proceed to the next. A stability

study could be done by administering each test a second time to each child one to two weeks after the child took it the first time. If the resulting stability coefficients were high, much more confidence could be placed in the test results. A reliability (equivalence) study could only be made if the publishers produced a second form of each test. Such a step would have merit insofar as providing teachers with an additional teaching and evaluation aid.

Fourth, the results of the interrelationships of the Scott-Foresman tests indicate in most instances that any one of the individual tests is a fairly good predictor of scores in succeeding tests. Teachers may place confidence in a child's test score and assume that if he is in one position at a particular time, he probably will hold a similar position in future reading achievement. This would be true of those teachers devoting approximately the same amount of time to each of the different reading groups within a single room but may not be as true of situations in which more help and time is accorded to the slower readers and retarded groups.

Fifth, the children's scores on the Scott-Foresman tests may not be completely indicative of whether children are ready to advance into the next reader. Reliability studies would allow teachers to place greater confidence in

the results. With confidence in the results, when a child scores low he is probably not ready to advance to the next reader. If they are allowed to go on, the low scores on subsequent tests may be due, in part, to inadequate background before undertaking the next book. Such children may need additional learning on parallel level books. If improvement on teaching and evaluation materials could be made available as suggested above, it would seem that the Scott-Foresman tests have a distinct advantage over other types of standardized reading achievement tests in this respect.

Sixth, a previous study indicated that the commonness of test-text vocabulary seemed to have no influence on reading achievement test results. However, it was mentioned that if any advantage did exist, among the three text series included, it would lie with the Scott-Foresman basic reading series. It is believed that a worthwhile study for the future would be a further test of this hypothesis. If definite advantages do lie in the use of the Scott-Foresman readers and the advantages are reflected in higher test scores on achievement tests other than the Scott-Foresman tests teachers and administrators should understand this and take it into account in their evaluations of the child's reading success.

Seventh, it appears that a wide range of possible future studies has been created with adoption by school systems of tests such as the Scott-Foresman Company publishes. Such studies would seem to be an important part of a school's evaluative procedures if they are depending on such test results as a basis for grouping, future reading success, additional work at a parallel level, or grading. The following questions are raised as examples: (1) How do various reading groups - high, middle, and low - compare with each other on test results of the same instruments? Does the time spent on a reader influence the test scores to any extent? How does a composite Scott-Foresman score predict reading success as measured by other standardized instruments? How do intelligence test scores predict Scott-Foresman test results, individually and totally? How do the reliabilities of the various sub-tests compare with each other?

Finally, to properly interpret the test results of instruments such as the Scott-Foresman tests requires a knowledge of the accuracy of test scores. Every user of test scores knows that no test is perfectly accurate and that a score is affected by the inaccuracy of the test itself. There is no way to determine the precise amount of error in an individual case but the use of the standard

error of measurement is one of the most useful.[1] It would seem practical, in schools that use the Scott-Foresman tests or similar instruments, for teachers to have a thorough understanding of the accuracy of test scores since they must be the judge of whether children are to be allowed to pursue an advanced reader or whether they should be given additional work at a parallel level. It would appear that such knowledge could eliminate some of the doubts in teachers' minds as to the progress of certain children. Many people will agree that if tests are to be used as an evaluation technique they must be used properly and with all the skill that is at hand.

Continuous studies on the evaluative instruments used by schools is necessary to effectively assess progress toward school objectives in reading or other areas.

---

[1] "How Accurate is a Test Score?", Test Service Bulletin (New York: The Psychological Corporation, Number 50, June, 1956), p. 1.

CHAPTER V

SUMMARY AND CONCLUSIONS

A variety of instruments and techniques are employed in measuring school achievement. The use of standardized tests, however, is usually recommended as the most objective and best single means of measurement. Tests in reading have been most numerous and although most are good, some are inadequately standardized and lack information on reliability and validity. Among the tests on which additional data is desired are the New Scott-Foresman Basic Reading Tests and the Lee-Clark Reading Tests.

There were four purposes for the study:

1. To determine the interrelationships existing between the primary grade tests within the New Scott-Foresman Basic Reading Test series.

2. To determine the relative efficiency of two reading readiness tests, the Lee-Clark test and the Gates test, in predicting children's later success in reading.

3. To determine the relative efficiency of the Lee-Clark Reading Tests in predicting subsequent success in reading.

4. To establish norms for the New Scott-Foresman Basic Reading Tests that may be used in the community in

which the study was conducted.

The study is important because it attempts to establish additional validational data on several standardized reading tests, to provide information that may assist teachers in predicting children's future success in reading, and to establish local norms for the Scott-Foresman tests which will enable teachers using them to better understand the relative reading accomplishments of children under their guidance.

The Scott-Foresman tests are a series of tests designed to measure reading achievement of pupils using the Scott-Foresman readers. The tests are to be administered upon completion of the accompanying Scott-Foresman reading book. In this respect, they are different than the ordinary standardized achievement tests ordinarily recommended for use in a fall-spring testing program. The Scott-Foresman tests may be regarded as somewhat similar to teacher-made objective tests only in a more standardized way since they are published and have norms established.

The Lee-Clark Reading Readiness Test and the Gates Reading Readiness Test are well-established instruments in the readiness field. Both have information published as to reliability and both have been correlated with other reading achievement tests as measures of prediction of future

reading success.

The Lee-Clark Reading Tests: Primer and First Reader are two reading tests designed to test the reading achievement of first and second grade children. Reliabilities have been published for the tests but additional information is desired on validational studies.

Information for the study was gathered in the primary grades of two public elementary schools in the community of Ellensburg, Washington. Teachers in these schools largely use the group method of teaching reading. Three reading groups - high, medium, and low - are usually maintained in each room.

Relationships between the standardized reading tests were established by use of the Pearson product-moment correlation coefficient. A standard error for each correlation was also computed. A total of thirty correlations was made. The Lee-Clark Reading Readiness Test had been administered to children one year previous to the time of the study, the Gates Reading Readiness Test had been given to children two years previous to the time of the study, and the Lee-Clark Reading Tests had been given to children one year previous to the time of administration of the Scott-Foresman tests.

A review of the literature revealed that the Gates

Reading Readiness Test and the Lee-Clark Reading Readiness Test have proven to be good predictors of future reading success. The Lee-Clark readiness test predicts reading success quite well as high as the sixth grade. The Lee-Clark Reading Tests had little information relative to validity studies but one study in which they were used contained information relative to the Scott-Foresman readers. In that study it was revealed that if any advantages exist in using basal readers that contain the greatest percentage of words in common with standardized reading tests, the advantage appears to be in favor of using the Scott-Foresman readers.

The Scott-Foresman tests lack information both as to reliability and validity but this is due to the philosophy of the test author regarding the use of her tests. It is believed, however, that since the tests are being used to measure reading achievement the necessary information as to reliability and validity should be published in order to make them a more effective instrument. It was on this premise that the writer undertook the study.

The results of the correlations and accompanying conclusions are as follows:

1. Scores of the Lee-Clark Reading Readiness Test and the Gates Reading Readiness Test do not predict reading

success for first and second graders when measured by scores of the Scott-Foresman tests. The readiness scores do predict reading success, however, when compared to the number of basic readers, designated for a given grade level, that are completed during the year. This is believed to be a more valid criteria of reading success than the Scott-Foresman scores.

2. Scores on any of the individual Scott-Foresman reading tests appear to be good predictors of scores on future Scott-Foresman tests at any grade level. The correlations for tests given to first graders ranged from .34±.16 to .78±.05; for second graders at the Lincoln School only, from .57±.12 to .64±.10; for second graders at the Lincoln and Washington Schools combined, .72±.07 to .77±.05; for third graders at Lincoln and Washington Schools combined, from .56±.23 to .78±.10.

3. Scores of the Lee-Clark Reading Tests are fairly good predictors of future reading success when measured by the Scott-Foresman tests. Correlations ranged from .02±.23 to .91±.02 for the Lee-Clark Primer Test with an average correlation, using Fisher's z, of .56±.10. Correlations ranged from .44±.18 to .50±.13 for the Lee-Clark First Reader Test, with an average correlation of .47±.15.

Several interpretations and implications resulted from the study.

1. Although the readiness tests did not predict how well children performed on single Scott-Foresman tests this should not reflect on the validity of the readiness tests since the comparisons were being made with an instrument that is itself of questionable validity. Also, the low correlations may have been influenced from the fact that children took the tests at varying intervals of time during the year.

2. Teachers may place reasonable confidence in a child's Scott-Foresman test score and assume that if he is in one position at a particular time, he probably will hold a similar position in future reading achievement.

3. A future reliability (stability) study of the Scott-Foresman tests would allow one to place greater confidence in decisions resulting from analysis of the results.

4. Children's scores on the Scott-Foresman tests may not be completely indicative of whether children are ready to advance into the next reader. Reliability studies would allow teachers to place greater confidence in the results.

5. Teachers using the Scott-Foresman tests for placement of children in reading groups need to understand how to properly interpret the test scores for the most

effective results in their reading program. It is believed that an orientation in the use and interpretation of the standard error of measurement might greatly enhance a school's reading program in which these instruments are used.

6. It appears that a wide range of possible future studies has been created with adoption by schools of tests such as published by the Scott-Foresman Company. Some of the questions that are raised are: (1) How does the time spent on a reader influence the test scores, if at all? (2) How do various reading groups compare with each other on test results? (3) How do intelligence test scores correlate with Scott-Foresman test results?

Continuous studies on the evaluative instruments used by schools are necessary in order to effectively assess progress toward school objectives in reading or other areas.

BIBLIOGRAPHY

BIBLIOGRAPHY

A. BOOKS

Anastasi, Anne. Psychological Testing. New York: The Macmillan Company, 1954.

Buros, Oscar (ed.). The Third Mental Measurements Yearbook. New Brunswick: Rutgers University Press, 1949.

Cook, Walter. "Achievement Tests," Encyclopedia of Educational Research. Edited by Walter S. Monroe. Revised edition. New York: The Macmillan Company, 1950.

Cronbach, Lee J. Essentials of Psychological Testing. New York: Harper and Brothers Publishers, 1949.

Edwards, Allen L. Statistical Methods for the Behavoral Sciences. New York: Rinehart and Company, 1954.

Freeman, Frank S. Theory and Practice of Psychological Testing. Revised edition. New York: Henry Holt and Company, 1955.

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, 1950.

Jordan, A. M. Measurement in Education. New York: McGraw-Hill Book Company, 1953.

Lacey, Oliver L. Statistical Methods in Experimentation. New York: The Macmillan Company, 1953.

Witty, Paul and David Kopel. Reading and the Educative Process. Boston: Ginn and Company, 1939.

B. PERIODICALS

Gates, Arthur. "A Further Evaluation of Reading Readiness Tests," Elementary School Journal, Volume 40 (April, 1940), pp. 577-591.

_____. "An Experimental Evaluation of Reading Readiness Tests," Elementary School Journal, Volume 39 (March, 1939), pp. 497-508.

Henig, Max S. "Predictive Value of a Reading Readiness Test and of Teacher's Forecasts," Elementary School Journal, Volume 50 (September, 1949-50), pp. 41-46.

Lee, J. Murray, Willis W. Clark, and Dorris M. Lee. "Measuring Reading Readiness," Elementary School Journal, Volume 34 (May, 1934), pp. 656-666.

Moreau, Margaret. "Long Term Prediction of Elementary Reading Achievement," California Journal of Educational Research, Volume 1 (September, 1950), pp. 173-176.

Wilson, Frank T. "Early Achievement in Reading," Elementary School Journal, Volume 42. (April, 1942), pp. 609-615.

C. PAMPHLETS

Eight More Years of Research in Reading. Educational Records Bulletin Number 64. New York: Educational Records Bureau, January, 1955.

Gates, Arthur. Manual of Directions for the Gates Reading Readiness Tests. New York: Bureau of Publications, Columbia University, Teachers College, 1939.

Monroe, Marion. General Manual for The New Scott-Foresman Basic Reading Tests. Chicago: Scott-Foresman Company, 1954.

Test Service Bulletin. Number 50. New York: The Psychological Corporation, June, 1956.

C. UNPUBLISHED MATERIALS

Nason, Doris E. "The Influence of Vocabulary Common to Test and Textbook on Primary Reading Scores," Unpublished Doctor's Dissertation, Boston University, Boston, 1951.

Robinson, Agnes S. "A Study of the Relative Values of Two Standardized Tests in Predicting First-Grade Achievement in Reading," Unpublished Master's Thesis, Sacramento State College, Sacramento, 1952.

E. CATALOG

Catalog of Standardized Tests. Los Angeles: California Test Bureau, 1955.

F. LETTERS

Letter from Wallace High, Research Associate of California Test Bureau, Los Angeles, dated April 5, 1956.

Letter from Mary Steuteville, Scott-Foresman Company, Chicago, dated February 24, 1956.

APPENDIX A

FORMULA USED FOR COMPUTATION OF THE PEARSON PRODUCT-MOMENT COEFFICIENT

FORMULA USED FOR COMPUTATION OF THE STANDARD ERROR OF **r**

SAMPLE CORRELATION COMPUTATION

FORMULA USED FOR COMPUTATION OF THE PEARSON PRODUCT-MOMENT COEFFICIENT*

$$r_{xy} = \frac{\Sigma xy}{(\Sigma x^2)(\Sigma y^2)}$$

where $$\Sigma xy = \left[(\Sigma x'y') - \frac{(\Sigma x)(\Sigma y)}{N}\right] i_x i_y$$

and $$\Sigma x^2 = \left[(\Sigma fx'^2 - \frac{(\Sigma fx')^2}{N}\right] i_x^2$$

and $$\Sigma y^2 = \left[(\Sigma fy'^2 - \frac{(\Sigma fy')^2}{N}\right] i_y^2$$

FORMULA FOR THE COMPUTATION OF THE STANDARD ERROR OF r

$$S.E._r = \frac{1 - r^2}{\sqrt{N-1}}$$

---

*Allen L. Edwards, Statistical Methods for the Behavoral Sciences (New York: Rinehart and Company, 1954), pp. 145-155.

SAMPLE CORRELATION COMPUTATION USING THE LEE-CLARK PRIMER TEST AND THE SCOTT-FORESMAN TEST "THE NEW FRIENDS AND NEIGHBORS." (GRADE TWO, LINCOLN SCHOOL)

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}}$$

$$\Sigma xy = \left[\Sigma x'y' - \frac{(\Sigma x)(\Sigma y)}{N}\right] i_x i_y = \left[3517 - \frac{(613)(349)}{64}\right] 3 \cdot 3$$

$$= [3517 - 3342.76]9 = 274.24 \cdot 9 = 2468.16$$

$$\Sigma x^2 = \left[(\Sigma fx'^2 - \frac{(\Sigma fx')^2}{N}\right] i_x^2 = \left[6265 - \frac{(613)^2}{64}\right] 3^2$$

$$= \left[6265 - \frac{375769}{64}\right]9 = [6265 - 5871.39]9$$

$$= 393.61 \cdot 9 = 3542.49$$

$$\Sigma y^2 = \left[\Sigma fy'^2 - \frac{(\Sigma fy')^2}{N}\right] i_y^2 = \left[2129 - \frac{(349)^2}{64}\right] 3^2$$

$$= \left[2129 - \frac{121801}{64}\right]9 = [2129 - 1903.14]9 = 2032.74$$

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} = \frac{2468.16}{\sqrt{(3542.5)(2032.7)}} = \frac{2468.16}{\sqrt{7200961.12}}$$

$$= \frac{2468.16}{2685.2} = .9117 = .91$$

$$S.E._r = \frac{1 - r^2}{\sqrt{N-1}} = \frac{1 - (.91)^2}{\sqrt{64-1}} = \frac{1 - .8312}{7.95} = \frac{.169}{7.95} = .021$$

$$r = .91 \pm .02$$

APPENDIX B

DISTRIBUTION OF SCORES ON THE NEW SCOTT-FORESMAN BASIC READING TESTS IN LINCOLN SCHOOL

DISTRIBUTION OF SCORES FOR THE NEW SCOTT-FORESMAN BASIC READING TEST: THE NEW PRE-PRIMERS GIVEN TO FIRST GRADERS AT LINCOLN SCHOOL

Score	f	Centile*
69-70	5	99
67-68	16	95
65-66	26	77
63-64	18	49
61-62	12	30
59-60	9	17
57-58	3	7
55-56	1	4
53-54	2	3
51-52	0	1
49-50	1	1

N = 93

M = 63.78

S.D. = 3.75

*These centiles refer only to the upper limit of the class intervals on the raw scores.

DISTRIBUTION OF SCORES FOR THE NEW SCOTT-FORESMAN BASIC READING TEST: THE NEW FUN WITH DICK AND JANE GIVEN TO FIRST GRADERS AT LINCOLN SCHOOL

Score	f	Centile*
69-70	8	99
67-68	14	88
65-66	11	69
63-64	7	54
61-62	10	44
59-60	7	29
57-58	3	19
55-56	4	14
53-54	4	10
51-52	2	4
49-50	1	1

N = 71

M = 62.8

S.D. = 5.2

*These centiles refer only to the upper limit of the class interval on the raw scores.

DISTRIBUTION OF SCORES FOR THE NEW SCOTT-FORESMAN BASIC READING TEST: THE NEW OUR NEW FRIENDS GIVEN TO FIRST GRADERS AT LINCOLN SCHOOL

Score	f	Centile*
69-70	10	99
67-68	3	67
65-66	3	57
63-64	4	47
61-62	1	33
59-60	6	30
57-58	0	10
55-56	1	10
53-54	0	7
51-52	2	7

N = 30

M = 63.5

S.D. = 6.82

*These centiles refer only to the upper limit of the class interval on the raw scores.

DISTRIBUTION OF SCORES FOR THE NEW SCOTT-FORESMAN BASIC READING TEST: THE NEW FUN WITH DICK AND JANE GIVEN TO SECOND GRADERS AT LINCOLN SCHOOL

Score	f	Centile*
67-68	3	99
65-66	1	89
63-64	2	86
61-62	5	78
59-60	0	61
57-58	0	61
55-56	1	61
53-54	5	57
51-52	3	39
49-50	2	29
47-48	1	21
45-46	0	18
43-44	2	18
41-42	0	11
39-40	1	11
37-38	0	7
35-36	1	7
33-34	0	4
31-32	1	4

N = 23

M = 54.3

S.D. = 9.66

*These centiles refer only to the upper limit of the class interval on the raw scores.

DISTRIBUTION OF SCORES FOR THE NEW SCOTT-FORESMAN BASIC READING TEST: THE NEW OUR NEW FRIENDS GIVEN TO SECOND GRADERS AT LINCOLN SCHOOL

Score	f	Centile*
67-68	5	99
65-66	5	93
63-64	6	86
61-62	9	77
59-60	10	64
57-58	2	50
55-56	7	47
53-54	5	37
51-52	0	30
49-50	2	30
47-48	4	27
45-46	2	21
43-44	1	18
41-42	1	17
39-40	4	16
37-38	1	10
35-36	1	9
33-34	0	7
31-32	0	7
29-30	2	7
27-28	1	5
25-26	1	3
23-24	0	1
21-22	1	1

N = 70

M = 54.5

S.D. = 10.5

*These centiles refer only to the upper limit of the class interval on the raw scores.

DISTRIBUTION OF SCORES FOR THE NEW SCOTT-FORESMAN BASIC READING TEST: THE NEW FRIENDS AND NEIGHBORS GIVEN TO SECOND GRADERS AT LINCOLN SCHOOL

Score	f	Centile*
69-70	10	99
67-68	10	88
65-66	11	77
63-64	15	63
61-62	12	46
59-60	5	32
57-58	4	27
55-56	2	22
53-54	2	20
51-52	2	17
49-50	3	15
47-48	1	12
45-46	0	10
43-44	0	10
41-42	1	10
39-40	1	9
37-38	5	8
35-36	0	2
33-34	0	2
31-32	2	2

N = 86

M = 59.9

S.D. = 9.52

*These centiles refer only to the upper limits of the class intervals on the raw score.

DISTRIBUTION OF SCORES FOR THE NEW SCOTT-FORESMAN BASIC READING TEST: THE NEW MORE FRIENDS AND NEIGHBORS GIVEN TO SECOND GRADERS AT LINCOLN SCHOOL

Score	f	Centile*
69-70	3	99
67-68	4	91
65-66	5	81
63-64	3	67
61-62	9	58
59-60	7	33
57-58	3	14
55-56	0	5
53-54	2	5
	N = 36	
	M = 62.38	
	S.D. = 4.15	

*These centiles refer only to the upper limits of the class intervals on the raw scores.

DISTRIBUTION OF SCORES FOR THE NEW SCOTT-FORESMAN BASIC READING TEST: THE NEW STREETS AND ROADS GIVEN TO THIRD GRADERS AT LINCOLN SCHOOL

Score	f	Centile*
67-68	2	99
65-66	5	96
63-64	7	86
61-62	5	72
59-60	9	63
57-58	6	45
55-56	3	33
53-54	4	27
51-52	1	19
49-50	1	17
47-48	2	16
45-46	2	12
43-44	3	7
41-42	0	2
39-40	1	2

N = 51

M = 57.5

S.D. = 7.15

*These centiles refer only to the upper limit of the class interval on the raw scores.

DISTRIBUTION OF SCORES FOR THE NEW SCOTT-FORESMAN BASIC READING TEST: THE NEW MORE STREETS AND ROADS GIVEN TO THIRD GRADERS AT LINCOLN SCHOOL

Score	f	Centile*
69-70	2	99
67-68	2	94
65-66	6	88
63-64	1	69
61-62	4	67
59-60	5	55
57-58	4	39
55-56	3	27
53-54	3	18
51-52	1	9
49-50	1	6
47-48	1	3

N = 33

M = 59.98

S.D. = 5.72

*These centiles refer only to the upper limit of the class interval on the raw scores.