1968

A Comparative Study of Reading Attainment Between Matched Groups of Graded and Non-Graded Students

Joseph Raymond Delaney
Eastern Illinois University

PAPER CERTIFICATE #3

To: Graduate Degree Candidates who have written formal theses.

Subject: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements.

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

Aug. 8, 1968
Date Author

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because _____

_____

_____

_____ _____
Date Author

A Comparative Study of Reading Attainment Between Matched Groups of Graded and Non-Graded Students

(TITLE)

BY

Joseph Raymond Delaney

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Master of Science in Education Degree

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

1968
YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

8-8-68
DATE

8-8-68
DATE

DEPARTMENT HEAD

ACKNOWLEDGEMENTS

The writer wishes to express his gratitude and sincere appreciation to the following persons: Dr. William J. Crane, advisor, for his patience and guidance in the preparation of this paper; Dr. Paul D. Overton for his excellent counsel and inspiration; Dr. Donald Moler for his participation on the committee; Mr. Joseph E. Foster, Superintendent of Livingston County Community Unit #2, in allowing this study to be made by permitting the use of school records. The writer also wishes to express his gratitude and thanks to his wife, Joanna, whose willingness to type and patience to understand contributed to the completion of this study.

TABLE OF CONTENTS

LIST OF CHARTS

Chapter I

INTRODUCTION AND STATEMENT OF THE PROBLEM

Preliminary Statement

When human beings first recognized differences among themselves in mental abilities and other significant traits is unknown, but the scientific recognition of such differences and, later, their scientific measurement came about during the nineteenth century. These discoveries and the derivation of statistical methods needed for the summarization and analysis of these differences were necessary foundations for subsequent developments in measurement and evaluation.[1] During this century testing has reached such tremendous proportions that a new breed of individual has emerged on the scene. This is the testing specialist who has found a fertile paradise in the present era.

Some form of evaluation is necessary to determine whether or not progress, either backwards or forwards, is being made. This is the reason for the increased interest in measurement in our school systems today. Huge expenditures are being made by school districts and, to determine whether the program justifies the price, evaluations are necessary. Most evaluations involving measurement are useful due to the

---

[1] J. Raymond Gerberich, Harry A. Greene and Albert N. Jorgensen, Measurement and Evaluation in the Modern School (New York: David McKay Company, Inc., 1962) p. 21.

fact that a definite aim is sought and is stated in such a way as to be meaningful. By the proper use of instruments for measuring, it is possible for the individual to know what he has accomplished. They also aid in helping school officials determine or discover when emphasis has been misplaced.

Actually, then, the recent development of modern educational instruments of measurement and evaluation may be regarded as an extension and improvement of an old practice. The modern education measuring instrument presents a picture of the course objectives as well as an analysis of the underlying skills, knowledges, concepts, understandings, and other outcomes upon which accomplishment in different subject areas depends. Educational tests and the information resulting from their use in the classroom have come to be almost universally identified with good teaching practice.[2] It is with this point in mind that an evaluation of reading progress in this school system is in order.

Philosophy of the Non-Graded Approach

The basic concept behind the non-graded or un-graded school is that all individuals are different and consequently education should recognize these differences and organize accordingly. "The wide range of differences among students of the same chronological age and the differences in understanding and achievement from subject to subject for a single student do not lend themselves to easy compression into the lock step of grade levels."[3] Children entering the first grade of school

---

[2] Ibid., p. 5.

[3] John I. Goodlad, "Classroom Organization," Encyclopedia of Education Research, (New York: The Macmillan Company, 1960), p. 222.

are by no means ready to learn together or do not possess the same qualities. The fact that home background plays such a vital role during the first five years of life indicates that even by age five the mental age differences may vary considerably. The conventional method of grading children was the easy way out during the years when education was greatly expanding. It became so ingrained in the minds of people that it became the only way to organize a school.

Because all students at a certain chronological age were placed in the grade level corresponding to it, the bright students and the handicapped students were really left out of the picture. This obviously would frustrate the individuals who found themselves out of the so-called average group in the class. Couple with this the problem that these children fail to achieve the next grade level and social problems are placed on top of the educational problem which already existed and you have the potential dropout after the first year of school. "What is needed today and for the future is an educational system so organized that every child can be provided for in keeping with his ability, his interests, and his time-table of development."[4] In the nongraded school, no child is forgotten. At least some attempt is made to throw off the yoke of placing a child because of chronological age and disregarding his mental capabilities. In its place is a system whereby every child works at a level he is capable of mastering, and he can do so at his own rate of speed.

Although the non-graded school has been growing steadily in many

---

[4] Vincent DiPasquale, "Schools Without Grades," Better Homes and Gardens, 33:28, 1955.

areas of the country, there has not been enough evidence compiled to reach any conclusive results as to whether this plan of school organization is any more effective than the traditional graded method. It would seem necessary that definite, significant differences be exposed if actual proof of quality is to be realized. Most opinions which have been expressed are simply reports based on opinion rather than fact. Williams conducted a study of academic achievement between students in a graded and non-graded school setting. The experimental groups consisted of thirty-eight pupils who had attended a non-graded school for three years. The control group consisted of thirty-eight pupils who had attended a graded school for three years. The students were matched on the basis of age, sex, and intelligence. Grade equivalents were used and a total score was used. The test used was the Stanford Achievement Test and scores from the sections dealing with Language and Arithmetic were used. Her study reached the following conclusions:

1. The results do not show that the slower pupils profit more from the non-graded structure. Since the slower pupils of the graded school achieved significantly higher than the slower pupils of the non-graded school, this study refutes the claim made by some that the graded structure is responsible for reading failures and mental health problems.

2. In pupil achievement, the pupil-teacher ratio may be more important than graded or non-graded organization.

3. Thia study also confirms the statement that the graded schools are aware of the differences in children's abilities and allow for these differences in planning and instruction.

4. When the entire study is considered, however, there does not appear to be a significant relationship between school organization and pupil achievement. In both schools the pupils were achieving above the norms provided by the test.[5]

---

[5] Wilmajean Williams, "Academic Achievement in a Graded School and in a Non-Graded School," Elementary School Journal, 67:135-139, Dec., 1966.

Research in the area of performance between the two methods of organization is comparatively scarce. Since 1959 to the present time, six comparative research studies in reading achievement have been reported. The summaries, with no attempt to evaluate the quality of the investigations, are as follows:

1. Four studies found the performance of the non-graded pupils significantly superior to that of the graded pupils;
2. One found no difference;
3. One found the graded control group significantly better than the non-graded experimental group.[6]

While the research in the area of this study doesn't overwhelmingly endorse the non-graded organization, it is quick to point out the lack of sufficient data to rule out qualities contained in its makeup. As was further stated in the article, "most of the studies were based on new or relatively short experiences with non-grading. The variations in the program make it impossible to treat them as six replications of the same treatment."[7]

Grouping students, which is the backbone in the non-graded organization, deserves to be mentioned due to the criticism it has received. The general concensus is in favor of some type of grouping, if grouping seems warranted at all. Some violently oppose grouping, whether by intelligence or ability, on the basis of discrimination and labeling. Johnston even goes so far as to say:

> Grouping by intellectual ability, as it is generally practiced, is intellectual segregation, which has had effects on the bright groups

---

[6] Louis T. DiLorenzo and Ruth Salter, "Co-operative Research on the Non-Graded Primary," Elementary School Journal, 65:273-274, Feb., 1965.

[7] Ibid., p. 274.

as well as the slow groups. Ability grouping is neither necessary or desirable. Intellectual segregation of this kind may be as damaging to personality as racial segregation or other kinds of segregation and for many of the same reasons. If democracy is to be fostered in our schools, such barriers to self-respect and the sharing of ideas must be removed.[8]

The basic philosophy is sound and realistic. The problems are establishing the limits of the non-graded program and determining the point of emphasis as far as the subject areas are concerned. These of course will differ as far as area of the country is concerned and to what ends or objectives the total program is aimed. The fact that our present educational program should always be under scrutiny for improvement lends credence to this system of teaching.

Purpose of the Survey

It is the purpose of this survey to investigate the relationship between students who have spent three years in a non-graded school situation and students who did not have the opportunity of the non-graded classroom situation. The relationship to be measured involves the differences in reading achievement. More specifically it is designed to:

1. Show the relationship between two groups of students; one group in the graded situation and one group in the non-graded situation, as far as intelligence test scores are concerned;

2. Show the relationship between two groups of students; one in the graded situation and one group in the non-graded, as far as achievement tests scores in reading are concerned;

3. Note the progress, if any, between the two groups as far as mean scores in reading after the fourth, fifth, and sixth grades of school in the graded and non-graded school situations.

The results of this survey are valid only for the two specific groups involved. It is important to note that undoubtedly some variables

---

8 A. Montgomery Johnston, "Intellectual Segregation," Elementary School Journal, 67:212, Jan., 1967.

which influenced the results were not measured. Since most of the students involved were of basically the same socio-economic group and since other variables, such as number of library materials available were not taken into the study, we can conclude that the results are valid only for the test scores.

Other terms which will be mentioned and explained are:

Group A - Thia refers to the group of students who spent grades two, three, and four in the regular graded situation.

Group B - This refers to the group of students who spent grades two, three, and four in a non-graded situation.

Intelligence test score - as used in this study refers only to the I.Q. score yielded by the Lorge-Thorndike Intelligence Teat, Level two, Form A, Primary and Intermediate Battery. This test was given in Grades two, three, five.

Achievement teat score - as used in this study refers only to a score yielded by the Stanford Achievement teats, Intermedidate batteries one and two, forms X, Y, Z. This test ia given yearly by the district.

S.R.A. Primary Mental Abilities - a test given during the month of September in the Forrest-Strawn-Wing Unit Schools. In a review of this test, John E. Milholland states:

> Five "primary mental abilities" labes are used to designate the subjects. The presence of and emphasia given to each of the abilities in the various levels reflect the judgement of the authors with respect to the relative importance of these abilities at the indicated grade levels. The advice is also given that a pupil making a total atanine score of seven or above has an indication that he has "the mental ability to do well on college entrance examinatione and to do good college work." Those with stanine scores of five or six may find it "difficult to gain admission to a highly selective college." Those with stanine scores below four, however, are advised that it is somewhat doubtful that other factors could compensate sufficiently to enable you to do college work succesafully.[9]

Need for the Survey

It seems that a program of any type must be periodically analyzed

---

[9] Oscar K. Buros, Editor, Mental Measurements Yearbook 6th edition, The Gryphon Preas, Highland Park, New Jersey, 1965.

to determine its effectiveness. This is doubly important when it comes to the reading program in a public school. Harris emphasizes this point when he says:

> The importance of reading is clearly recognized by the elementary school. In the primary grades, more time and effort is spent on teaching reading than on any other phase of the school program. More money is spent on reading matter than on any other type of school supplies. Poor reading is recognized as the most important single cause of retardation in the elementary school (although it, in turn, may be due to low intelligence). 10

The fact that elementary school, particularly the primary grades, places so much importance on basic skills and because these skills represent the foundation for the learning of all other skills, the need is axiomatic. It is also true that what constitutes the "best" reading program has never been standardized in the school systems of the United States. What is effective in one district doesn't always achieve the same results in other districts. Consequently, during this century, emphasis on what is the best method has shifted drastically from the beginning of this century. As Smith points out in a recent article:

> During the period from 1950-1960, for the first time in history, reading instruction in American schools underwent harsh and severe criticism by laymen. Some people maintained that the criticisms were unfair and rose to the defense of their methods through articles, speeches, discussions, and investigations. Several comparative studies of "then and now" were made. These studies, on the whole, showed that we were teaching reading as well as or better than in preceeding years.
>
> Insofar as progress is concerned the criticism by laymen probably had three good effects: it caused school people to examine their present methods more carefully; it stimulated the interest of parents and others in reading instruction; it offered motives and opportunities to school people to explain the research, psychology, and philosophy on which present methods are based. So in this situation, as is often the case in other

---

10 Albert J. Harris, How to Increase Reading Ability, 4th edition, David McKay Company, Inc., 1961, p. 3.

situations, even criticism caused reading to move forward.[11]

The problem here constitutes the query as to what is the best method of instruction. This forms the basis for the evaluation contained in this survey.

No one can say, particularly with the wealth of material available today and the importance of better comprehensive reading skills, that his program is the most effective and doesn't need to be changed. Positive progress comes about through farsighted thinking and planning. A more thorough knowledge of new systems available is the progressive attitude necessary in education. We must always look for better methods of doing things. This is particularly true in a field as basic to education as reading. Smith upheld the questioning attitude when she said:

> What I am trying to say is that while our accomplishments have been very great, indeed, it may be that we have only penetrated the first layer, the troposphere, so to speak. Undoubtedly, brilliant new insights will be revealed, ingenious new techniques of experimentation will be involved. Possibilities of such developments portent opportunities for unlimited achievement in the future.[12]

Forrest-Strawn-Wing school district has tried to upgrade its reading instruction by the adoption of the non-graded primary levels with emphasis on reading. A district, in order to be progressive and forward looking, must search for better ways of doing the job it has set out to do. But it fails miserably if it changes for the sake of change and does not question whether or not the new road is in fact a better road to travel. Many studies have proven that individual pupils can make gains under all approaches used in reading. The question is if the approach that has been

---

[11] Nila B. Smith, "What Have We Accomplished in Reading?" Teaching Reading: Selected Materials, ed. William Barbe (New York: Oxford University Press, 1965), pp. 43-44.

[12] Ibid., p. 45.

adopted is fulfilling the expectations of all concerned. This survey will attempt to shed some light on this question. Schubert says that:

> It is a mistake to assume that just because children are free from discernible sensory impairments that they will learn as easily when one method is employed as another.[13]

---

13 Ibid., p. 126.

Chapter II

CHOICE OF GROUPS AND STATISTICS

Groups Chosen and Methods of Choice

The non-graded philosophy of reading was introduced into the Forrest-Strawn-Wing curriculum during the 1963-64 school year. The first problem was to determine which groups to use in the survey. Upon examining the information available, it became quite apparent that the size of the group and the total amount of test information would determine which groups would be compared. Also, a better group in the non-graded class would be one which had the benefit of being a truly representative group of the non-graded philosophy. The first years undoubtedly would be handicapped by lack of teacher knowledge about the new approach and the students themselves would be better adjusted to the classroom organization after the first year. The size of the group would be important in that the larger the group the more valid the results.

Since the Unit 2 school district is rather small, 1967 enrollment 811 students, and because a certain number of these students are transfers in and out of the district, the amount of cumulative test information was important in the selection of Group A, or the graded class. The present class of sophomores was chosen and after eliminating those students with incomplete records the group left was 50 students. The present sixth grade class was chosen as Group B because, not only were they a group who had progressed completely through the non-graded classes,

but information gathered on them could be used immediately by the teachers involved in the teaching of these students. After all students had been eliminated who had incomplete data, this group boiled down to a group of 56. Since fifty was the number of students contained in Group A, by chance a number were selected to reduce the number of Group B to 50.

The next step was to show the relationship between Groups A and B according to intelligence. Chart A shows both groups and their intelligence data on the Primary Mental Abilities Tests and the Lorge-Thorndike Intelligence Tests. The P.M.A. was given to both groups in their first year of school. The Lorge-Thorndike was given to Group A during their second and third years in school, and to Group B during their third and fifth years in school. It is quite apparent that the two groups, without a specific attempt to match score for score, are well suited for the survey. A mean difference of only 1.5 points exists between them. Walker says:

> Remember that very small differences between scores should not be considered as inconsistencies. Discrepancies of two, three, or five points are not generally worth concern. It's the larger ones you need to look into - and the larger they are, the more likely it is that they reflect the true situation.[14]

Need For Evaluative Measures

The data on which evaluation is based come from many sources. Commonly these include teachers' grades based on classroom recitation, teacher-made tests or quizzes, or subjective judgments on the part of the teacher. Standardized tests then become supplemental devices which aid the teacher in determining the point to which a child has progressed. The

---

[14] Robert Walker, "How to Understand and Use Test Results," Grade Teacher, (April, 1968), p. 10.

INTELLIGENCE TEST DATA

Chart A

	P.M.A. Grade 1	Lorge-Thorndike 2	Lorge-Thorndike 3	Mean Intelligence
Median	108	111	111	
Mean	108.2	111.9	110.4	110.1
Range of Scores	88-137	83-133	85-135	

Group A

	P.M.A. 1	Lorge-Thorndike 3	Lorge-Thorndike 5	Mean Intelligence
Median	112	108	109	
Mean	109.8	109	107.2	108.6
Range of Scores	82-130	80-135	71-134	

Group B

best of these reflect careful curriculum research, pre-standardization tryouts, and standardization on representative population.[15]

Evaluations, or some method of appraisal, are needed because instruction produces outcomes. Because work samples are varied, and because there can be so many areas of instruction, different methods of appraisal are necessary. The English teacher may require a theme which will be evaluated in terms of specific predetermined objectives. The physical education teacher might require a physical fitness test which has certain established goals of fitness at various age levels. The speech teacher might compare or evaluate a student's speech based on a rating scale. All of the previous examples necessitate a common system which would be able to permit a comparison of one student with another student, or permit a comparison of one student with other students, or permit a comparison between groups of students involved in the same area of instruction.

It becomes difficult to establish a system which will meet all of the desired or required objectives. Durost lists the following requirements for a system of comparable measures:

1. It must be simple and easy to understand;
2. It must permit statistical manipulation, including the weighting and combining of the various separate elements (scores) into a composite or total score, with each element receiving the weight judged most appropriate;
3. It must also be possible to express the degree of the relationship between one element and another, either graphically or statistically;

---

15 Walter N. Durost, "The Characteristics, Use, and Computation of Staninea," Test Service Notebook, No. 23, p. 1. (Harcourt, Brace, and World, Inc., 1961).

4. It must have a rational basis that will staisfy persons with widely divergent backgrounds of training and experience.[16]

The responsibilities of administrators, teachers and guidance workers for knowing about their students and the curriculum in which these students function cannot be over-emphasized. The quastion becomes which method of evaluation will most accurately and most clearly enable those in charge to determine the effectiveness of their programa. The key to all of this knowledge is the ability to learn how to serve better the needs of the students in the school system.

Intelligence Testing

Intelligence tests, according to moat authorities, measure the reaults of hereditary plus environmental factors to a much greater degree than do achievement tests. These influences are rather general in nature and are common to all children who live in what might be termed normal environments and conditions. A note of caution should be sounded at this point. Intelligence test scores are vital to educational planning because they are, in some measure, predictive of probable success in school and discriminate between the ablest and the least able. However, educators hold a healthy skepticism about any single test score and make use of it as an estimate rather thsn a literally true and exact measure of intelligence. Educators study the complete record of the child and are slow to draw any but tentative conclusions until the evidence is overwhelming. A single I.Q. is a potentially dangerous piece of information

---

16 Durost, Ibid.

unless its values and limitations are fully understood.[17]

It must be understood that an I.Q. score is just that. It simply represents what a student scored on that particular test and will vary somewhat with each test. For this reason no less than three I.Q. scores on each student involved in Group A and B were used and a mean score was derived.

Achievement Tests

The Stanford Achievement Test is the designation of a series of comprehensive achievement tests developed to measure the important knowledge, skills, and understandings commonly accepted as desirable outcomes of the major branches of the elementary curriculum. The tests are intended to provide dependable measures of these outcomes, comparable from subject to subject and grade to grade, for use in connection with improvement of instruction, pupil guidance, and evaluation of progress. The tests have been planned with a view toward simplicity of administration, scoring, and interpretation, so that they may be used effectively by persons with little or no formal training in the use of standard tests.[18]

What is the most meaningful kind of achievement? The answer to thia queetion must be determined by the values of both the individual and the society. For example, most of us will agree that the achievements of a skilled medical doctor are praiseworthy. In the realm of educational measurement, the most meaningful achievement is almost certainly academic

---

17 Walter N. Durost, "How To Tell Parents About Standardized Test Results," Test Service Notebook, No. 26, p. 2. (Harcourt, Brace, and World, Inc., 1961).

18 Manual Stanford Achievement Test, (New York: Harcourt, Brace, and World, Inc., 1964), p. 2.

success. Achievement tests which are different measures of school subjects are the most significant kind of tests simply because success in school is important. The individual's self-esteem may depend greatly on this success, and his progress in school is a substantial indication of his potential for becoming a person who will make positive contributions to society. It is obvious that achievement tests on school subjects are the most important standardized tests.[19]

Intelligence and Achievement Test Score Relationships

Because elementary schools use intelligence and achievement tests to a greater degree than all other tests, relationships between the two types should be established for evaluative purposes. The child goes to school to learn to use language correctly, and to understand the social system in which he lives. These specific skills and knowledge develop normally in the child to the extent to which he responds to instruction. A comparison of the ability measure and the achievement measure will help reveal the degree of consistency between the child's school performance and his measured ability.[20]

On the other hand, it must be remembered that some problems also arise when trying to compare the relationships between scores on intelligence and achievement tests. Blythe C. Mitchell, a member of the test department of Harcourt, Brace, and World Publishing Company, stated the following sources of error involved in the comparison. They are:

1. There is often a lack of comparability of the paired achievement and intelligence measures. An educational age derived from age

---

19 Quentin Stodola and Kalmer Stardahl, Basic Educational Tests and Measurements, (Chicago: Science Research Associates, Inc., 1967), pp. 195-196.

20 Herman J. Peters, Guidance In The Elementary Schools, (Chicago: Rand McNally and Company, 1965), p. 115.

norms of an achievement battery may not be compared with the mental age taken from an entirely different population.

2. The use of other types of relative measures (percentile ranks and stanine levels) does not afford proper comparison unless the ranks for achievement and for intelligence are based on the same reference population.

3. A requirement often ignored is the need to take account of the varying part that intelligence plays in the specific areas of achievement. Correlations with reading and science, for example, are generally found to be higher than those with spelling and arithmetic computation. The school that expects identical achievement in all subjects for a given level of intelligence is failing to take account of these differentiated relationships. The achievement expected or predicted for a given level of intelligence must be established separately for each subject test in an achievement battery.[21]

In theory, ability tests are especially designed to measure potential for future achievement, whereas achievement tests are aimed at measuring present achievement. The distinction between the two is primarily one of purpose rather than content, since the material in the two types often overlaps. Some intelligence tests, for example, may have approximately the same content as certain tests of reading and arithmetic achievement. In a sense, all tests are achievement tests in that they measure previous learning. Actually, there might be some advantage in considering achievement and ability tests to be in the same broad category. Attention would then be focused on the fact that all test scores are dependent upon previously learned behavior, whether they measure present status or predict future achievement.[22]

With the listing of some advantages and disadvantages, encouragements,

---

[21] Blythe C. Mitchell, "A Comparison of the Achievement-Intelligence Relationship for Pupils With That for School Systems," The Journal of Educational Research, Vol. 57 No. 4 (December, 1963) p. 179.

[22] Quentin Stodola and Kalmer Stardahl, op. cit., p. 167.

and cautions, it seems that what is specifically touched upon is the untenable ability of people, tests, and testing companies to feel that one method of evaluating a program is certainly the best and most accurate. From the evidence available there does exist some relationship between intelligence test scores and achievement test scores. Aware of the pitfalls and cautions, this method of comparison was chosen as possibly the best method, considering the facts available from the school records on the students involved.

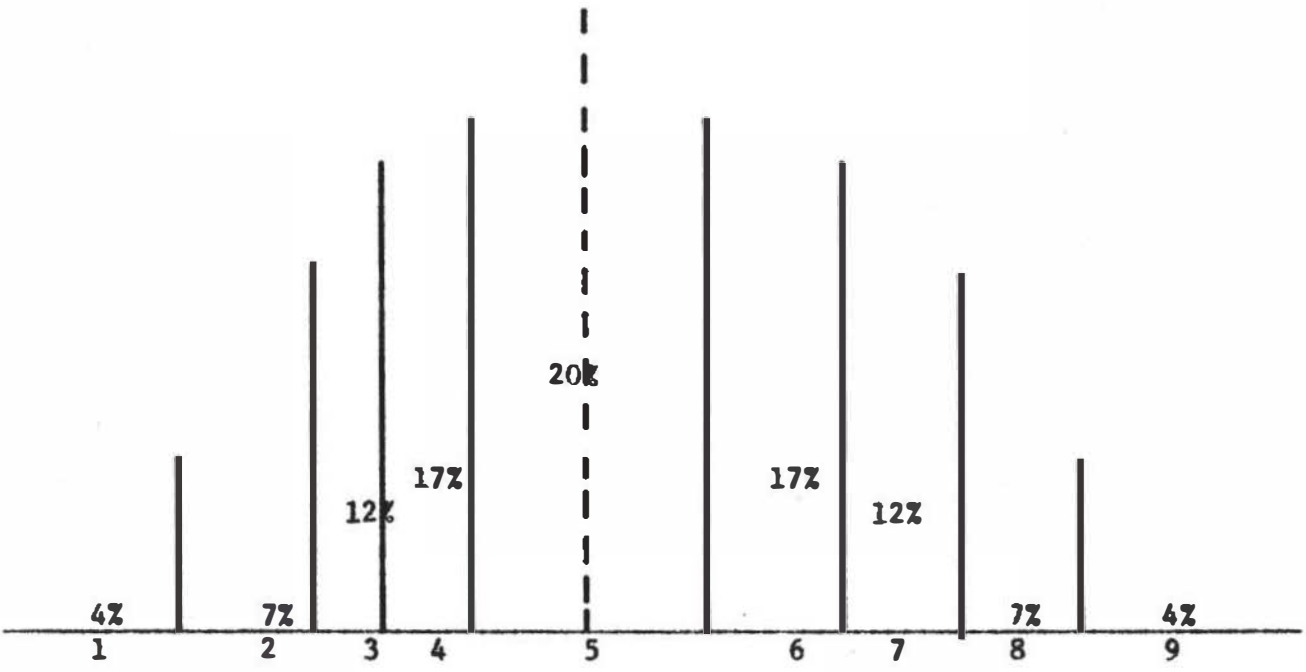
Use Of Stanine Scores

A special comment should be made about stanine (an abbreviation for standard nine) scores, which are slightly different from those based on standard deviations. In the stanine system the area under the normal curve is arbitrarily divided into nine segments or stanines, with the first stanine representing the lowest segment and the ninth the highest. In general, the procedure is to assign to stanine five the area lying one-fourth of a standard deviation above and one-fourth below the mean, and then to divide the remainder of the distribution into units of one-half standard deviation above and below stanine five. Thus each stanine except one and nine is half a standard deviation in width; one and none are at the ends of the distribution and include all cases falling at the extremes. As can be seen in Chart B, a stanine represents a band of scores. Chart B also illustrates the percent of cases represented by each stanine in a normal distribution.[23]

---

[23] Quentin Stodola and Kalmer Stardahl, op. cit., p. 85.

Percentage Of Cases Included In Each Stanine Score



Stanines

Chart B

Durost, on page 14, listed the requirements for a system of comparable measures. Because the use of stanine is relatively new, the amount of material written on their use is scant. With the adoption of achievement score profile sheets to include stanines, it would seem obvious that test publishers see a great future in their use. The scores from achievement tests are now graphed in grade scores, percentile ranking, and stanines. If test results are to be used for evaluative purposes, the need for an easy reporting system is apparent. He has the following comments on the use of stanines:

The requirements outlined previously seem to be adequately met by a system of standard scores called stanines. While test scores and numerical data of various kinds have long been transformed from raw scores derived from original measures to standard scores, the notion of using stanines did not come into extensive use until World War II. At that time the Air Force psychology program sought a means of translating its test data into

simple workable form that would permit the mass analysis of hundreds of thousands of test scores with minimum labor. For this purpose the stanine system was adopted and used effectively. Since the war the use of stanines has gradually increased, both as a research device and as a means of interpreting test data.[24]

In view of the shortcomings of reporting scores by the common methods, I.Q., grade equivalents, percentile rank, it is suggested by Durost that the best method of reporting standardized test results to people, other than those whose knowledge includes a background in the area, is in terms of stanines. He gives the following reasons for this method of reporting:

1. They are more dependable than any of the other common methods of reporting scores because they are broader units, although precise enough for our purposes.

2. Stanines make the test results comparable for the individual from test to test as long as the group on which they are based is the same.

3. They are relatively easy to explain to parents.[25]

Stanine usage for reporting scores and making comparisons of groups has become quite popular in recent years. Because the scores are easier to interpret and understand, it seems reasonable to assume their usage will be increased.

---

[24] Durost, 23, p. 1.

[25] Durost, 26, p. 2.

Chapter III

REPORT OF FINDINGS

The purpose of this study was to investigate the relationship between students who have spent three years in a non-graded school situation and students who did not have the opportunity of the non-graded classroom as far as reading ability was concerned.

The two groups that were selected for this comparison were the sophomore class of 1967-68 and the unit sixth grade classes of 1967-68 in the Forrest-Strawn-Wing Community Unit #2 in Livingston County. Because the size of the unit in population is small, the number in each group involved was only fifty students. The sophomore class had no benefit of the non-graded classes while the two sixth grades involved have progressed through three years of school in the non-graded classes.

Intelligence quotient scores of all the students in both groups were used so that a mean score of intelligence could be arrived at for the purpose of determining the similarity of both groups according to intelligence. If the mean difference was small, it would seem that this ideally would make the two groups more suitable for the purpose of comparison. The mean difference was 1.5 points which made the two groups compatible for comparison as far as intelligence scores are concerned. Each of the students in both groups had three I.Q. scores from which the mean for each group was derived.

Achievement test scores in reading from the Stanford Achievement

Test taken by each student during their fourth, fifth, and sixth years of school were the next scores that were gathered on each student. These scores were derived from the sections of the test involving word meaning and paragraph meaning. Chart C shows the position of each grade as far as mean or median grade equivalent scores for the total group over the three-year period.

All evaluations and interpretations on the following charts will be in terms of stanine rankings rather than by percentile or mean score differences. The reasons for this method of interpretation have already been discussed previously in the paper. There is no attempt to definitely state which group is superior in reading or which group is weak. The attempt is to point out what generally seems to be apparent or similar between Group A and Group B.

Chart C

Stanford Achievement Test

Year Grade		1961-4	1962-5	1963-6
Median Score	Word Meaning	4.6	5.4	6.7
	Paragraph Meaning	4.3	5.9	6.9
Mean Score	Word Meaning	4.77	5.61	6.85
	Paragraph Meaning	4.42	6.04	6.91
Range of Scores	Word Meaning	23-86	27-108	36-102
	Paragraph Meaning	18-71	34-107	32-105

Group A

Stanford Achievement Test

Year-Grade		1965-4	1966-5	1967-6
Median Score	Word Meaning	4.1	5.4	5.9
	Paragraph Meaning	4.1	5.3	6.4
Mean Score	Word Meaning	4.16	5.54	6.19
	Paragraph Meaning	4.10	5.31	6.48
Range of Scores	Word Meaning	23-70	25-88	39-93
	Paragraph Meaning	18-80	24-95	32-110

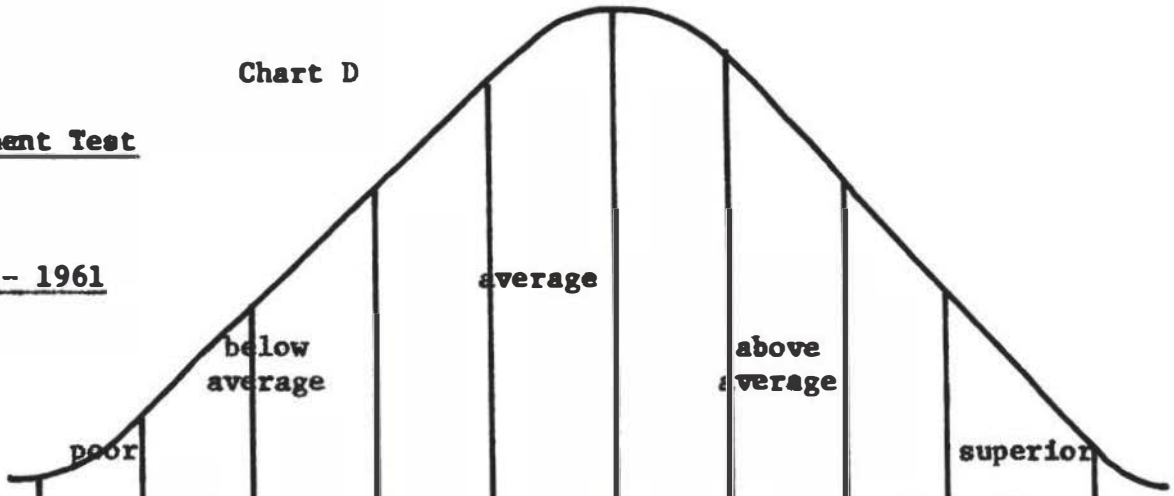
Group B

Chart D

Stanford Achievement Test

Sophomore Class

Word Meaning 4th - 1961



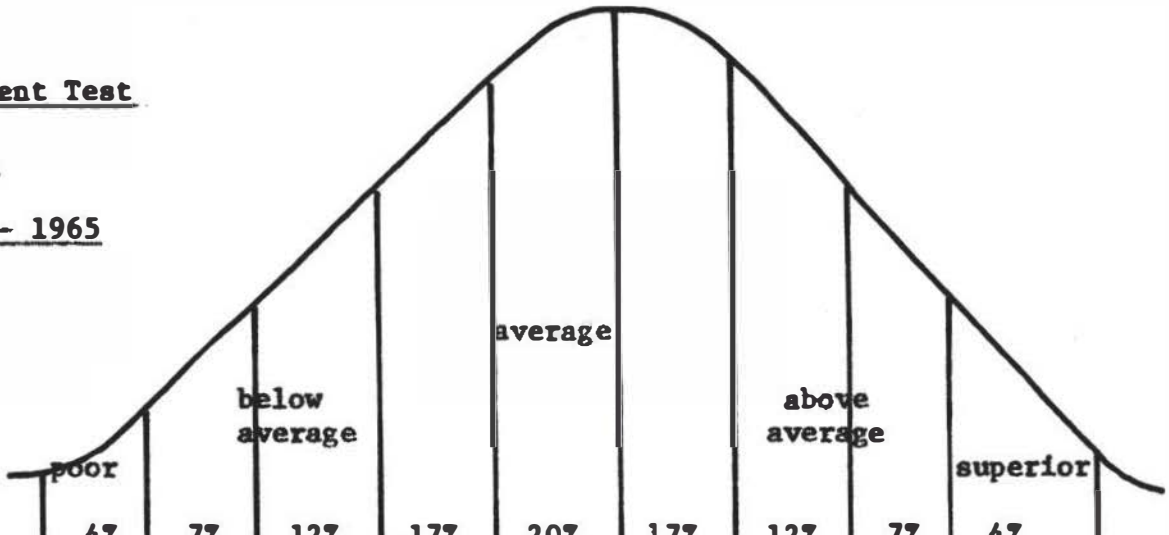
	4%	7%	12%	17%	20%	17%	12%	7%	4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	2	1	9	17	14	4	2	0	1
Range of Scores	23-25	26-30	31-38	39-48	49-61	62-71	72-79	80-83	84-86

GROUP A MEAN SCORE - 47.7

Stanford Achievement Test

Sixth Grade Class

Word Meaning 4th - 1965



	4%	7%	12%	17%	20%	17%	12%	7%	4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	1	3	10	12	13	6	3	0	2
Range of Scores	23-24	25-27	28-33	34-41	42-50	51-58	59-64	65-67	68-70

GROUP B MEAN SCORE - 41.6

Chart D - The range of responses for Group A was sixty-three points and the number of students who fell into the average range was thirty-five. Twelve students ranged in the below average distribution while three fell into the above to superior range. The mean stanine for Group A was four. Group B on the other hand had a response range of forty-seven with thirty-one students in the average category, fourteen in the below-average to poor range, and five in the above-average to superior section. The mean stanine for Group B was five.

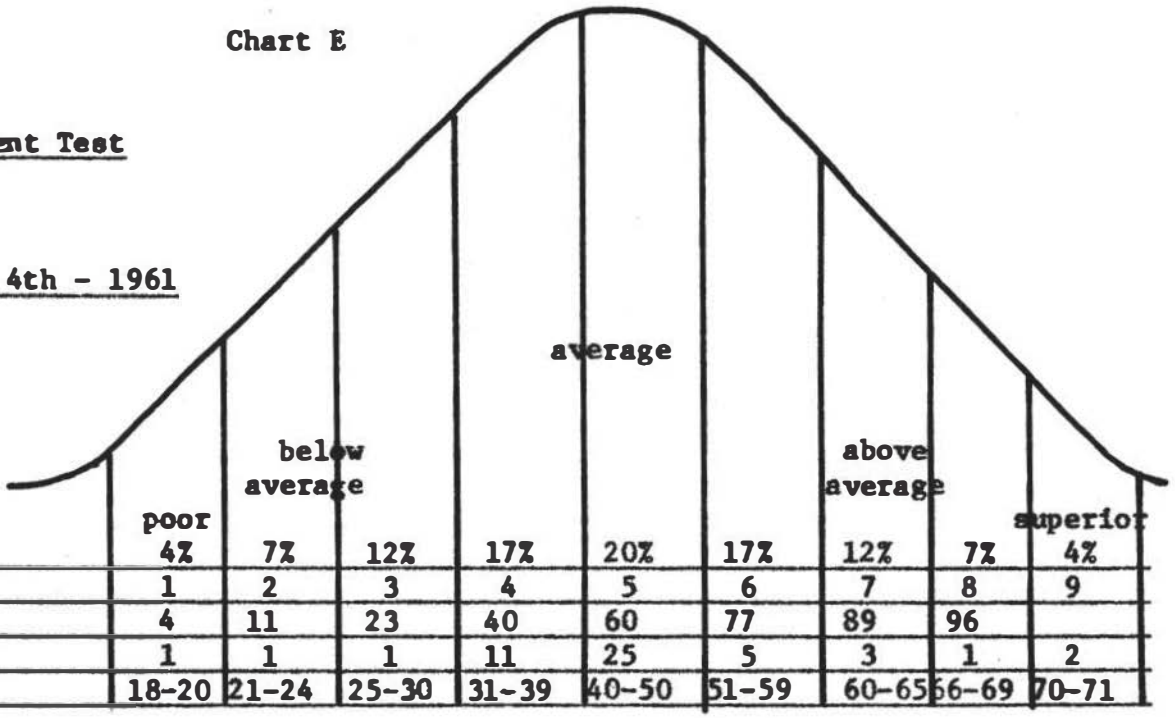
Due to the closeness of scores, mean stanines of four for Group A and five for Group B, the difference is insignificant. It would seem that both groups on this test during the year involved showed no apparent difference in reading achievement.

Chart E

Stanford Achievement Test

Sophomore Class

Paragraph Meaning 4th - 1961



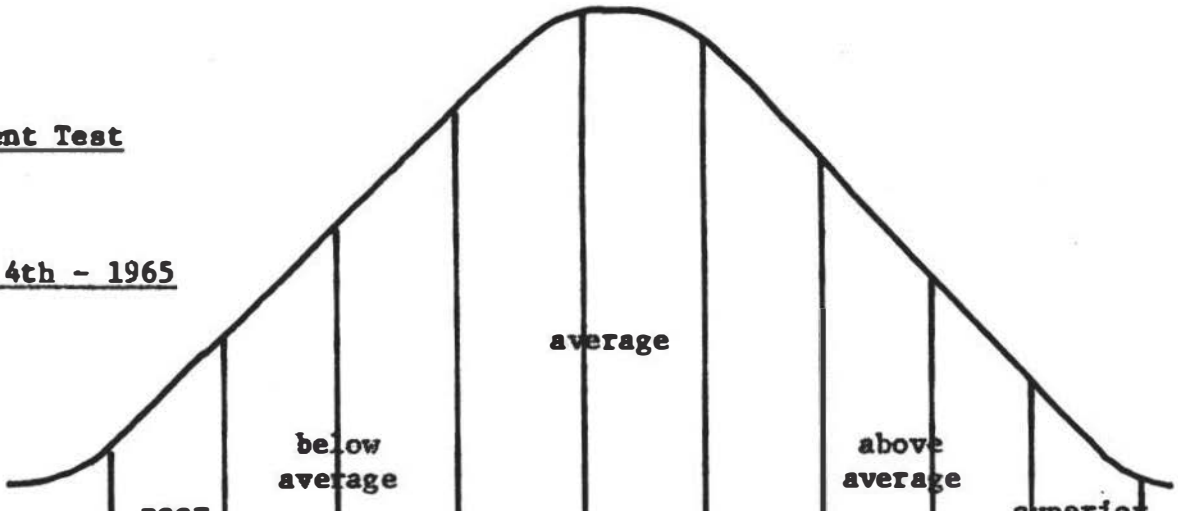
	poor 4%	7%	12%	17%	20%	17%	12%	7%	superior 4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	1	1	1	11	25	5	3	1	2
Range of scores	18-20	21-24	25-30	31-39	40-50	51-59	60-65	66-69	70-71

GROUP A MEAN SCORE - 44.2

Stanford Achievement Test

Sixth Grade Class

Paragraph Meaning 4th - 1965



	poor 4%	7%	12%	17%	20%	17%	12%	7%	superior 4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	2	2	12	15	12	6	0	0	1
Range of scores	18-20	21-25	26-32	33-43	44-55	56-66	67-73	74-78	79-80

GROUP B MEAN SCORE - 41.0

Chart E - The range of responses for Group A on this teat was fifty-three. The number of students who fell into the average classification was forty-one while three can be found in the below-average to poor range and six in the above-average to superior range. The mean stanine for Group A was five. Group B had a response range of sixty-two with thirty-three students in the average range, sixteen in the below-average to superior range. The mean stanine for Group B on this test was four.

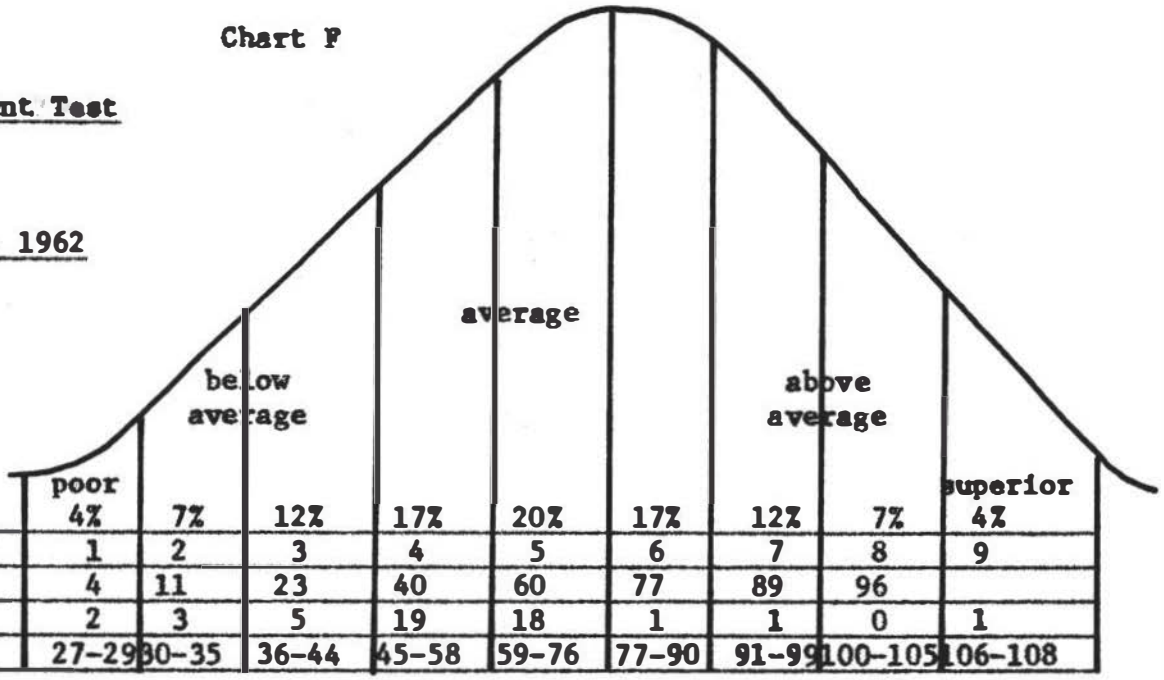
A general summation of the two groups on their reading understanding and vocabulary at Grade four seems to be that while Group A on the test of paragraph meaning placed more students in the above-average to superior range, and Group B placed more students in the above-average to superior range on the vocabulary test, each offset the other and their mean staninea are equal. Group A might be somewhat better with comprehension of reading material, but Group B commands a slight lead in the vocabulary section. No general difference seems apparent on this comparison.

Chart F

Stanford Achievement Test

Sophomore Class

Word Meaning 5th - 1962



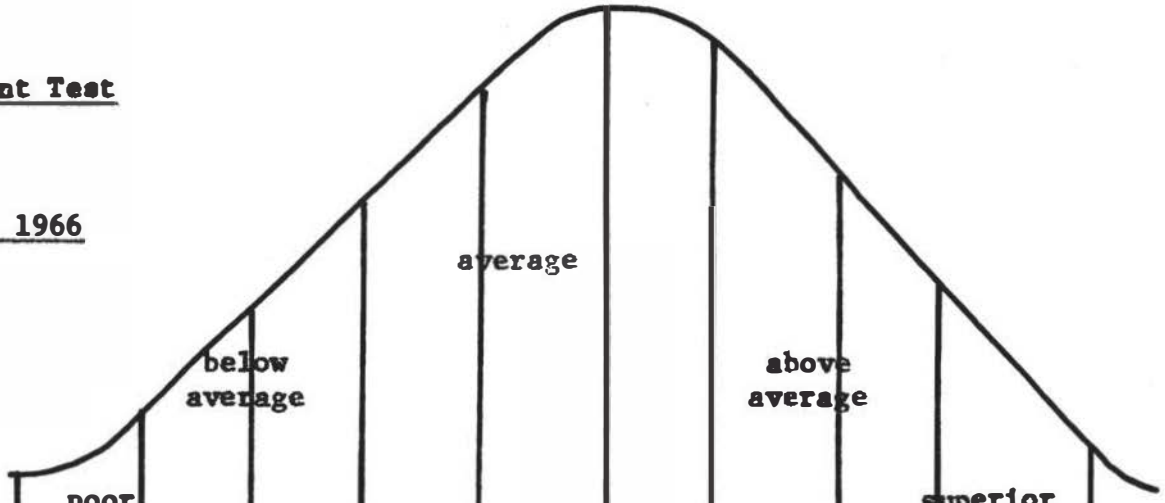
	poor 4%	7%	12%	17%	20%	17%	12%	7%	superior 4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	2	3	5	19	18	1	1	0	1
Range of scores	27-29	30-35	36-44	45-58	59-76	77-90	91-99	100-105	106-108

GROUP A MEAN SCORE - 56.1

Stanford Achievement Test

Sixth Grade Class

Word Meaning 5th - 1966



	poor 4%	7%	12%	17%	20%	17%	12%	7%	superior 4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	1	0	6	13	15	7	4	3	1
Range of scores	25-27	28-31	32-38	39-49	50-62	63-73	74-80	81-84	85-88

GROUP B MEAN SCORE - 55.4

Chart F - The range of responses for Group A on this teat was eighty-one. The number of students falling into the average stanine range was thirty-eight. Group B had a response range of sixty-three, eighteen points less than Group A. The number of students found in the average group for Group B was thirty-five. The difference in number of students in the average range is only three. Group A placed ten students in the below-average to poor range, while three students achieved the above-average to superior category. Group B placed ten students in the above-average to superior range with seven students in the below-average to poor range. The mean stanine for Group A was five, for Group B four.

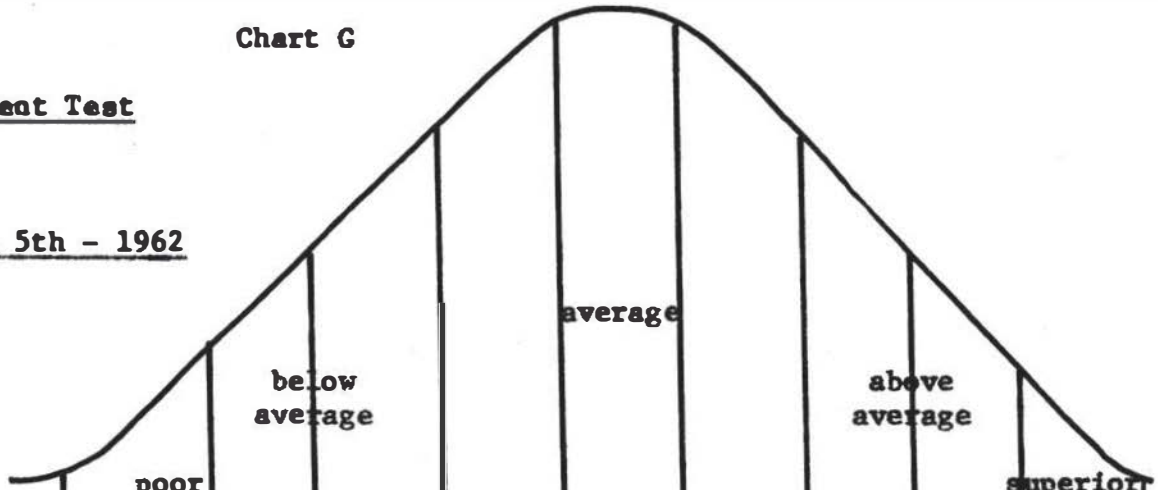
On this teat the two groups involved had a more even distribution of scores than appeared during the compilation of test scores showing their fourth-grade performance in this same area. Once again they seem well matched in level of performance.

Chart G

Stanford Achievement Test

Sophomore Class

Paragraph Meaning 5th - 1962



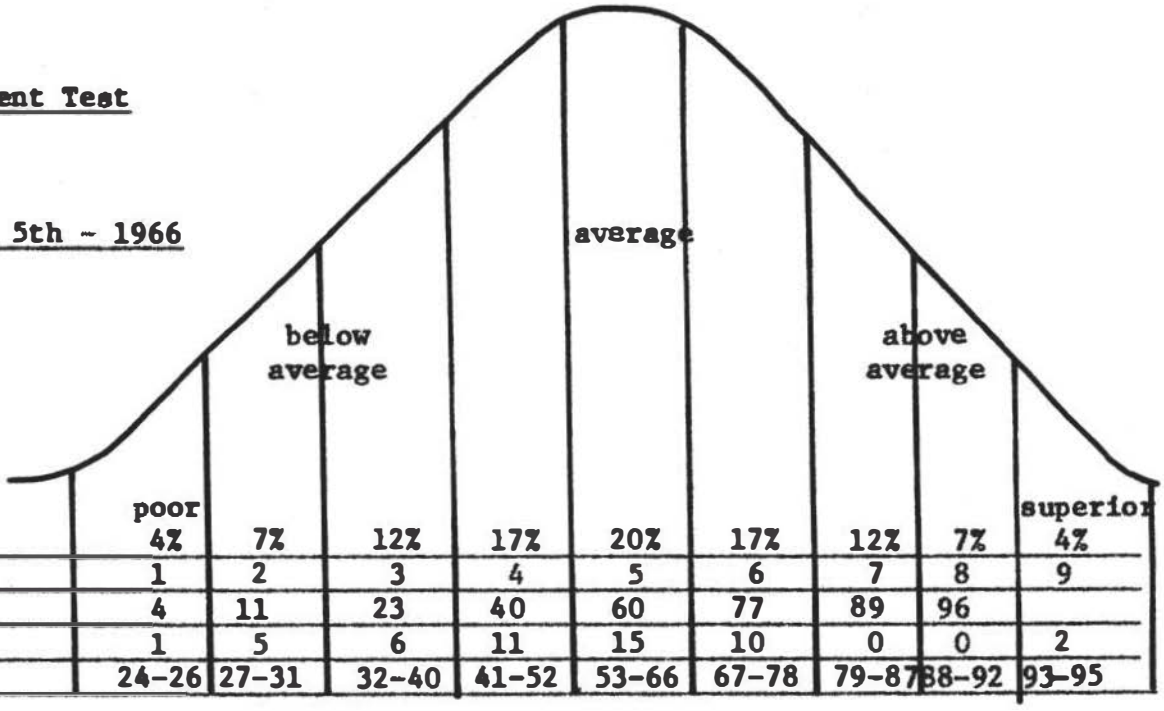
	poor 4%	7%	12%	17%	20%	17%	12%	7%	superior 4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	2	2	5	20	18	2	0	0	1
Range of scores	34-36	37-41	42-50	51-62	63-77	78-89	90-98	99-103	104-107

GROUP A MEAN SCORE - 60.4

Stanford Achievement Test

Sixth Grade Class

Paragraph Meaning 5th - 1966



	poor 4%	7%	12%	17%	20%	17%	12%	7%	superior 4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	1	5	6	11	15	10	0	0	2
Range of scores	24-26	27-31	32-40	41-52	53-66	67-78	79-87	88-92	93-95

GROUP B MEAN SCORE - 53.1

Chart G - This chart illustrates the performance of two groups during their fifth year of school involving that section of the test which tests comprehension. This test, like the facts shown on Chart E, which involves the test the same year, shows remarkable similarities between the groups. The range of scores for Group A was seventy-three, for Group B the range was seventy-one. Group A placed forty students in the average stanine area and Group B placed thirty-six in the average area. Group B had two students in the above-average to superior range; Group A placed one student. Nine students fell into the below-average to poor range in Group A while Group B had twelve. The mean stanine on this test for Group A was four, for Group B five.

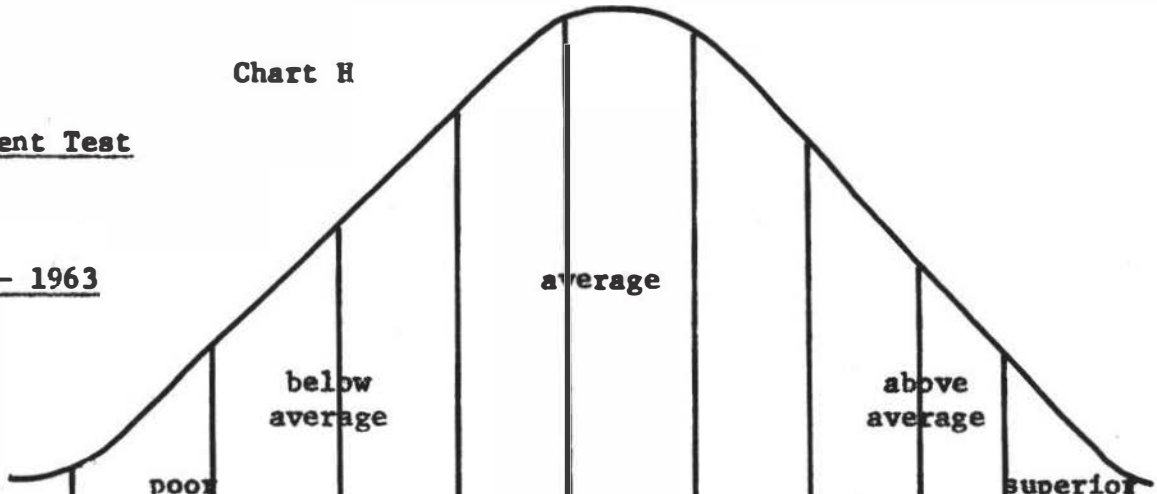
In summarizing the test scores for their fifth year in school it seems that any differences which existed at the fourth year have narrowed even further during their fifth year. If we can state no general differences at the fourth year of school, we are almost compelled to say no differences exist at the fifth level. The similarity of the two groups at this point is striking.

Chart H

Stanford Achievement Test

Sophomore Class

Word Meaning 6th - 1963



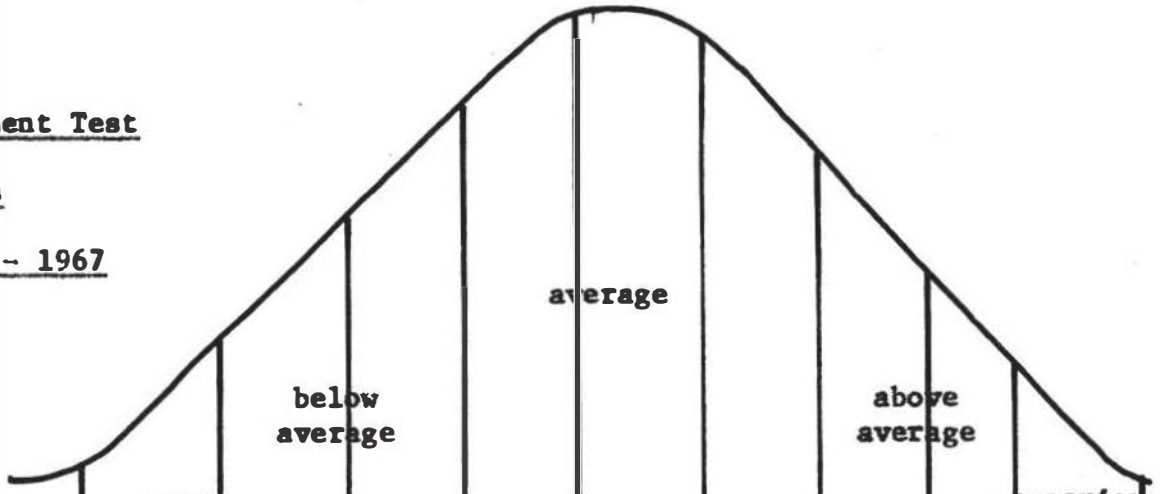
	4%	7%	12%	17%	20%	17%	12%	7%	4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	3	1	5	12	13	6	5	4	1
Range of scores	36-38	39-43	44-51	52-62	63-75	76-86	87-94	95-99	100-102

GROUP A MEAN SCORE - 68.5

Stanford Achievement Test

Sixth Grade Class

Word Meaning 6th - 1967



	4%	7%	12%	17%	20%	17%	12%	7%	4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	1	3	5	17	12	7	4	0	1
Range of scores	39-40	41-44	45-51	52-60	61-71	72-80	81-87	88-91	92-93

GROUP B MEAN SCORE - 61.9

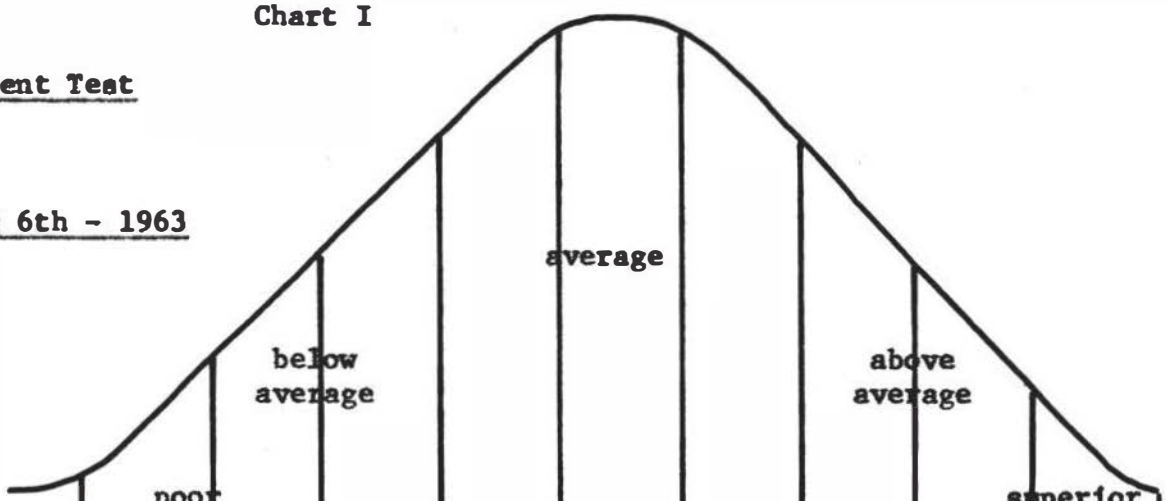
Chart H - The range of scores for Group A on this test was sixty-six points. Thirty-one students fell into the average range. Nine students were rated on the scale as below-average to poor while ten ranked in the above-average to superior range. The mean stanine scores for Group A was five. Group B had a range of scores a little below Group A. Their range was fifty-five points. Thirty-six of these fell into the average stanine range. Nine students ranked in the below-average to poor range and five ranked above-average to superior. The mean stanine score for Group B was five.

Chart I

Stanford Achievement Test

Sophomore Class

Paragraph Meaning 6th - 1963



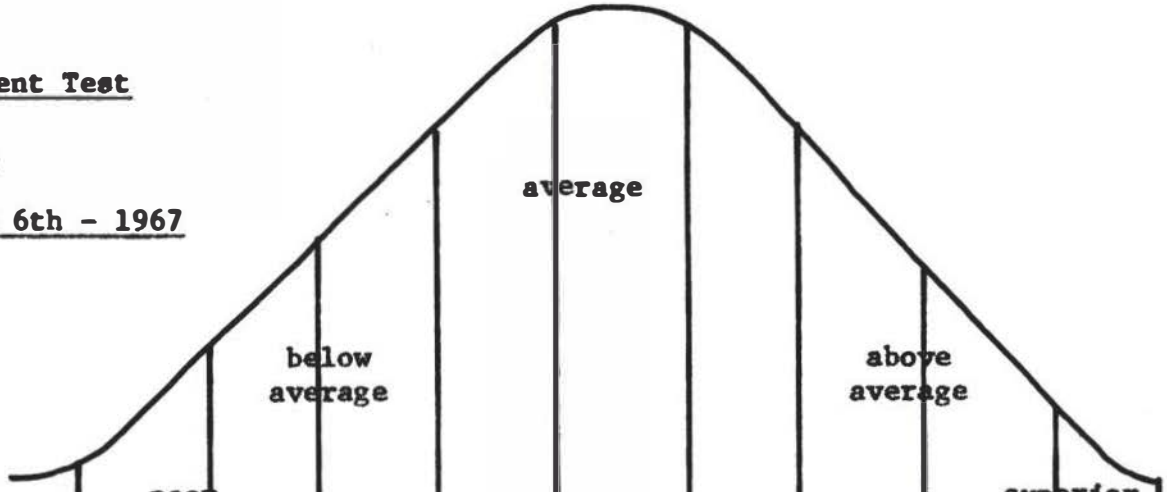
	poor 4%	7%	12%	17%	20%	17%	12%	7%	superior 4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	2	0	3	7	21	12	2	0	3
Range of scores	32-34	35-39	40-48	49-60	61-75	76-87	88-96	97-102	103-105

GROUP A MEAN SCORE - 69.1

Stanford Achievement Test

Sixth Grade Class

Paragraph Meaning 6th - 1967



	poor 4%	7%	12%	17%	20%	17%	12%	7%	superior 4%
Stanine	1	2	3	4	5	6	7	8	9
Percentile	4	11	23	40	60	77	89	96	
Number of pupils	2	0	7	15	18	4	1	2	1
Range of scores	32-34	35-40	41-49	50-62	63-78	79-91	92-100	101-106	107-11

GROUP B MEAN SCORE - 64.8

Chart I - The range of scores for Group A on this test was seventy-three. Group B had a range of seventy-eight. Forty students of Group A fell into the average stanine classification with five students below-average to poor and five students placing in the above-average to superior range. Group B had thirty-seven students in the average range with nine falling into the below-average to poor range and four in the above-average to superior range. The mean stanine for both Group A and Group B was five.

Chapter IV

SUMMARY AND CONCLUSIONS

To properly summarize and conclude this comparative study, the purpose should once again be clear so that any conclusions can be checked for congruency with that purpose. Let it be stated at this point that the reliability of a study of this nature can always be questioned and should always be questioned as to whether it proves anything. From the outset it was clearly established that the purpose of this survey was to investigate the relationships between students who have spent three years in a non-graded school situation and students who had not had the opportunity of the non-graded classroom situation. There was no attempt made to state which was better. The non-graded method of reading is in progress now and the question which involved the administrators and board members of the Livingston County Community Unit #2 was the effectiveness of their adopted program.

The sixth-grade classes and the sophomore class of 1967-68 were used in the study. All cumulative records were pulled and test scores involving intelligence and reading were pulled from the vast amount of material. The sophomore class was the smaller of the two groups, therefore it was necessary to compile all available information on this group first so that the largest possible population could be achieved. This group boiled down to fifty in number. All students with incomplete data were removed from the group. Next, the sixth grade folders were gone over with the same

purpose in mind and this group totaled fifty-six. In order that the population of the two groups would be the same, a random number were eliminated to reduce the group to fifty.

The relationship between the intelligence level of both groups was the next fact which seemed essential. How did they compare in I.Q.'s with each other? The relationship here was almost too good to be true. A mean difference of only 1.5 points existed between them. It would seem that this would aid in comparing the groups. This mean difference was based on three different intelligence test scores for each individual involved in the study.

Reading scores obtained by the students on the Stanford Achievement Test from the two sections involving reading, paragraph meaning and word meaning, were then compiled. A frequency distribution on each group of scores for each test was made and from this both groups were assigned stanine levels according to that particular group for that particular test. Mean grade equivalent scores and stanine scores were found for both groups. A comparison was then made between the two groups on each of the tests.

Specific conclusions about the results of this comparative survey are not forthcoming. There seems to be no basic difference between the two groups insofar as test results are concerned. This might lead one to conclude that no apparent progress is being made under the non-graded reading program. This might be true. On the other hand, we might also conclude that this means the non-graded reading program works just as well and is doing the job as well as before. Whether or not the children are benefitting to a greater degree because of different books used, individual differences taken into greater consideration, and grouping into homo-

geneus groups, is something that is impossible to clearly measure. This survey made no attempt other than to place before the people concerned some comparison of reading achievement to determine effectiveness. The conclusions are left to the reader.

BIBLIOGRAPHY

Bibliography

Books

Buros, Oscar K. Mental Measurements Yearbook. 6th edition, The Gryphon Press, Highland Park, New Jersey, 1965.

Gerberich, J. Raymond, Greene, Harry A., Jorgensen, Albert N. Measurement and Evaluation in the Modern School. New York: David McKay Company, Inc., 1962.

Harris, Albert J. How To Increase Reading Ability. 4th edition. New York: David McKay Company, 1961.

Peters, Herman J. Guidance In The Elementary School. Chicago: Rand McNally and Company, 1965.

Smith, Nila B. "What Have We Accomplished In Reading," Teaching Reading: Selected Materials. ed. William Barbe, New York: Oxford University Press, 1965.

Stodola, Quentin and Stardahl, Kalmer. Basic Educational Tests and Measurements. Chicago: Science Reasearch Associates, Inc., 1967.

Articles and Periodicals

DiPasquale, Vincent. "Schools Without Grades," Better Homes and Gardens, 33:28, 1955.

Durost, Walter N. "The Characteristics, Use, and Computation of Stanines," Test Service Notebook. No. 23, New York: Harcourt, Brace and World, Inc., 1961.

_____. "How To Tell Parents About Standardized Test Results," Test Service Notebook. No. 26, New York: Harcourt, Brace, and World, Inc., 1961.

Walker, Robert. "How To Understand and Use Test Results," Grade Teacher. April, 1968.

Encyclopedias and Journals

DiLorenzo, Louis T. and Salter, Ruth. "Co-operative Research on the Non-Graded Primary," Elementary School Journal. Vol. 65 (February, 1963), pp. 273-274.

Goodlad, John I. "Classroom Organization," Encyclopedia of Education Research. New York: The Macmillan Company, 1960.

Johnston, A. Montgomery. "Intellectual Segregation," Elementary School Journal. Vol. 67 (January, 1967), p. 212.

Mitchell, Blythe C. "A Comparison of the Achievement-Intelligence Relationship for Pupils with that for School Systems," The Journal of Educational Research. Vol. 57, No. 4 (December, 1963).

Williams, Wilmajean. "Academic Achievement in a Graded School and in a Non-Graded School," Elementary School Journal. Vol. 67 (December, 1966), pp. 135-139.