8-1-1954

A Study of Achievement in Language Arts of the Intermediate Grades of the Grand Forks Public Schools

Lloyd D. Volker

A STUDY OF ACHIEVEMENT IN LANGUAGE ARTS OF THE INTERMEDIATE GRADES OF THE GRAND FORKS PUBLIC SCHOOLS

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

By

Lloyd D. Volker

In Partial Fulfillment of the Requirements

for the Degree of

Master of Science In Education

August, 1954

This thesis, offered by Lloyd D. Volker as a partial fulfillment of the requirements for the Degree of Master of Science in Education, is hereby approved by the Committee under whom the work has been done.

Erich Selke
Chairman

C. J. Bjork

L. M. Martin

Darrell E. Kuder
Dean of the Graduate School

ACKNOWLEDGMENTS

The writer is grateful to Dr. Erich Selke, Professor of Education at the University of North Dakota, for his courteous and valuable assistance in writing this thesis.

The writer wishes to express his appreciation to Dr. Elroy Schroeder, Superintendent of the Grand Forks public schools, for his cooperation in securing the data for this study. The writer is indebted to him and all staff members whose work with the testing program has made the study possible.

Special thanks is extended to the writer's wife, Darel J. Volker, for her patience and understanding.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Today, more than ever before, our public schools are under attack. The main charge seems to be that they are failing to teach the academic subjects properly and with any degree of success. Many national publications have carried articles (often to stimulate sales) condemning the teaching methods and achievement of our public school system. One of the most recent examples of this is a series of articles by Howard Whitman in Colliers magazine during the past few months. Frequently overheard among friends and acquaintances are such statements as: "Children just can't spell" or "Children just can't read." These and similar remarks are usually the erroneous generalizations made by persons who are aware of certain individual failures which to them become typical. Some of the attacks on our schools come from people who may carry a grudge against education in general. Others may be employed by subversive elements to perpetuate suspicion and dissatisfaction. However, much of the criticism of public education comes from responsible, honest and conscientious citizens who sincerely believe that our schools are falling down on the job. In any event, it is up to us as teachers and administrators to answer the charges and prove, if possible, that they are false.

Purpose of the Study

One of the methods being used by educators all over the nation to refute some of these unjustifiable attacks on our educational system is that of making comparative studies of achievement test results over a period of years. For years many school systems have been using standardized tests annually. A study of scores over a period of years should provide a good index of academic progress. It is, therefore, the purpose of this study to determine by similar method whether or not the Grand Forks public elementary schools might be justifiably criticized for inferior achievement in the language arts subjects; namely, reading, English and spelling. The answer to this problem will be developed along two lines: (1) the relative achievement of grades three, four, five and six will be compared, and (2) the annual achievement of each class will be compared with the test norm.

The problem may be stated more simply in the following manner. The available data was examined to determine the answers to these questions:

1. What is the status of annual achievement in reading, English and spelling for the intermediate grades in Grand Forks? Is it lower, about the same, or higher than it used to be?

2. How does this status compare with the test norms? Are we doing as well in Grand Forks as the rest of the nation?

Source of Data

In 1934, Superintendent Elroy Schroeder appointed Clifford Sawyer, a teacher with training in this field, to be in charge of the testing program for the Grand Forks public schools. The program was especially concerned with standardized intelligence and achievement examinations. Prior to this time the usual testing had been done by individual teachers who constructed their own examinations as they were needed. This step by Superintendent Schroeder marked the beginning of the use of standardized tests on a city-wide basis.

The test selected for the elementary grades was the 1929 edition of the New Stanford Achievement Test. It was used annually until the spring of 1940 when the Modern School Achievement Test replaced it for that year only. Since that time the 1940 edition of Stanford was used for twelve years until the latest edition was adopted in 1953.

Clifford Sawyer resigned his position on July 1, 1951. Walter Loomer, Coordinator of Elementary Education, was in charge of testing for the 1951-52 term. The following year, Alfred Nehring, Principal of Winship School, was appointed to perform this function and has continued to do so up to the present time.

Practically all of the data for this study was taken from records contained in the files of the Director of the Testing Program. Some information was obtained from annual reports to the Board of Education by the Superintendent of Schools, the Coordinator of Elementary Education, and the Director of the Testing Program.

CHAPTER II

REVIEW OF OTHER ACHIEVEMENT STUDIES IN LANGUAGE ARTS

With the turn of the century came the beginning of the modern achievement and intelligence testing movement. The ensuing development of this field has provided educators with valuable tools for the measurement of pupil accomplishment. School teachers and administrators have generally made use of standardized test results to improve their instruction and curriculum, but it is difficult to find published reports of such studies. Comparative examination investigations are especially scarce in the areas of language and spelling. Much more pertinent information is available in the field of reading. This is very likely the result of the wide range of teaching methods that have been applied, and the continuous controversy it has caused throughout the nation.

A brief perusal of the history of education immediately underscores the proposition that the present assault on modern achievement in the fundamentals is another example of history repeating itself. Apparently, each generation tends to criticize each succeeding one, and the cry for the quality of the "good old days" will be forever heard. Over one hundred years

ago in Boston, the Grammar School Committee complained that pupils memorized their lessons but did not understand basic principles.[1] Today, in contrast, we often hear the lament that modern education is devoid of drill thereby causing a lack of retention. Usually there is no defense against such criticism except argument based on opinion. However, the value of comparative examinations for this purpose was put into practice by John L. Riley of Springfield, Massachusetts, in 1906. He discovered in an old attic a set of examinations which had been given in the Springfield schools sixty years earlier. Upon giving them to children in the same city, the results showed a convincing improvement by the latter group. One example was ninth grade spelling where the 1906 pupils averaged 51.2 per cent as compared to 40.6 per cent for the 1846 group.[2]

However, a more recent survey of spelling ability in western Illinois did not show such favorable results. In 1945, the Buckingham-Ayres list was given to 921 pupils in grades four through eight. Their scores were compared to those of a similar group of thirty years earlier. The results are reported in Table I. Table I indicates that the 1945 group was considerably

---

[1]Hollis L. Caswell and A. Wellesley Foshay, Education in the Elementary School, American Book Company, Chicago, 1950, pp. 223-224.

[2]C. C. Ross, Measurement in Today's Schools, Prentice-Hall, Inc., New York, 1947, p. 52.

Table I[1]

Spelling Ability as Shown by Present and Former Studies, 1945

Grade and Column of Words Spelled	Average Per Cent of Words Correctly Spelled	
	Present	Former
Grade IV		
Col. Q	18.3	42.0
Col. R	17.1	34.7
Grade V		
Col. S	29.5	42.3
Col. T	20.5	39.3
Grade VI		
Col. U	34.5	47.3
Col. V	19.6	39.3

below the others in ability to spell the words of the Buckingham-Ayres list. Years ago, the columns of work of each test were of about equal difficulty. However, this is no longer true in 1945.

Material of more ancient vintage was used recently when several children's essays written in 1845 were dictated to present day school children. The children of 1845 made four times as many mistakes in spelling even though many of the words were in more common usage at that time.[2]

[1]Calvin S. Sifferd, "A Survey of Spelling Ability," Elementary School Journal, XLVII (February, 1947), pp. 340-346.

[2]"101 Questions About Public Education," National Parent Teacher, XLVII (November, 1953), p. 8.

A cursory examination of the preceeding information seems to mean that spelling ability today is superior to that of a hundred years ago but nevertheless has declined in the past quarter century.

A recent magazine article cites several comparative studies including a brief summary of the Austin, Minnesota, Survey of 1951. In the intermediate grades, pupils of 1921 were approximately a year behind in reading achievement but were ahead of their 1951 counterparts in spelling.[1] A similar situation presented itself in Dearborn, Michigan. In 1951 youngsters were found to lead corresponding 1926 classes in reading, arithmetic, and written and oral English. Only in spelling were they lower than the former generation. The article concludes that today's grade children score as high as, and generally higher than, those of earlier days except for spelling which apparently is not up to former standards.[2]

A very thorough comparative study of reading was made in the Springfield, Missouri, public schools. This was part of a complete reading survey made in response to a local minority group which was quite critical of present day reading instruction. In 1931, the Thorndike-McCall Reading Scales had been given to

---

[1] "The Truth About Our Public Schools," Changing Times, XLIX (June, 1954), p. 9.

[2] Ibid., p. 10.

144 sixth grade pupils. The same test was administered to a group of 198 sixth graders in 1948. Such factors as age distribution, transfer students, home background, and previous test experience were taken into consideration in the evaluation of results. The scores showed that the average sixth grader of 1948 was a slightly better reader than the average sixth grader of 1931. The summary results are shown in Table II.

Table II[1]

Scores Made in 1948 and 1931 by Sixth Grade Pupils Tested on the Thorndike-McCall Reading Scale

Year	Number of Pupils	Mean Score	Standard Dev.
1948	198	23.32	6.32
1931	144	22.54	4.02

The authors believed the increase of variability of scores, as shown in Table II, probably meant that reading instruction in 1948 was more effective than in 1931, and that most sixth grade children read better in 1948 than did their predecessors. Their conclusions are stated as follows:

[1]F. H. Finch and V. W. Gillenwater, "Reading Achievement Then and Now," Elementary School Journal, XLIX (April, 1949), pp. 446-454.

> In so far as the results attained in the six schools covered by this comparison are indicative of the effectiveness of reading instruction in the Springfield school system, there is reasonably good evidence that the teaching of reading is now more successful in producing the outcomes we have measured than it was seventeen years ago... This report is based on records collected in a single school system and furnishes no basis for conclusions concerning conditions prevailing, or outcomes being achieved, in other school systems. Educational workers who are familiar with developments in the teaching of reading throughout the past four decades may see in the practices that have been introduced reasons for predicting that gains will be found when comparative studies of this type are made in other communities. Such studies might well be undertaken in other schools where records adequate for the purpose have been preserved.[1]

The above recommendation is interesting because it points out that educators realize the need for this type of study in schools throughout the country. Not only does it provide an intelligent answer to critics, as was its original purpose in Springfield, but it also provides an objective estimate of progress which may be used in planning the educational program.

Paul Witty replied to reading critics in a recent article wherein he reports on several research projects in the field. The following statements are brief summaries of some of them.

1. In 1947, Ann Crosby under the supervision of D. A. Worcester of the University of Nebraska, gave 5,106 pupils the same silent reading test that had been administered to 5,680 pupils in 1921. The pupils in 1947 scored significantly higher.

---

[1] Ibid., p. 454.

2. In 1948, William S. Gray in Grand Rapids, Michigan, gave the same oral and silent reading tests as in 1916. Notable progress has been made in comprehension, but little or no progress has been made in oral reading and the speed of silent reading.

3. In 1948, Judith I. Krugman and J. Wayne Wrightstone reported results of a city-wide testing program. Data was taken from tests given in sixth and eighth grades from 1935 and 1941, and seventh and eighth grades from 1941 and 1946. They concluded that there was no evidence in the results to substantiate the claim that reading had become poorer. The averages indicated that the reading level remained about the same, that it fluctuated close to the national norm, tending generally to be slightly above.

4. In 1949, Ernest W. Tiegs made a study of achievement in basic skills before and after 1945. Stanford and Progressive Achievement Tests in sixty communities in seven states were analyzed. He found that achievement of public school pupils is not falling. In fact, the data showed a slight though probably not statistically significant gain.[1]

A few other surveys are reported in which the investigators were primarily concerned with the immediate status of a group compared to the standardized test norms. The main purpose of these studies has been to determine relative standings with regard to corresponding levels on a national basis.

A state wide study of reading ability of 15,206 sixth grade pupils in 648 schools in Indiana was made in 1945. The reported results showed that there were great individual

---

[1]Paul Witty, "Are Children Learning to Read?" School and Society, LXXV (May 10, 1952), pp. 289-294.

differences in reading proficiency. Most groups had a difference of four or more years between the pupil testing highest and the one testing lowest, and that as much as seven years was not uncommon. However, average achievement approximated the test norms.[1]

A spelling survey was also conducted in Indiana in 1947. The Stanford Achievement Test spelling list Form H was given to the 275,000 pupils in all grades. Of the 107,747 township pupils, about forty-six per cent were retarded, twenty-six per cent normal, and twenty-eight per cent accelerated. As a general trend, the percentages of retarded cases increased at higher grade levels with a definite rapid increase at grade five.[2]

In 1949, a study of the teaching of reading was made in the white schools of Morgan County, Alabama. The results of the Iowa Silent Reading Tests indicated that the average pupils in grades four through eight were somewhat below the national norms as shown in Table III. The summary of this study expressed the belief that the pupils were not ready for fourth grade reading at the time they entered the grade. They were weak in rate, comprehension, and word meaning, and the amount of retardation increased as they progressed through school.[3]

---

[1] "A State-Wide Survey of Reading Achievement," The Elementary School Journal, XLVI (May, 1946), pp. 480-482.

[2] William H. Fox, "Spelling Proficiency in Township Schools in Indiana," The Education Digest, XII (April, 1947), pp. 5-6.

[3] Orville Wheeler, "Evaluation of Reading in Grades I-XII, in the Public Schools of Morgan County, Alabama," Peabody Journal of Education, XXVI (May, 1949), pp. 329-337.

Table III

Grade Equivalents, Grade Retardation, Average Gain per Year of Pupils in Grades IV-VIII in Word Meaning, Morgan County, Alabama

Grade	Grade Equivalent	Grade Retardation	Ave. Yearly Gain
IV	2.9	1.1	
V	3.2	1.8	0.3
VI	4.4	1.6	0.75
VII	5.4	1.6	0.83
VIII	6.0	2.0	0.77

The foregoing information clearly illustrates that while, in certain geographic locations of our country, elementary school achievement in the academic subjects may be more advanced than in others, on the whole, progress of today's pupils is superior to that of pupils of years gone by. Only in the area of spelling does there seem to be any reasonable grounds for the criticism - things aren't what they used to be.

CHAPTER III

PRESENTATION OF THE INVESTIGATION

In spite of what some critics of education may infer, it is still the purpose and function of the elementary school to teach the basic skills in the academic areas. Therefore, if the amount of academic progress, or lack of it, is to be ascertained in a given school, that school must maintain an adequate testing program. Aside from the usual measurement practices employed by teachers, the Grand Forks school system has conducted such a program for the past twenty years through the annual administration of standardized intelligence and achievement tests.

Limitation of the Study

During the first four years of standardized achievement testing in Grand Forks, compilations of results either were not made or have been lost. Such distributions for Stanford Achievement Tests are available for the school terms of 1937-38 and 1938-39, and for The Modern School Achievement Test which was given during the term of 1939-40. In these years, the tests had been conducted at the end of the first semester because of the existing policy of mid-year promotion. This policy was dropped the following year when the Stanford (1940 edition) was adopted for use near the end of the second semester. Because of the variations in tests and time of testing, this study is limited to the years 1941 through 1954.

The subjects investigated by this study are commonly referred to as the language arts. They are reading, language, and spelling. The study is restricted to the achievement in these subjects by the Grand Forks public school children in grades three, four, five and six. The medians are used.

Preliminary Information

It has been stated above that for the years considered, the standardized examinations were given near the end of the second semester. Each year they have been given approximately one month before the school term was completed. According to the test manual, this places the test norm for grade three at 3.8; for grade four at 4.8; for grade five at 5.8; and grade six at 6.8. These are the respective norms used in calculating deviations from the norm from all tables containing such information.

One other fact relative to norms is pertinent. The 1951 edition of the Stanford Achievement Test was used in the years 1953 and 1954. Consequently, the norms for the last two years have been recently established and provide a more up-to-date standard of national achievement.

When the class medians of two or more groups are compared, it is important that the groups are of reasonably similar quality. None of the class medians presented in this study are that of any synthetic homogeneous group. Rather, each class is a normal heterogeneous group representing the usual distribution

of abilities and backgrounds. It is conceivable that any given class could contain a relatively large number of either bright or dull pupils. However, no correlating study of intelligence is made, and for the purposes of this investigation it is assumed that the classes of a given grade are closely similar in the ability to achieve.

Relative class size is another important factor to consider when the median of one group is compared to that of another. It is difficult to make a valid comparison between a small class and a large one. This could also be true of very small classes of similar size. Although such classes were heterogeneous in composition, their small numbers could easily preclude a normal distribution of abilities.

An examination of Table IV reveals the annual class size of each grade is approximately 250 pupils until the increased birth rate of the war years altered conditions. This is evident in the table beginning in grade three of 1949. The succeeding enrollments show an approximate average of 300 pupils per grade. However, in spite of the sudden increase in enrollment in 1949 and following, the difference in class size is assumed not to be enough to have a significant influence on the comparative study of achievement. This assumption is based on the belief that the smallest class in Table IV is of sufficient size to provide a normal distribution of achievement.

Table IV

Annual Class Size by Grade

Year	Number of Pupils Grades 3	4	5	6
1954	298	286	319	296
1953	282	312	298	310
1952	327	302	318	306
1951	313	307	323	267
1950	320	320	258	259
1949	329	250	251	213
1948	253	255	212	256
1947	263	216	273	248
1946	230	259	251	266
1945	262	244	259	254
1944	237	249	253	256
1943	241	240	238	246
1942	249	251	251	272
1941	278	248	278	245

Reading

Reading makes use of several basic skills. A good reader must first of all be able to comprehend what he reads. To do this he must have an adequate vocabulary. He should be able to

attack new words with the aid of context clues, phonics, or analysis techniques. He must be able to read rapidly. If he reads to others he should do so clearly and with expression. In the past, more emphasis was placed on oral reading, and it was usually a part of the reading test. Today, the greater emphasis is on comprehension, and standardized achievement tests usually omit oral reading. This is probably due in a large extent to the difficulty of setting up usable standards of evaluation, plus the fact that oral tests must be individually administered.

The Stanford Achievement Test Battery divides the reading test into two parts; paragraph meaning (comprehension), and word meaning (vocabulary). Each part has a specified time limit, but this is largely for the purpose of standardization. The amount of time is adequate for all but the exceptionally slow pupil to complete the test. This fact is true for all of the parts of the Stanford battery. It is basically a power test.

Paragraph Meaning

Table V shows annual median grade equivalents and the deviations from the test norm in paragraph meaning for grades three, four, five and six.

In 1941, the median grade equivalent for grade three was 3.8 which was identical to the test norm. This was the lowest

Table V

Paragraph Meaning - Grades III, IV, V, VI
Median Grade Equivalents and Deviations from Norms

Year	Grade Equivalent Grade				Deviation Grade			
	3	4	5	6	3	4	5	6
1954	3.9	4.9	5.9	6.6	.1	.1	.1	-.2
1953	4.0	5.2	6.0	6.6	.2	.4	.2	-.2
1952	3.9	4.8	5.6	7.4	.1	.0	-.2	.6
1951	3.9	4.8	5.6	7.4	.1	.0	-.2	.6
1950	4.0	4.8	5.8	7.0	.2	.0	.0	.2
1949	3.9	4.8	5.8	6.8	.1	.0	.0	.0
1948	4.3	4.8	5.9	7.4	.5	.0	.1	.6
1947	4.2	4.6	6.0	7.0	.4	-.2	.2	.2
1946	4.0	4.9	5.6	7.0	.2	.1	-.2	.2
1945	4.3	4.6	5.6	7.2	.5	-.2	-.2	.4
1944	4.0	4.8	5.9	7.0	.2	.0	.1	.2
1943	3.8	4.6	5.6	7.2	.0	-.2	-.2	.4
1942	3.9	4.8	6.0	7.0	.1	.0	.2	.2
1941	3.8	4.8	5.2	6.8	.0	.0	-.6	.0

grade equivalent recorded in grade three. The highest is 4.5 which occurred in 1945. In 1954, it was 3.9. The medians tend to fluctuate with a slight trend to increase annually until 1949.

Here there is a drop followed by a leveling off with achievement remaining a tenth or two above 1941. The noticeable drop in 1949 coincides with the increase in third grade enrollment that year as shown in Table IV. However, there is no corresponding trend or drop in grade four.

The 1941 median for grade four is 4.8. It drops to 4.6 in the years of 1943, 1945 and 1947. In 1948, it is again 4.8 which it maintains the four succeeding years, rising to its peak of 5.2 in 1952. The 1954 median is 4.9. During the last seven years the fourth grade held steady with a slight increase in paragraph meaning achievement the last two.

Grade five has a low median grade equivalent of 5.2 in 1941. Its best mark of 6.0 was reached in 1942, 1947 and 1953. The 1954 median was 5.9. There is no discernible trend toward either poorer or better achievement.

Fluctuation of achievement in grade six is similar to that of grade five. The 1941 median was 6.8. It rises and falls slightly, indicating no particular trend over the years, but shows a noticeable drop in the low of 6.6 in 1953 and 1954.

By examining the deviations from the test norms as shown in Table V, grade three stands out as being consistently above the national norm. Grade four does quite well, dropping below the norm in only three of the fourteen years. Grade five makes

the poorest showing by being below the norm six different years. It is six-tenths of a grade below in 1941. Grade six has usually achieved very well ranging up to six-tenths above the norm in 1951 and 1952. The last two years are the only ones in which grade six dropped below the norm.

Summary of Paragraph Meaning

In general, the achievement in paragraph meaning for grades three, four, five and six of the Grand Forks public schools shows a normal amount of fluctuation from year to year. There is little evidence of any general trend up or down. However, grades four, five and six have a somewhat better record the last seven years than the first seven. Annual deviations from the norm in each grade show that the Grand Forks medians are more often slightly above the national norms.

Word Meaning

Table VI shows the median grade equivalents and deviations from test norms for grades three, four, five and six in Grand Forks public schools.

In 1941, the median grade equivalent for grade three in word meaning was 3.4. This is the lowest median for the period. Thereafter it increases gradually until it reaches 4.1 which is maintained from 1952 through 1954. This clearly indicates a trend of better achievement in vocabulary over the years.

Table VI

Word Meaning - Grades III, IV, V, VI
Median Grade Equivalents and Deviations From Norms

Year	Grade Equivalents Grade 3	4	5	6	Deviations Grade 3	4	5	6
1954	4.1	5.2	6.1	7.0	.3	.4	.3	.2
1953	4.1	5.1	6.1	6.7	.3	.3	.3	-.1
1952	4.1	4.8	5.4	6.8	.3	.0	-.4	.0
1951	4.0	4.8	5.8	7.0	.2	.0	.0	.2
1950	3.9	4.9	5.8	7.0	.1	.1	.0	.2
1949	3.9	4.8	5.6	6.6	.1	.0	-.2	-.2
1948	3.9	4.9	5.6	7.2	.1	.1	-.2	.4
1947	3.9	4.5	5.6	6.4	.1	-.2	-.2	-.4
1946	3.9	4.7	5.4	6.6	.1	-.1	-.4	-.2
1945	4.0	4.4	5.6	6.8	.2	-.4	-.2	.0
1944	3.8	4.4	5.6	6.8	.0	-.4	-.2	.0
1943	3.8	4.4	5.1	7.2	.0	-.4	-.7	.4
1942	3.8	4.5	5.8	6.6	.0	-.3	.0	-.2
1941	3.4	4.5	5.2	6.4	-.4	-.3	-.6	-.4

Grade four varies somewhat more than grade three. The 1941 median was 4.5. The low was 4.4 in 1943, and the peak was 5.2 in 1954. A rather steady gain is noted in the last seven years.

The 1941 fifth grade median was 5.2. The low of 5.1 appears in 1943 and the high of 6.1 in 1953 and 1954. At first glance this seems to show present achievement at a level one full year above what it was a decade ago. However, this generalization is not true for the 1942 median was 5.8. Nevertheless, Table VI shows a decided gain for grade five, especially during the last five years.

Grade six shows a median of 6.4 in 1941 and 7.0 in 1954. This is relatively the picture. With some fluctuation there is indication of a slight increase in word meaning particularly over the last five years.

The deviations from the norms in Table VI show that except for 1941, grade three was consistently above the national norm. Its best achievement was during the last five years. Grade four annually was below the norm from 1941 through 1947. From 1948 on, it was at or above the norm, making its best rating in 1954. For nine of the fourteen years, grade five was below the norm. It was above the norm in 1953 and 1954. Grade six is more consistent in approximating the test norm than either grades four or five. Its medians over the years vary above and below the norm but have been below only once in the last five years.

Summary of Word Meaning

In general, the level of achievement in word meaning by grades three, four, five and six of the Grand Forks public

schools has risen in recent years. Before 1950, medians were usually below the national norm. However, since 1950, achievement has been consistently above the national average.

Summary of Reading

The complete picture of reading was obtained by combining the information from Tables V and VI.

In grade three, gain in both paragraph meaning and word meaning was made during the fourteen years. However, in the more recent years, paragraph meaning has leveled off while word meaning achievement has continued a gradual climb. Reading achievement in grade three is slightly better today than it was fourteen years ago and is slightly above the national average as compared with the test norm.

Grade four is relatively consistent in maintaining normal annual progress in paragraph meaning. In recent years the gain in achievement in word meaning is quite marked. Reading achievement in grade four definitely is better today than it was in the early years of this study. In comparison with test norms, Grand Forks' fourth graders are now achieving slightly better than the average for the nation.

Annual paragraph meaning achievement in grade five has remained about the same over the years. However, notable progress has been made in word meaning during recent years. Fifth

graders in Grand Forks are doing better reading today. Although the earlier years of this study show that fifth grade medians are very often below the norm, the more recent years indicate that the average fifth graders in Grand Forks are now reading as well, or slightly better, than the national average.

Annual achievement in grade six has fluctuated over the years showing a present decrease in paragraph meaning and an increase in word meaning. The present Grand Forks' sixth graders are probably neither much better nor much poorer readers than those of the 1940's. Grade six has been slightly below the national norm the last two years. However, during most of the years included in this survey they have been at or above the national average.

Language

The teaching of language in the elementary school has many ramifications. However, the fundamental aim in language instruction is to develop the pupil's ability to communicate with others. To do this well, he must learn to understand and apply the mechanics of oral and written expression. The Stanford Achievement Test in languages measures such ability by including items in capitalization, punctuation, sentence sense, and word usage. However, the primary battery used in Grand Forks from 1941 through 1952 did not include a language test. Consequently, Table VII does not contain median grade equivalents for grade three in those years.

Table VII

Language - Grades III, IV, V, VI
Median Grade Equivalents and Deviations from Norms

Year	Grade Equivalents Grade 3	4	5	6	Deviations Grade 3	4	5	6
1954	3.9	5.1	5.4	7.0	.1	.3	-.4	.2
1953	4.2	5.0	6.1	6.7	.4	.2	.3	-.1
1952		4.9	5.6	7.0		.1	-.2	.2
1951		4.7	5.8	7.2		.2	.0	.4
1950		4.8	5.9	7.0		.0	.1	.2
1949		4.6	5.6	6.4		-.2	-.2	-.4
1948		4.8	5.9	7.4		.0	.1	.6
1947		4.2	6.0	6.4		-.6	.2	-.4
1946		4.8	5.6	7.0		.0	-.2	.2
1945		4.7	5.9	7.0		-.1	.1	.2
1944		4.7	5.9	6.4		-.1	.1	-.4
1943		4.6	5.4	7.2		-.2	-.4	.4
1942		4.6	5.6	7.0		-.2	-.2	.2
1941		4.7	5.5	6.6		-.2	-.4	-.2

Table VII shows the median grade equivalents and deviations from the norm in language for grades three, four, five and six.

The median for grade three was 4.2 in 1953. In 1954, it dropped to 3.9. However, achievement for both years was above the national norm.

In 1941, the median for grade four was 4.7. It was 5.1 in 1954. There is some fluctuation in the intervening years with a low of 4.2 in 1947. Achievement in recent years is clearly superior to that of former years.

The fifth grade median in 1941 was 5.5. The high of 6.1 was reached in 1953, followed in 1954 with the low of 5.4. The evidence indicates that fifth grade achievement over the years tends to remain about the same.

In 1941, the median for grade six was 6.6. The 1954 mark of 7.0 was also attained in 1942, 1945, 1946, 1950 and 1952. The high of 7.4 was set in 1948, and the low of 6.4 was made in 1944 and 1949. This information shows that present language achievement in grade six is approximately the same as it has been in past years.

In comparison with national norms, grade three has been above the norm both years in which it was included in the test battery. During the early years, as shown in Table VII, the fourth grade medians fell below national standards. But the past four years they have been above the norm, making their best comparative achievement in 1954. Although the fifth grade median was four-tenths of a grade below the norm in 1941 and 1954, it was at, or above, the norm in seven of the fourteen years. Grade six has made relatively good achievement, being above the norm nine times.

Summary of Language

It has been only during the last two years of this study that the Stanford Achievement Test of language was available for grade three. However, in both years, achievement was above the national norm. Grade four has made progress and is achieving better today than during the early 1940's. Fifth graders today are doing as well as they have in past years. The fifth grade medians fluctuate but, as a whole, achievement approximates the norm. There has been little change over the years in sixth grade achievement, but it has usually been slightly above the norm.

Spelling

The Stanford Achievement Test in spelling for the years 1941 through 1952 consisted of a list of fifty words dictated by the teacher. The newer edition of the Stanford Achievement Test in spelling has a similar test for grades three and four. However, the test for grades five and six consists of seventy-two multiple-choice questions in which the pupil chooses the correct spelling from three words or marks "N G" if the spelling is not given. The test manual states that results from this method correlate substantially with results from the dictation-type test.*

*See appendix. Directions for Administering Intermediate and Advanced Batteries.

Table VIII shows the spelling median grade equivalents and deviations from the norms for grades three, four, five and six.

In 1941, the median for grade three was 3.8. The highest median was 4.2 in 1954, and the lowest was 3.4 in 1942. A slight amount of fluctuation occurs over the years but, as a whole, achievement has remained approximately the same.

The median for grade four was 4.8 in 1941 and 5.3 in 1954. The low of 4.6 occurred in 1942, 1943 and 1945. Table VIII shows a slight improvement for the past seven years over the earlier years of this study.

The fifth grade median in 1941 was 5.4. In 1954, it was 6.0. The high was 6.2 in 1943, and the low was 5.2 in 1946. As a whole, there is little evidence to indicate any trend toward improved achievement.

Table VIII shows grade six has been most inconsistent in spelling achievement. The medians range between the low of 5.4 in 1948 and the high of 7.6 in 1953. The median was 6.4 in 1941 and 6.9 in 1954. Achievement in 1953 and 1954 was a decided improvement over achievement in any of the preceeding years.

In comparison with national norms, grade three has been at or above the norm every year except 1942. It was four-tenths of a grade above in 1954. Grade four also has usually done quite well. It has been below the norm only six of the fourteen years, attaining five-tenths of a grade above in 1954. Grade five

Table VIII

Spelling - Grades III, IV, V, VI
Median Grade Equivalents and Deviations From Norms

Year	Grade Equivalents Grade 3	4	5	6	Deviations Grade 3	4	5	6
1954	4.2	5.3	6.0	6.9	.4	.5	.2	.1
1953	4.0	4.7	6.0	7.6	.2	-.1	.2	.8
1952	3.8	4.8	5.4	6.4	.0	.0	-.4	-.4
1951	3.9	4.8	5.4	6.4	.1	.0	-.4	-.4
1950	4.0	4.8	5.8	6.4	.2	.0	.0	-.4
1949	3.8	4.8	5.5	6.2	.0	.0	-.3	-.6
1948	3.9	4.8	5.2	5.4	.1	.0	-.6	-1.4
1947	3.9	4.7	5.6	5.8	.1	-.1	-.2	-1.0
1946	4.1	4.8	5.2	6.2	.3	.0	-.6	-.6
1945	4.1	4.6	5.6	6.4	.3	-.2	-.2	-.4
1944	3.9	4.7	5.2	5.9	.1	-.1	-.6	-.9
1943	3.9	4.6	6.2	6.6	.1	-.2	.4	-.2
1942	3.4	4.6	5.6	6.4	-.4	-.2	-.2	-.4
1941	3.8	4.8	5.4	6.4	.0	.0	-.4	-.4

was two-tenths of a grade above the norm in 1953 and 1954. But, in ten of the fourteen years it was under the norm, falling six-tenths below in 1944, 1946 and 1948. Over the years, grade six

has compared poorly with national achievement. It has been below the norm in twelve of the fourteen years. In 1948, it fell one and four-tenths grades below. Only in the last two years has it been above. In 1953, it was eight-tenths higher than the norm.

Summary of Spelling

Achievement in spelling in grade three has remained relatively constant over the years. This achievement shows a tendency to be slightly above the national norm. Grade four shows a slight gain in achievement during recent years. During these years, achievement also rose to approximate the test norm. In the last two years, grade five has made noticeable gain. Achievement in these two years was above the norm, whereas before, it was consistently below. Grade six has shown a sudden increase the last two years. Prior to 1953, medians in grade six varied from year to year, but the difference was relatively small. Sixth grade achievement surpassed national norms only in 1953 and 1954.

CHAPTER IV

SUMMARY AND CONCLUSIONS

The Stanford Achievement Test battery has been given annually to elementary pupils in the Grand Forks public schools from 1941 through 1954. The achievement in language arts in the intermediate grades, as determined by this study, is based on median grade equivalents and deviations from the norms. For the purposes of this investigation, the quality of the groups studied have been assumed to be relatively the same. This assumption is based on two facts: (1) each class represents a heterogeneous group, and (2) each class is composed of over 200 pupils, which should provide a normal distribution of abilities. Reading achievement is based on test results in both paragraph and word meaning.

Summary of Findings

The findings of this investigation are summarized by grade as follows:

Grade III

Reading achievement is slightly better today than it was fourteen years ago. It is also slightly better than the national average.

Standardized language tests have included grade three only the last two years. Achievement in these years is above the national norm.

Spelling achievement over the years has remained relatively the same. It is, and usually has been, above the norm.

Grade IV

Reading achievement in recent years shows a definite increase. The present level is above the national norm.

Language achievement shows decided growth and is presently above the norm.

Growth in spelling achievement is slight but evident. In recent years, it has held steadily at the norm and was five-tenths of a grade above in 1954.

Grade V

Reading achievement is better today with more improvement in word meaning. Present achievement is slightly above the national norm.

Language achievement varies a good deal. It is about the same today as in past years. As a whole, this indicates a level approximating the national. In 1954, however, it was four-tenths of a grade below the norm.

Spelling achievement has risen noticeably in the last two years. It is now above the norm, although it was consistently below in earlier years.

Grade VI

Reading achievement has fluctuated over the years. There was a sudden drop in paragraph meaning the last two years. This was partially countered by a rise in word meaning in 1954. As a whole, there is insufficient evidence to indicate a trend, and achievement today is above the same as the national average.

Language achievement varies but, over the years, there is little change. It is and has usually been above the norm.

Spelling achievement has varied unusually. The largest gain was made in 1953. It has been above the national norm only in 1953 and 1954.

Conclusions

The purpose of this study has been to examine achievement in language arts in grades three, four, five and six of the Grand Forks public schools to determine: (1) how present achievement in these areas compares with that of former years, and (2) how achievement compares with national norms.

The conclusions based on the results of this study are:

1. Reading achievement is slightly better today than it was in the 1940's. It has tended to be near or above the national norms and is presently slightly above.

2. Language achievement is presently about the same as it has been over the years. As a whole, it has remained near the national norms.

3. Spelling achievement today shows an increase over former years to a level above the national norms. However, it is likely that the spelling test in the newer Stanford Achievement Test battery adopted in 1953 was an important factor in this progress. This contention is based on the fact that medians in grades five and six have usually fallen below national norms.

4. For the years included in this study, Grand Forks public elementary schools cannot be justifiably criticized for failing to maintain and improve standards in the language arts.

Recommendations

Grand Forks is one of the larger school systems in the state of North Dakota. This fact immediately places it in a position of educational leadership, leadership which is expected through its ability to provide superior administrators, teachers, facilities and equipment. As a leader, Grand Forks cannot be satisfied with slight or normative achievement but must be constantly striving for superior results. Therefore, in spite of

the conclusion that the Grand Forks public elementary school has maintained and improved achievement in the language arts, there is need for further appraisal and analysis of the findings of this investigation.

The conclusions are based on a study of annual median grade equivalents which, over the years, provide an indication of the trends in achievement. This points out, generally, areas of strength or weakness, but it does not provide a complete picture of annual achievement. Achievement testing should do more than indicate the level or direction of attainment. It should also be a tool used by the administrator and teacher for the diagnosis of individual deficiencies. One author points out the need for recognition of this purpose as follows:

> Measurement, to be of maximum value should have as its fundamental purpose the better understanding of the individual pupil and his progress. There is serious danger that this primary purpose is defeated when too much emphasis is placed on average scores, total scores, or composite scores... It is indeed unfortunate that administrators or teachers sometimes complacently accept the fact that the average score of a certain group compares favorably with a national norm on a standardized test as evidence that the school is satisfactorily accomplishing its purpose.[1]

Although the foregoing discussion is primarily intended as a caution in the use of average scores, it applies equally well in the interpretation of median scores. Rausch goes on to conclude:

---

[1]Oscar P. Rausch, "Average Scores - Use with Caution," Elementary School Journal, Vol. XLVII (December, 1947), p. 225.

> Administrators and teachers must, therefore, analyze the results of measurements as they pertain to individual pupils in individual areas in order to obtain the maximum diagnostic and instructional value from the tests.[1]

A more accurate picture of class achievement is obtained by determining the relative number of pupils falling above or below the norm, the positions of the high and low scores, and any concentration in the distribution of scores. In the final analysis, the teacher should scrutinize the individual's scores in relation to his ability to achieve, his relative standing in the group, and his achievement in comparison to the test norm. In order to improve the use of achievement tests for diagnostic purposes the following recommendations are made:

1. Either give standardized tests near the end of both semesters or change the date of current practice from the end of the school year to the end of the first semester. This will provide the teacher with a better understanding of her class and its individual members when she will have ample remaining time to concentrate on discovered weaknesses. It is true that this information can be obtained from the test results of the preceeding year, but such information is relatively cold and is too easily ignored by teachers in fulfilling their regular busy daily schedule. Through mid-year testing the pupils, as well as the

---

[1] Ibid., p. 228.

teacher, can evaluate the results of their own efforts during the first semester, and they will be more enthusiastic about improving the quality of their work. In effect, mid-year achievement testing would provide an opportune stimulus to both teacher and pupils to attack individual shortcomings.

2. Expand the analysis of annual test results by the Principal's Council, utilizing the breakdown of scores which includes an indication of the achievement level attained in each subject by the largest number of pupils. This study should contain a regular appraisal of all contributing factors toward test results.

3. Establish a policy of staff meetings within the various schools for the purpose of cooperative analysis of achievement results for the individual buildings with emphasis on ways and means for improvement. Such meetings should strengthen instruction and morale through the mutual exchange of ideas and a better understanding of one another's problems.

Many factors influence achievement. It is believed that the following more tangible suggestions will assist progress in the language arts.

1. Existing library facilities should be improved as much as possible. A well supplied central library under the direction

of a trained librarian is desirable in every school. Although this goal is not immediately accessible in every case, much can be done by increasing the quality and quantity of present library resources.

2. In 1950, the Grand Forks schools held a city-wide spelling contest with all elementary children participating. This stimulated an increased interest and industry in spelling which was reflected that spring by a noticeable rise in spelling achievement scores. Since this study shows a lagging in spelling over the years, it is recommended that an annual spelling contest be held each spring in the elementary grades as one method of directing emphasis in that area. Details of such a contest should be worked out by a committee of teachers and administrators.

Many other suggestions will be forthcoming from all staff members as the study of achievement is carried on. It is hoped that as a result, when the present time becomes the "good old days" the evidence will be available to show that progress in education is continuous.

BIBLIOGRAPHY

Books

Caswell, Hollis L, and Foshay, A. Wellesley, Education in the Elementary School, Chicago: American Book Company, 1950, pp. 223-224.

Ross, C. C., Measurement in Today's Schools, New York: Prentice-Hall, 1947, p. 52.

Periodicals

"A State-Wide Study of Reading Achievement," The Elementary School Journal, Vol. XLVI (May, 1946), pp. 480-482.

Finch, F. H., and Gillenwater, V. W., "Reading Achievement Then and Now," Elementary School Journal, Vol. XLIX (April, 1949), pp. 446-454.

Fox, William H., "Spelling Proficiency in Township Schools in Indiana," The Education Digest, Vol. XII (April, 1947), pp. 5-6.

"101 Questions About Public Education," National Parent Teacher, Vol. XLVII (November, 1953), p. 8.

Rausch, Oscar P., "Average Scores - Use with Caution," Elementary School Journal, Vol. XLII (December, 1947), pp. 225-228.

Sifferd, Calvin S., "A Survey of Spelling Ability," Elementary School Journal, Vol. XLVII (February, 1947), pp. 340-346.

"The Truth About Our Public Schools," Changing Times, Vol. XLIX (June, 1954), p. 9.

Wheeler, Orville, "Evaluation of Reading in Grades I-XII, in the Public Schools of Morgan County, Alabama," Peabody Journal of Education, Vol. XXVI (May, 1949), pp. 329-337.

Witty, Paul, "Are Children Learning to Read?", School and Society, Vol. LXXV (May 10, 1952), pp. 289-294.

APPENDIX

TEST 4 *Language* (Continued) ◀8a

DIRECTIONS: In each sentence, decide which of the numbered words is correct. Then mark the answer space at the right which has the same number as the word you have chosen.

SAMPLE: Apples 1 is / 2 are good. 1 2

Soon it 1 began / 2 begun to rain. 1 2 35

I 3 said / 4 says to Jim, "Just try it." 3 4 36

5 My father he / 6 My father told me to come. 5 6 37

My little sister 1 seen / 2 saw a bear. 1 2 38

Yesterday Jack 3 came / 4 come home early 3 4 39

I 5 brung / 6 brought my lunch today. 5 6 40

Miss Brown 1 sat / 2 set over there. 1 2 41

I used to 3 could / 4 be able to sing better. 3 4 42

Sam 5 ain't / 6 isn't here today. 5 6 43

1 Let / 2 Leave me have a turn now. 1 2 44

Bob and 3 I / 4 me painted the scenery. 3 4 45

Give Tom 5 that there / 6 that sandwich. 5 6 46

Ned wants to do it 1 hisself. / 2 himself. 1 2 47

Will you 3 teach / 4 learn me to jump rope? 3 4 48

Sally 5 drawed / 6 drew a picture of a cow. 5 6 49

Don't you want 1 no more / 2 any more ice cream? 1 2 50

I 3 broke / 4 busted my fishing pole. 3 4 51

Jane 5 swam / 6 swum across the pool. 5 6 52

Please 1 take / 2 bring this note to your mother. 1 2 53

Three of 3 we / 4 us boys got caught. 3

Stand 5 hear / 6 here beside me. 5

We all 1 clumb / 2 climbed over the fence. 1

John didn't give us 3 any / 4 no paper. 3

Mary has 5 gone / 6 went to the park. 5

A boy 1 doesn't / 2 don't like to sit still. 1

Where did you buy 3 them / 4 those socks? 3

Where 5 are / 6 is my books? 5

My mother should 1 have / 2 of told me. 1

I've 3 did / 4 done my arithmetic. 3

There 5 were / 6 was five cookies in the jar. 5

My aunt gave me 1 a / 2 an apple. 1

The children have done 3 there / 4 their jobs. 3

Sit down and rest 5 your / 6 you're feet. 5

All of us wanted to go 1 bad. / 2 badly. 1

Nancy can certainly read 3 good. / 4 well. 3

The grass has 5 grown / 6 growed an inch. 5

1 Hadn't you ought to / 2 Shouldn't you use a broom? 1

We have already 3 choosed / 4 chosen sides. 3

Have you 5 written / 6 wrote to Helen? 5

Our team will win this game 1 easy. / 2 easily. 1

Stop.

No. right ( ) × 2 (

No. omitted or double marked (

Sum (

Subtract

DIFFERENCE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Gr. score	below 10									10	11	12	13	14	15	15	16	17	18	19	20	21	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	

DIFFERENCE (Cont'd)	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
Gr. score	39	40	41	42	43	45	46	47	49	50	52	53	55	57	58	60	62	64	65	67	69	71	73	74	76	78	80	83	85	88	92	97	102	108

DIFFERENCE ____

List of Words (Continued)

31.	berries	Come with us to pick ***berries.***	**berries**
32.	dirt	Wash the ***dirt*** from your hands.	**dirt**
33.	feeling	Betty was ***feeling*** better at noon.	**feeling**
34.	narrow	The space is very ***narrow.***	**narrow**
35.	laugh	The funny clowns make us ***laugh.***	**laugh**
36.	easy	The work is ***easy*** today.	**easy**
37.	boxes	The toys were packed in ***boxes.***	**boxes**
38.	else	Who ***else*** will stay?	**else**
39.	seemed	It was not really long, but it ***seemed*** long.	**seemed**
40.	family	There are three children in the ***family.***	**family**
41.	explore	How would you like to ***explore*** a cave?	**explore**
42.	elephant	Jumbo was a circus ***elephant.***	**elephant**
43.	dried	The clothes ***dried*** in the warm air.	**dried**
44.	package	He carried the ***package.***	**package**
45.	grocery	Food is sold at the ***grocery*** store.	**grocery**
46.	potatoes	We had baked ***potatoes*** for dinner.	**potatoes**
47.	touch	Don't ***touch*** the hot iron.	**touch**
48.	geography	We learn about other countries in ***geography.***	**geography**
49.	settled	The sand ***settled*** to the bottom.	**settled**
50.	several	She read ***several*** pages.	**several**

TEST 3 *Spelling*

1. _____ 26. _____
2. _____ 27. _____
3. _____ 28. _____
4. _____ 29. _____
5. _____ 30. _____
6. _____ 31. _____
7. _____ 32. _____
8. _____ 33. _____
9. _____ 34. _____
10. _____ 35. _____
11. _____ 36. _____
12. _____ 37. _____
13. _____ 38. _____
14. _____ 39. _____
15. _____ 40. _____
16. _____ 41. _____
17. _____ 42. _____
18. _____ 43. _____
19. _____ 44. _____
20. _____ 45. _____
21. _____ 46. _____
22. _____ 47. _____
23. _____ 48. _____
24. _____ 49. _____
25. _____ 50. _____

No. RIGHT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Gr. score	14	15	16	17	18	19	20	21	21	22	23	24	25	26	27	28	29	29	30	31	32	32	33	34	35	35	36	37	38	38	39	40	41	42	43	44	45	46	47	49	50	52	54	55	57	59	61	63	65	68

TEST 1 *Paragraph Meaning* (Continued)

Here is the way to lay a brick walk in a garden. Dig a path 4 inches deep. Pack and roll down 2 inches of sand. Lay in place __37__ $2\frac{1}{2}$ inches thick. Your finished walk will be just a little __38__ ground level.

37. **cement** **boards** **bricks** **dirt**
38. **above** **below** **nearer** **beneath**

When we become angry or afraid, our hearts begin to beat rapidly. Our muscles feel tight. Our bodies get ready to fight or run, even though we do not really need to do either. Afterward, we feel as tired as though we had actually __39__ or __40__.

39. **slept** **eaten** **run** **awakened**
40. **rested** **fought** **slept** **read**

Wool is clipped from live sheep by a process called shearing. The entire mat of fleece from each animal comes off in one piece. With electric clippers one man can __41__ from 150 to 200 __42__ a day. After shearing, the __43__ is rolled up in bundles and sent to the mill.

41. **clip** **run** **kill** **feed**
42. **pounds** **lambs** **pelts** **sheep**
43. **skin** **hide** **fleece** **cotton**

A bottle used to be made by a glass blower with a long pipe through which he blew air into a bubble of hot liquid glass. Now the work is done by a machine which revolves over a pot of melted __44__, sucks up the amount needed, shapes it on a mold, and blows it out. A workman operating a __45__ can produce ten times as many __46__ in an hour as an old-fashioned glass blower could.

44. **metal** **iron** **glass** **ice**
45. **blower** **machine** **factory** **pipe**
46. **pipes** **balls** **bottles** **glasses**

A few years ago most freight was carried by railroad trains. Now such things as furniture and automobiles are sent across country on trucks. Goods sent by __47__ can go only where __48__ have been laid, but goods sent by __49__ can reach any point to which a __50__ runs.

47. **truck** **rail** **freight** **express**
48. **roads** **paths** **tracks** **highways**
49. **truck** **freight** **rail** **express**
50. **drive** **trail** **track** **road**

Stop.

No. RIGHT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Gr. score	below 10			10	12	14	15	16	17	18	20	21	22	23	24	25	26	27	28	29	30	30	31	32	33	34	35	36	37	38	39	40	42	43	44	46	48	50	52	55	57	60	63	67	71	76	81	86	93	99	

TEST 1 *Paragraph Meaning*

DIRECTIONS: Find the word that belongs in each space, and draw a line under it. Do *not* write in the spaces.

SAMPLE:

Wheat grows on farms. Most bread is made from wheat. If farmers did not plant __51__, most people would have no __52__ to eat.

51. **corn** **potatoes** **rice** **wheat**

52. **oranges** **bread** **carrots** **eggs**

Mary and John live in a big __1__.

1. **tree** **house** **farm** **yard**

See them laugh.
Something is __2__.

2. **funny** **red** **big** **out**

Frank wanted to go out to play, but his mother said it was too wet outdoors. Frank looked out the window and saw that his mother was right. The __3__ was falling fast.

3. **night** **rain** **storm** **cold**

The little boy can throw a ball, but he cannot __4__ it.

4. **make** **catch** **swing** **eat**

We have a small pony.
We always try not to __5__ it.

5. **ride** **see** **hurt** **feed**

Helen was sick. The girls at school wrote her a letter. "Dear Helen," they said, "We hope you will soon feel __6__ enough to come back to __7__."

6. **well** **happy** **nice** **glad**

7. **church** **visit** **school** **town**

Mother frogs lay their eggs in the wat
The __8__ hatch into tiny tadpoles that c
breathe under the __9__ the way fish do.

8. **frogs** **toads** **eggs** **animals**

9. **rocks** **water** **neck** **body**

The children went to the circus. Th
saw elephants, monkeys, and many oth
animals. There were many clowns and l
of popcorn and peanuts. The children sa
that they wished a __10__ would come eve
day.

10. **parade** **clown** **circus** **monkey**

You can often find shells along the edges
rivers and lakes. An even better place
pick up __11__ is by the ocean.

11. **seaweed** **shells** **rocks** **sand**

Tom and Jane had for a pet a white mou
called Mickey. The children were fond
Mickey and took him on their vacation tri
They both took care of him. It was Tom
job to keep the cage nice and clean, and it w
__12__ duty to see that the __13__ got plen
of the right kind of food.

12. **his** **their** **Mickey's** **Jane's**

13. **mouse** **children** **mice** **kitten**

When Mary was ten years old, she w
given ten cents a week. Her brother To
who was twelve, got twenty-five cents a we
Mary asked her father why she could not ha
as much as Tom. Her father replied, "Wh
you are as old as Tom is now, you may ha
just as much as he gets now." Two ye
later, when Mary reached her __14__ birthd
her father said, "Now you may have __15__
cents a __16__."

14. **next** **tenth** **eleventh** **twelfth**

15. **five** **ten** **twenty** **twenty-five**

16. **day** **month** **week** **year**

Go on to the next pa

groups; moreover, it tends to free the norms of the effects of both retardation and acceleration. The modal-age group for a grade is more precisely defined as the group of pupils in the 12-month age range in which the greatest number of pupils is included. To identify the modal-age group for each grade, the mode of the distribution of ages for the grade was determined mathematically, and a range of six months above and six months below this mode was taken as defining the "modal-age range" for the grade. From the same distributions of chronological ages within grades it was possible to determine so-called "age-grade couplets," which are the associated age and grade values that indicate the typical, or modal, age of pupils of specified grade status. These age-grade couplets are shown in Table 4; they are the bases for assigning age equivalents to scores in the modal-age norm system.

Percentile norms. The percentile norms for end-of-grade testing were derived directly from the distributions of scores for modal-age pupils, and for intermediate points by a process of interpolation between the successive end-of-year results.

Construction

A major goal in the preparation of this edition of *Stanford* was to insure that the content of the test would be in harmony with present objectives, and measure what is actually being taught in today's schools. To make certain that the test content would be valid in this sense, the construction of the new edition (as of each earlier edition) was preceded by a thorough analysis of the most widely used series of elementary textbooks in the various subjects, of a wide variety of courses of study, and of the research literature pertaining to children's concepts, experiences, and vocabulary at successive ages or grades. On the basis of this analysis, the authors prepared detailed outlines of the content to be covered by all subtests at all grade levels. These outlines specified the relative proportion of content to be devoted to the various skills, knowledges, and understandings within each area and served as blueprints for the tests that were ultimately to emerge. At this stage, as well as throughout the whole developmental process, reliance was placed on the judgment of subject-matter specialists in the several areas.

The actual task of preparation of test material was completed early in 1951. Sufficient material was developed for seven experimental forms of each test at each level, each test being of approximately the same length as it was intended that the final form of the test should be. Inasmuch as it was planned that there would be not more than five final forms of the material, this arrangement provided for tryout of about 40 per cent more material than was ultimately to be retained for the final forms of the test. Experimental editions of the seven forms were prepared, these forms being designated Forms T–1, T–2, etc., to T–7. The experimental editions corresponded in page size, layout, typography, etc., as closely as possible to the anticipated final format.

The experimental editions were administered to approximately 12,000 pupils within about a month of the closing of school in spring, 1951. Because of the importance of the decisions with respect to elimination or retention of items that were to be based on results of this tryout, an effort was made to have the tryout sample a representative one with respect to such characteristics as regional distribution, size of school system, and rural *vs.* urban character.

Each coöperating system was asked to list the textbooks in use in each subject, at each grade level. Analysis of these lists of instructional materials indicated that the systems were w divergent in this respect, which was, of course, desirable fron standpoint of avoiding text-related bias in the sample.

Administration of the tests was done in practically all inst by the classroom teacher, in order that the administration w correspond most nearly to the typical regular administrati the tests. In addition to the experimental edition of *Stan* every pupil was given an intelligence test in order that data w be available for checking on the equivalence of ability of the gr taking the several forms (necessary for comparability of i difficulty values); and for checking on the extent to which item-analysis sample was a typical one with respect to ge ability level. The *Stanford* tests were administered essen without time limit in order that all pupils would have an op tunity to attempt all items. Provision was also made for pupils to record in each test the item which they had reach stipulated amounts of time, which information was used in d mining time limits for the final forms of the tests.

In general, the several forms of the experimental editions distributed sequentially within a classroom (i.e., in the order T–2, T–3, T–4, T–5, T–6, T–7, T–1, T–2, etc.) so that presum equivalent groups of pupils took each of the seven forms. equivalence was, as pointed out above, checked by a compa of the intelligence test scores for the pupils taking the va forms.

The experimental administration was designed to provide i analysis information not only for those grades in which a g battery was intended to function in its final form but also f least one grade below and one grade above this intended ra This was done in order to assist in the selection of items that w discriminate among the pupils at the extremes of the ranges in regular use of the tests.

For each of the approximately 10,000 questions in the ex mental forms, a count was obtained of the number of pupi each grade level answering the item correctly, and, in the ca multiple-choice items, the number of pupils selecting each of incorrect responses to the item. The numbers answering rectly were converted to per cents, and these per cents for cessive grades for a given item were considered to constitut "item profile," revealing the extent to which an item correl with progress through school. These item profiles were consid one of the most important indices of item validity, and cons able weight was attached to them in the selection of items fo final forms. Results of this item tryout permitted identifica of ambiguous items, of items either too easy or too difficult fo grades for which they were intended, and of items unsatisfac in other respects. Such items were eliminated from considera for retention in the final forms.

Each teacher participating in the administration of the ex mental editions was asked for comments, criticisms, or suggest for improving the tests. Teachers were asked to record sys atically on a form provided for the purpose their comments respect to clarity of questions and directions, appropriatenes content, format, typography, suitability of item types, and o aspects of the test.

The content of the five final forms of the test was selected f the total body of material tried out experimentally in such a that the final tests conform to the specifications with respec content, relative emphases, etc., originally established; that are of appropriate difficulty for the grades in which they ar tended to be used, and that they are highly comparable in con and difficulty.

TABLE 6. K-Scores Corresponding to Grade Scores for All Batteries

Grade Score	1 Par. Mean.	2 Word Mean.	3 Spell.	4 Lang.	5 Arith. Reas.	6 Arith. Comp.	7 Soc. Stud.	8 Sci.	9 St. Sk.
109	107.4	109.0	104.6	102.2	108.6	115.9	115.3	106.0	107.0
108	106.4	108.1	104.4	102.2	107.8	114.6	113.2	105.4	106.4
107	105.4	107.1	104.2	102.2	107.0	113.1	111.1	104.8	105.8
106	104.7	106.3	104.0	102.1	106.2	111.6	109.2	104.2	105.2
105	103.9	105.4	103.9	102.1	105.5	110.0	107.3	103.7	104.5
104	103.1	104.4	103.8	102.1	104.7	108.5	105.7	103.2	103.9
103	102.5	103.6	103.6	102.0	103.9	107.1	104.1	102.7	103.4
102	101.8	102.7	103.4	102.0	103.0	105.9	102.7	102.3	102.8
101	101.2	101.9	103.3	101.9	102.2	104.7	101.4	101.8	102.4
100	100.6	101.1	103.1	101.9	101.3	103.4	100.0	101.4	101.8
99	100.0	100.3	102.9	101.8	100.5	102.0	98.9	100.9	101.3
98	99.6	99.6	102.7	101.8	99.9	100.8	97.9	100.6	100.8
97	99.1	98.8	102.5	101.7	99.3	99.7	97.3	100.0	100.5
96	98.5	98.0	102.4	101.7	98.7	98.7	96.7	99.6	100.1
95	98.1	97.3	102.2	101.6	98.0	97.4	95.9	99.1	99.8
94	97.7	96.5	102.0	101.6	97.5	96.3	95.2	98.6	99.6
93	97.3	95.8	101.8	101.5	96.9	95.0	94.4	98.1	99.3
92	96.8	95.1	101.7	101.5	96.3	93.9	93.8	97.7	98.9
91	96.4	94.5	101.6	101.4	95.7	92.8	93.2	97.3	98.6
90	95.8	93.7	101.4	101.4	95.1	91.6	92.4	96.8	98.3
89	95.4	92.9	101.2	101.3	94.6	90.2	91.6	96.4	97.9
88	95.0	92.3	101.0	101.2	94.0	89.0	90.9	96.0	97.6
87	94.6	91.6	100.7	101.1	93.4	87.7	90.3	95.6	97.3
86	94.2	90.8	100.4	101.0	92.7	86.3	89.6	95.2	96.9
85	93.8	90.2	100.2	100.9	92.0	85.0	88.9	94.8	96.6
84	93.3	89.5	99.9	100.8	91.2	83.5	88.2	94.4	96.3
83	92.7	88.7	99.7	100.6	90.5	82.0	87.5	94.0	96.0
82	92.3	88.1	99.4	100.5	89.7	80.8	86.9	93.6	95.7
81	91.8	87.5	99.0	100.4	88.9	79.4	86.2	93.3	95.4
80	91.4	86.8	98.7	100.3	88.2	78.1	85.5	92.9	95.0
79	91.0	86.2	98.4	100.1	87.6	76.8	84.9	92.6	94.7
78	90.6	85.5	98.0	99.9	86.7	75.4	84.3	92.2	94.4
77	90.0	84.9	97.7	99.7	85.7	74.0	83.6	91.7	94.1
76	89.6	84.3	97.3	99.6	85.0	72.6	83.0	91.4	93.8
75	89.1	83.7	97.0	99.4	84.3	71.2	82.5	91.1	93.4
74	88.7	83.1	96.7	99.2	83.6	69.7	81.9	90.8	93.0
73	88.3	82.6	96.4	98.9	82.7	68.1	81.4	90.5	92.7
72	87.8	82.0	96.0	98.7	82.1	66.7	80.8	90.1	92.4
71	87.4	81.5	95.6	98.5	81.4	65.3	80.3	89.8	92.1
70	86.8	81.0	95.3	98.3	80.7	63.8	79.8	89.5	91.8
69	86.4	80.5	94.9	98.1	80.1	62.2	79.3	89.1	91.5
68	86.0	80.0	94.5	97.9	79.4	60.7	78.8	88.8	91.2
67	85.6	79.6	94.1	97.7	78.7	59.2	78.3	88.5	90.8
66	85.2	79.1	93.7	97.5	78.2	57.7	77.8	88.1	90.4
65	84.6	78.6	93.3	97.3	77.6	56.4	77.4	87.8	90.1
64	84.2	78.2	92.8	97.0	77.0	55.2	76.9	87.5	89.7
63	83.7	77.7	92.4	96.7	76.5	53.9	76.5	87.2	89.4
62	83.3	77.3	91.9	96.5	75.8	52.7	75.9	86.9	88.9
61	82.9	76.9	91.4	96.3	75.3	51.5	75.5	86.6	88.5
60	82.4	76.5	90.9	96.0	74.7	50.3	75.0	86.2	88.1

Grade Score (*Cont'd*)	1 Par. Mean.	2 Word Mean.	3 Spell.	4 Lang.	5 Arith. Reas.	6 Arith. Comp.	7 Soc. Stud.	8 Sci.
59	82.0	76.1	90.5	95.7	74.1	49.1	74.6	85.9
58	81.5	75.7	90.0	95.4	73.5	47.9	74.1	85.7
57	81.0	75.4	89.5	95.1	73.0	47.0	73.6	85.4
56	80.6	74.9	89.0	94.9	72.6	46.3	73.1	85.0
55	80.3	74.5	88.5	94.6	72.0	45.5	72.6	84.6
54	79.9	74.1	87.9	94.3	71.5	44.7	72.2	84.3
53	79.6	73.7	87.5	94.0	71.0	43.8	71.7	83.8
52	79.2	73.3	86.9	93.7	70.5	43.2	71.3	83.5
51	78.7	73.0	86.5	93.4	70.0	42.6	70.9	83.1
50	78.4	72.6	85.9	93.1	69.5	41.9	70.4	82.7
49	78.0	72.2	85.4	92.7	68.9	41.2	70.0	82.4
48	77.6	71.8	84.9	92.5	68.3	40.6	69.6	82.0
47	77.3	71.5	84.5	92.1	67.6	40.0	69.2	81.6
46	76.9	71.1	83.8	91.7	67.0	39.6	68.7	81.2
45	76.6	70.7	83.2	91.4	66.5	39.2	68.3	80.8
44	76.2	70.3	82.7	90.9	65.9	38.7	67.8	80.2
43	75.8	69.9	82.1	90.6	65.4	38.3	67.4	79.7
42	75.5	69.6	81.6	90.1	64.8	37.8	66.9	79.2
41	75.2	69.3	81.0	89.7	64.3	37.4	66.5	78.6
40	74.8	68.8	80.4	89.2	63.6	37.0	66.1	77.9
39	74.5	68.5	79.8	88.7	63.1	36.7	65.7	77.3
38	74.1	68.1	79.2	88.3	62.5	36.2	65.4	76.8
37	73.7	67.7	78.5	87.8	62.0	35.8	65.0	76.1
36	73.5	67.4	78.0	87.4	61.5	35.4	64.7	75.4
35	73.2	67.1	77.5	86.9	61.0	35.0	64.3	74.7
34	72.8	66.7	76.7	86.4	60.5	34.6	63.9	74.0
33	72.5	66.4	76.0	85.9	59.9	34.3	63.5	73.3
32	72.2	66.0	75.3	85.4	59.4	33.8	63.2	72.5
31	71.9	65.7	74.7	85.0	58.8	33.5	62.9	71.7
30	71.6	65.4	73.9	84.5	58.3	33.2	62.5	70.8
29	71.3	65.0	73.3	84.0	57.7	33.0	62.1	70.0
28	71.0	64.7	72.5	83.5	57.1	32.8	61.7	69.2
27	70.7	64.4	71.7	83.0	56.6	32.7	61.4	68.3
26	70.5	64.1	71.0	82.5	56.2	32.6	61.0	67.4
25	70.3	63.7	70.3	82.0	55.6	32.5	60.7	66.5
24	70.0	63.4	69.4	81.6	55.2	32.4	60.4	65.4
23	69.7	63.1	68.7	81.0	54.7	32.3	60.0	64.5
22	69.5	62.8	67.8	80.5	54.1	32.2	59.7	63.4
21	69.4	62.5	67.0	80.0	53.6	32.1	59.4	62.4
20	69.1	62.1	66.1	79.5	53.0	32.0	59.0	61.3
19	68.8	61.8	65.2	78.9	52.4	31.9	58.7	60.1
18	68.7	61.6	64.4	78.3	51.9	31.9	58.4	59.1
17	68.4	61.4	63.6	77.8	51.5	31.8	58.0	58.0
16	68.2	61.0	62.7	77.2	50.9	31.7	57.7	56.9
15	68.0	60.7	61.8	76.7	50.4	31.7	57.4	55.9
14	67.7	60.3	61.0	76.1	49.8	31.6	57.0	54.7
13	67.6	60.0	60.3	75.5	49.3	31.6	56.7	53.4
12	67.4	59.8	59.6	74.9	48.6	31.5	56.4	52.2
11	67.2	59.6	58.7	74.4	48.0	31.5	56.1	51.1
10	67.0	59.4	57.9	73.8	47.3	31.5	55.8	50.0

7 K-Scores

Grade or age equivalents have a limitation that is often not recognized: they do not constitute scales of equal units. A gain of one year of grade or age equivalent at one part of the scale does not represent the same amount of growth in ability as does gain of a year at another part of the scale. The units in a percentile system similarly suffer from a lack of equality. While, for many purposes, this lack of equality of units is not a serious deficiency, it is necessary for accurate measurement of growth to have scales of equal units. To meet the need for such scales, there have been developed for each of the *Stanford* subtests so-called *K-scales*, which have units that are approximately equal throughout the ent range. K-scales are such that for all functions average performance of pupils of Grade 10.0 ha value of 100 and the unit of measurement is equal one seventh the standard deviation of the scores fifth-grade pupils.[1] Table 6 gives the K-scores co sponding to grade scores for all subtests. It is necessary, in the ordinary interpretation of scores, convert results to K-scores, but for study of relat rates of growth in the various functions, true chan in variability in the functions over the range of mentary grades, and related problems, the K-sco are superior to raw or grade scores.

[1] The theory underlying K-scores and the procedures invol in their derivation are fully described in "The Determinatio Units of Measurement Which Are Consistent with Inter- and In Grade Differences in Ability" by Eric F. Gardner; unpublis Ed.D. dissertation, Harvard University School of Education, 1

Profile Chart. Interpretation of results is facilitated by use of the Profile Chart on the front page of each test booklet. A marked copy is reproduced below as Figure 1. On this Profile Chart there is a scale for each test in the battery, graduated in terms of grade score (or grade equivalent). Results on the several subtests are plotted on the Profile Chart by making a cross or large dot at the appropriate place on the respective scales. If these plotted points are connected, the resulting graph shows areas of strength and weakness, and the magnitude of the departure from typical performance in the various subjects. A line drawn across the Profile Chart at a point corresponding to the grade placement at the time of testing will serve as one reference point for evaluating the status in each subject. It is a useful practice to retain the completed Profile Chart in a pupil's record folder.

The short bar below the scale for each subtest indicates the magnitude of the *standard error of measurement.* In every test score there is a certain "error of measurement," by which is meant the difference between the score as obtained and the hypothetical "true" score or perfectly reliable measure. For any test the magnitude of such differences is indicated by the standard error of measurement of a score on the test. If one asserts that an individual's "true" score will fall in an interval extending one standard error of measurement on either side of his obtained score, such assertion will be correct, on the average, in two cases out of three.

Class Record and Class Analysis Chart. A Cl Record and Class Analysis Chart is provided w each package of tests. It permits a convenient tabu tion of the scores for a complete class and facilita summarizations of the scores.

Reliability. Reliability data for the tests in Elementary Battery are given in Table 5.

TABLE 5. Reliability Coefficients[1] and Related Data

Test	N	r_{11}[2]	Mean	Stand. Dev.[3]	S. E Mea
Grade 3					
Par. Mean.	240	.914	27.77	9.73	2.8
Word Mean.	240	.909	22.60	8.24	2.4
Spelling	240	.953	26.60	11.38	2.4
Language	240	.868	33.89	17.99	6.5
Arith. Reas.	240	.905	20.49	7.96	2.4
Arith. Comp.	240	.893	21.14	6.17	2.0
Grade 4					
Par. Mean.	251	.898	34.40	8.33	2.6
Word Mean.	251	.924	28.58	7.19	1.9
Spelling	251	.946	35.26	10.84	2.5
Language	251	.869	45.06	15.50	5.6
Arith. Reas.	251	.933	32.03	8.83	2.2
Arith. Comp.	251	.900	31.65	6.59	2.0

[1] Split-half reliability coefficients corrected by the usual Spearn Brown formula.

[2] All tests except Spelling are time-limit tests, but the time limits sufficiently generous to permit almost all pupils to attempt all items. is not believed that the speed factor is sufficient to cause these split- coefficients to be spuriously high.

[3] Pupils in a single class, school, or even school system would prob exhibit somewhat less variability in the various tests than these sam of pupils from numerous systems, and hence might yield reliability c ficients slightly lower than those here reported.

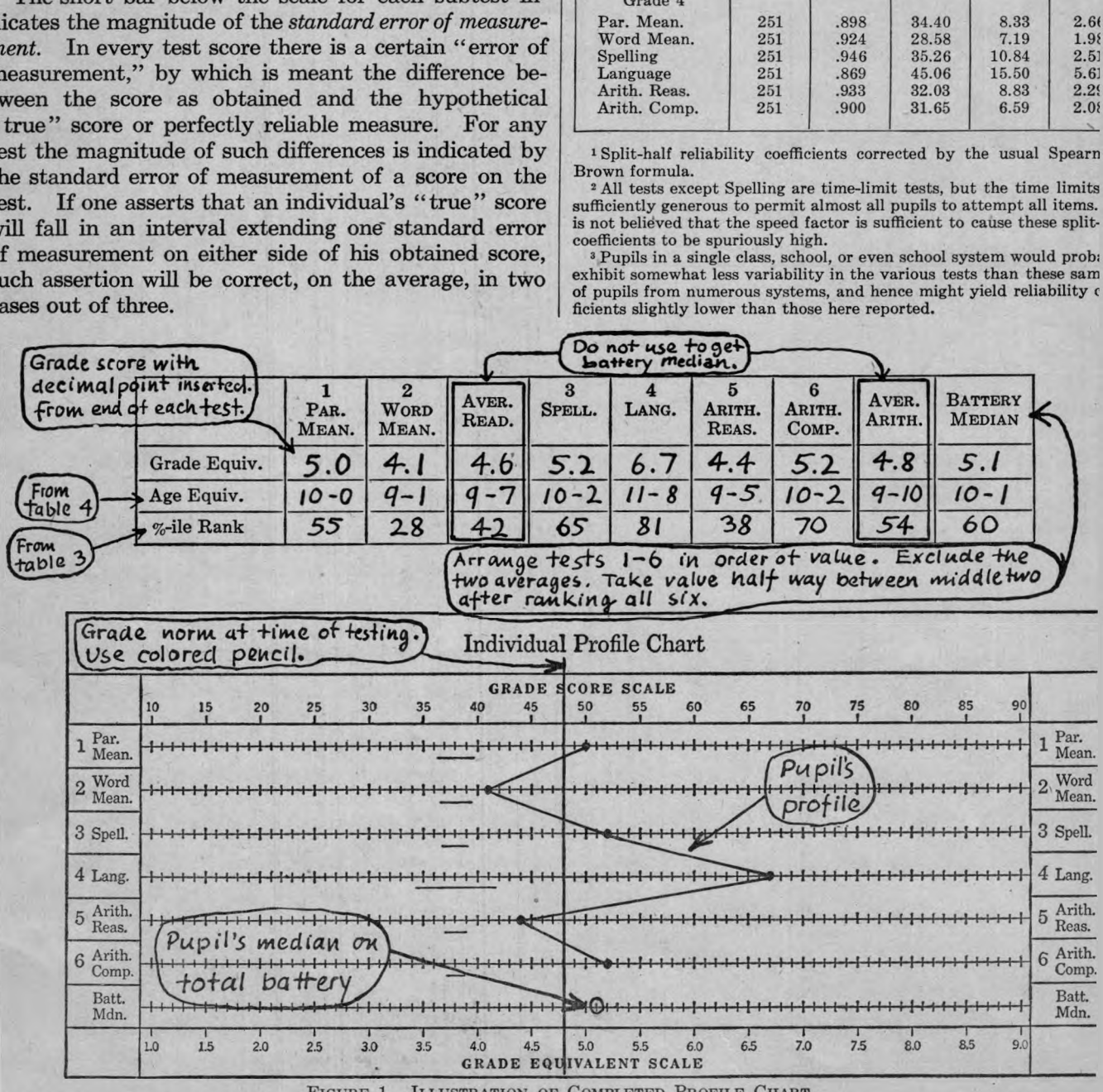

FIGURE 1. ILLUSTRATION OF COMPLETED PROFILE CHART

TABLE 3. Grade Scores Corresponding to Selected Percentiles for Modal-Age Groups

%-ile	Grade 3.2						%-ile	Grade 3.5						%-ile	Grade 3.8				
	Par. Mean.	Word Mean.	Spell.	Lang.	Arith. Reas.	Arith. Comp.		Par. Mean.	Word Mean.	Spell.	Lang.	Arith. Reas.	Arith. Comp.		Par. Mean.	Word Mean.	Spell.	Lang.	Arith. Reas.
98	57	53	54	70	45	39	98	63	59	58	73	49	43	98	72	64	61	76	52
95	50	49	49	63	42	38	95	56	53	53	66	46	42	95	63	59	57	69	49
90	46	43	44	55	40	37	90	50	48	47	59	43	40	90	57	54	52	63	47
85	42	41	42	50	38	36	85	46	44	45	54	42	39	85	53	51	49	58	45
80	40	39	40	46	37	35	80	44	43	43	50	41	39	80	49	48	46	54	44
75	38	37	38	43	36	34	75	42	41	41	47	40	38	75	47	46	45	50	43
70	36	36	37	41	35	34	70	40	40	39	44	39	38	70	45	44	43	47	42
65	35	35	35	39	34	33	65	38	39	38	41	38	37	65	43	42	42	45	41
60	34	34	34	36	33	32	60	37	37	37	39	37	36	60	42	41	40	42	40
55	33	33	33	34	32	32	55	36	36	36	37	36	36	55	40	40	39	40	39
50	32	32	32	32	32	31	50	35	35	35	35	35	35	50	38	38	38	38	38
45	31	30	31	29	31	30	45	34	33	34	32	33	34	45	37	36	37	36	37
40	30	29	30	26	30	30	40	32	32	33	30	32	33	40	36	35	36	33	36
35	29	28	29	24	29	29	35	31	31	32	28	31	33	35	34	33	35	31	35
30	28	27	28	21	28	29	30	30	29	31	25	30	32	30	33	32	33	28	33
25	26	26	27	18	27	28	25	29	28	29	21	29	31	25	31	31	32	25	32
20	25	25	26	15	26	28	20	27	27	28	18	28	30	20	29	29	30	22	31
15	23	23	24	11	24	27	15	25	25	26	15	26	29	15	27	27	28	18	29
10	21	21	21	10–	22	26	10	23	23	23	10–	24	28	10	25	25	26	13	26
5	18	19	19		19	23	5	20	21	21		22	26	5	21	22	22	10–	23
2	15	17	17		17	20	2	16	18	18		18	24	2	17	19	19		19

%-ile	Grade 4.2						%-ile	Grade 4.5						%-ile	Grade 4.8				
	Par. Mean.	Word Mean.	Spell.	Lang.	Arith. Reas.	Arith. Comp.		Par. Mean.	Word Mean.	Spell.	Lang.	Arith. Reas.	Arith. Comp.		Par. Mean.	Word Mean.	Spell.	Lang.	Arith. Reas.
98	76	66	63	79	56	53	98	80	69	65	83	62	57	98	86	71	68	86	69
95	68	62	59	73	52	51	95	72	65	62	76	58	55	95	79	69	65	79	65
90	61	57	55	66	50	49	90	65	61	58	71	54	52	90	71	65	62	74	62
85	56	53	53	62	49	47	85	61	57	55	66	52	51	85	66	62	60	70	59
80	53	51	50	58	47	46	80	57	55	54	62	51	50	80	62	59	58	66	57
75	51	49	48	54	46	45	75	55	52	51	58	50	49	75	59	57	56	64	55
70	48	46	46	51	45	44	70	52	50	50	55	49	48	70	56	55	54	60	53
65	46	44	45	49	44	43	65	50	48	48	52	48	47	65	54	53	52	57	52
60	44	43	44	46	43	42	60	48	47	47	49	47	46	60	52	51	50	54	51
55	43	42	43	44	42	42	55	46	45	46	47	46	45	55	50	50	49	51	50
50	42	41	42	42	41	41	50	45	44	45	45	45	44	50	48	48	48	48	48
45	40	40	40	39	40	40	45	43	43	43	42	44	43	45	46	46	46	46	47
40	38	39	38	37	39	39	40	41	42	41	40	42	42	40	44	44	44	44	46
35	37	37	37	35	37	38	35	39	40	39	38	41	41	35	43	43	43	41	45
30	35	35	35	32	36	38	30	38	38	38	35	39	40	30	41	42	41	39	43
25	34	33	34	29	35	37	25	37	36	37	33	38	39	25	39	40	39	36	41
20	32	32	32	25	34	36	20	35	34	35	29	37	38	20	37	38	37	33	39
15	30	30	30	21	32	35	15	32	32	33	26	35	37	15	35	35	35	30	37
10	28	28	28	17	29	33	10	30	30	30	21	32	36	10	32	32	32	25	35
5	24	24	24	10	26	30	5	26	26	26	14	28	32	5	28	28	29	17	31
2	20	21	20	10–	22	28	2	21	23	21	10–	24	29	2	23	25	23	10–	27

Percentile norms based on modal-age grade groups. For many purposes it is preferable to compare the scores of a pupil with the scores of other pupils of the same grade status rather than with those of pupils in other grades, which, in effect, is what a grade equivalent system of interpretation does. Accordingly, there are presented in Table 3 *percentile norms* for the subtests of the Elementary Battery. In this table the grade scores corresponding to selected percentiles for modal-age groups are given for several possible testing-dates—roughly for beginning-, middle-, and end-of-year testing. Percentiles for scores not included in the tables or percentiles for testing dates other than those included may be estimated by interpolation if such additional refinement is desired. The percentile rank corresponding to a grade score shows the percentage of pupils of the given grade placement hav scores less than the given score. For example, a gr score of 63 on the Paragraph Meaning Test ha percentile rank of 95 at Grade 3.8, which means t 95% of the modal-age pupils of this grade sta made grade scores less than 63. Use of percen norms tends to avoid erroneous inferences as to desirability of effecting changes in grade placeme such as are sometimes drawn when a pupil's gr equivalents differ greatly from his actual grade pla ment. Percentiles have the further advantage being easy to explain to parents or other teach Although the advantages of percentile norms most apparent in the upper grades, it is becom increasingly common to utilize this system for in pretation of scores at all levels.

4 Directions for Scoring

The directions for scoring accompany the keys which contain the correct answers. These directions for scoring should be read carefully and followed exactly, and the keys should be used for the greatest accuracy and ease of scoring.

Correction for possible guessing is made only in the Language Test. It is the only test with answers of two choices and is, therefore, more susceptible to unreliability through guessing. There will occasionally be negative raw scores in the Language Test results. All negative scores are treated as zero scores.

5 Interpretation of Scores; Norms

A raw score (number of right answers) on *Stanford Achievement Test*, as on most tests, has but little meaning in itself. Raw scores acquire meaning when they are related to some set of *norms* — that is, the scores made by some specified group, and converted to some type of score intrinsically more meaningful. There are many groups of pupils whose scores might be used as bases for comparison or interpretation of raw scores — e.g., pupils of given grade, age, sex, mental ability, curriculum, type of school, etc., but, practically, it is not possible to provide all the kinds of normative data that might be useful.

Scores on *Stanford Achievement Test* are interpreted chiefly by reference to two sets of norms, as follows:

a. *Modal-age grade norms*, recommended for interpretation of individual scores.
b. *Total-group grade norms*, recommended for interpretation of group averages.

The nature of these two sets of norms and their special characteristics are outlined below. This discussion will be more meaningful if one understands the nature of grade and age norms. *Grade norms* permit comparison of a score with the scores made by pupils of specified grade status. They yield a *grade equivalent* for a given score, which indicates the grade placement of pupils for whom the given score is the average score. Similarly, *age norms* permit comparison of scores with the scores obtained by pupils of specified age, and yield *age equivalents*.

Modal-age grade norms. The 1940 edition of *Stanford Achievement Test* introduced a refinement in the derivation of grade norms, the so-called *modal-age-grade* concept. Grade norms embodying this concept, instead of being based on the scores of all pupils in a given grade, are based on the scores only of tho pupils who are typical with respect to age. Thus, tl norm group is specified with respect to both grade ar age status. The most common single age group in grade is designated the "modal-age" group, when the name for these norms. Use of modal-age norn permits comparison of the typical child in a grac with children who are most nearly like him with r spect to both grade and year of age. These childre have been in each grade only one year and entere school at nearly the same age.

Approximately 65% of the children in any grac are in the modal-age group.[1] To compare thes children with the total population in each grac would be to reduce the accuracy of the compariso to the extent that the norms for the modal-age grou and the total group differ; these differences rang from approximately one tenth of a year in the earl grades to a half year, or more, in the higher grade The difference is negligible at grades two and three but of increasing magnitude in later grades. Th average performance of the total grade group i depressed by virtue of the fact that the total grou includes both accelerated and retarded pupils, i addition to those normally placed for age, and becaus there are more retarded, duller pupils than ther are accelerated, brighter pupils. Use of the tota group average as the norm for evaluation of an in dividual's performance, therefore, sets an unduly lo standard for the majority of pupils and, in the lon run, is likely to encourage acceptance of an unneces sarily low level of achievement.

The Profile Chart (see page 12) on the front of th test booklet is drawn in terms of modal-age grad norms. (This was also true of the Profile Chart o the 1940 edition.)

If results on *Stanford* expressed in terms of modal age norms are to be compared with results on othe achievement tests, which are not expressed in term of modal-age norms, it should be recognized that ther are these systematic differences between modal-ag grade and total-group grade norms. Appropriate ad justments should be made when different types o norms are involved.

Total-group grade norms. Total-group grade norms in contrast to modal-age norms, are based on the per formance of *all* the pupils in a given grade. Th total-group norms should be used to interpret averag scores of a total class, school, or school system becaus usually such a total includes both the approximat 65% of pupils who are at grade for age and the othe 35% who are retarded or accelerated.

The total-group grade norms are given in Table 1 This table gives total-group grade equivalents corre sponding to *grade scores* in the various subjects. A given score on any subtest nearly always has a highe grade equivalent according to total-group norm than it does according to modal-age norms.

[1] Exact per cent in each grade is given in Table 8.

FIRST SITTING — TEST 1. *Paragraph Meaning*

Say to the pupils: "Now open your booklet to Test 1, Paragraph Meaning, which is on page 2. (See that all pupils have the correct page.) Now fold the page back, like this, so that only page 2 is showing. (Demonstrate and see that all do this correctly.)

"Look at the top of the page, where it says 'DIRECTIONS.' (Hold up a booklet and point to the proper place.) The directions tell you what to do. They say: 'Find the word that belongs in each space, and draw a line under it. Do *not* write in the spaces.' Now look at the first sample story. It says: 'Wheat grows on farms. Most bread is made from wheat. If farmers did not plant . . . most people would have no . . . to eat.' What word goes in the space which has the number 51 in it? (Encourage reply.) Yes, the word that belongs in the space is 'wheat.' Look at the four words in the row that begins with 51. They are — corn, potatoes, rice, wheat. Since 'wheat' is the word that belongs in the space marked 51, a line has been drawn under 'wheat.' What word belongs in the space numbered 52? (Encourage reply.) Yes, the word is 'bread,' so you draw a line under the word 'bread,' which you see in the line numbered 52. All the stories on this page have one or more words left out. Each space with a number in it tells you where a word has been left out. For each space there is a row of words beside the same number as appears in the space. The word that goes in each space is one of the words in the row that has the same number as the space has.

"You are to read each story and find the words that have been left out. Begin with the first story and answer as many questions as you can. When you are not sure which answer is the right one, make the best choice you can, but do not make wild guesses. When you reach the bottom of the page, go on to pages 3 and 4. When you finish page 4, go back and see if you have done the best you can. Do not work on any other tests. Are there any questions about what you are to do? (Pause.) READY. GO!" (Record the starting time on the board.)

After *25 minutes*, say: "STOP! Put your pencil down. Close your booklet." Collect the test booklets immediately. (The first sitting should end here.)

SECOND SITTING — TEST 2. *Word Meaning*

When the second sitting begins, make sure that each pupil has his own booklet.

After the booklets have been distributed, say to the pupils: "Now open your booklet to Test 2, Word Meaning, on page 5. (See that all pupils have the correct page.) Now fold the page back, like this, so that only page 5 is showing. (Demonstrate and see that all do this correctly.) Look at the top of the page, where it says 'DIRECTIONS.' (Hold up a booklet and point to the proper place on the booklet.) The directions tell you what to do. They say: 'Draw a line under the one word that makes the sentence true, as shown in the first sample. Look at all four words and choose the best one.' Now look at the samples. (Hold up a booklet and point to the sample exercises.) The first sample says: 'The name of a color is — farm milk red pet.' Whi of the last four words makes the sentence true? (Wa for the class to answer.) Yes, the name of a col is red, so a line has been drawn under the wo 'red.' The second sample says: 'The day that com after Friday is — Monday Tuesday Saturday Su day.' What is the answer? (Pause for reply.) Ye 'Saturday' is the answer, so you draw a line under t word 'Saturday.' In each sentence on this page you a to draw a line under the word that makes the senten true.

"Begin with question number 1 and answer as mar questions as you can. When you are not sure whi answer is the right one, make the best choice you can, b do not make wild guesses. Are there any questio about what you are to do? (Pause.) READY. GO (Record the starting time on the board.)

After *8 minutes*, say: "STOP! Turn to the ne page, which is for Test 3, Spelling. (Pause until all hav found the place. If children seem tired or restles allow a few minutes of relaxation before starting Te 3.) Now fold the page back, like this, so that only pag 6 is showing." (See that all do this correctly.)

SECOND SITTING (*Cont'd*) — TEST 3. *Spelling*

"On this page you are to write some words. First shall give the number of a word and then say the wor Next I shall read a sentence with that word in it. The I shall say the word again. Then you write that word be side its number. Listen carefully and be sure to writ each word beside the right number."

The spelling words and sentences are on a separat list, which is included with your package of test Specific directions for administering the Spelling Te are given on the List of Words for Test 3, Spelling.

When the last word has been spelled, say: "Clos your booklet." Collect the test booklets immediatel (The second sitting should end here.)

THIRD SITTING — TEST 4. *Language*

When the third sitting begins, make sure that eac pupil has his own booklet.

After the booklets have been distributed, say to th pupils: "Now open your booklet to Test 4, Languag which is on page 7. (See that all pupils have the co rect page.) Now fold the page back, like this, so that onl page 7 is showing. (Demonstrate and see that all d this correctly.) Look at the top of the page, where it say 'DIRECTIONS.' (Hold up a booklet and point to th proper place on the booklet.) The directions tell yo what to do. They say: 'In each pair of words in heav type in the letter below there is an error in either capitali zation or punctuation. You are to decide which one c each pair has the correct capitalization and punctuatio Then mark the answer space at the right that has the sam number as the correct form.' Look at the first sample It says: 'This is 1 mr. Jones,' with a small M in Mr. an '2 Mr. Jones' with a capital M in Mr. The second for is right, since 'Mister' should have a capital M; so th

writing he does these things correctly. For the sake of simplification of measurement and scoring it is necessary to test these skills in a somewhat artificial fashion, perhaps with some loss of validity. Within the limits of objective measurement at these grade levels, however, it is believed that the Elementary Language Test affords a valid appraisal of mastery of those aspects of language which the test purports to cover.

The importance of standards of usage must be granted for a language which is the medium of communication in all parts of a large nation. Colloquialisms which may be very serviceable in a limited area cannot be credited in a nation-wide test. It must be recognized, however, that modern usage is occasionally at variance. Items on matters that are highly controversial have been avoided, but if every item about which there is some disagreement were to be excluded, there would be little left to test. Tradition has been the standard of correctness up to the point where controversy makes tradition untenable. Items beyond that have not been selected for use. As long as language usage is changing, there will be occasion for argument within the transition zone.

Language Test scores reflect a combination of home background, curriculum content, and possibly of intensity and persistency of instruction. Each school will need to judge first of all the language background of its pupils. In the light of the conclusions, it may gauge the scope of its language problem and estimate the possible degrees of improvement. Capitalization and punctuation skills and sentence sense are more amenable to change through instruction than is word usage, because of its closer relationship to out-of-school practice.

Arithmetic Reasoning. The Arithmetic Reasoning Test is divided into two parts. Part I measures reasoning in problems taken from life experiences. The general reading vocabulary has been kept much below the problem-solving level being measured. Computation difficulty has been controlled so that it is only a minor factor.

To be sure that the test would be representative of the many kinds of arithmetic problems that confront children, each of the 30 problems in Part I was defined under a two-way classification. Each problem was classified (1) in accordance with the four fundamental processes of addition, subtraction, multiplication, and division; and (2) in accordance with the kinds of measures used; namely, space (linear, area, volume), weight, time, temperature, and value.

From a consideration of these classifications a observed curriculum practice, problem types we assigned to the batteries in amounts judged to appropriate.

Though most problems were written in the form a simple direct statement and question, some we tried out with extra numbers, without numbers, wi two denominations of measures, with "hidden" nu bers, and in other ways commonly used to stimula reasoning. Seldom did any type other than the dire statement and question survive the tryout evaluati of items.

Part II tests two essential components of the abili to reason in arithmetic; namely, the informatio background of children and their understanding of t number system.

Arithmetic reasoning involves a combination of t language of quantitative comparison, arithmetic i formation, ability to read, the thinking aspects problem solving, some ability in computation, and oth factors which are merged into one score. Gener deficiency in this test, and in arithmetic computatio suggests shortcomings of performance, perhaps of i struction, which are as yet unidentified. Low te scores indicate possibilities so numerous and vari that the total range of teaching procedures in arit metic should be reëxamined by the teacher and supervisor.

Arithmetic Computation. The Arithmetic Comp tation Test consists of 42 exercises which cov chiefly fundamental operations with whole numbe A small number of items cover fundamental oper tions involving dollars and cents and one exerci involves addition of two fractions of like denon nator. Time limit for the test is generous, reduci the emphasis on computational speed. The exercis are representative of the usual curriculum and te book patterns of content at the third and fourth gra levels.

The five forms are closely parallel in content that each form consists of the same major types exercises. Most major types of examples may subdivided into several variants or minor typ and advantage was taken of this fact to avoid a rig patterning from form to form. This variation fr form to form makes the tests less susceptible coaching.

The free-response type of item is used in both arit metic tests of the Elementary Battery to simplify t work of children at this age.

The Intermediate and Advanced tests are also available in *Partial Batteries*, each of which includes, in a single 16-page booklet, the Paragraph Meaning, Word Meaning, Spelling, Language, Arithmetic Reasoning, and Arithmetic Computation tests. The reading tests and the arithmetic tests of the Elementary, Intermediate, and Advanced batteries also are published as separate booklets.

The Intermediate and Advanced Partial Batteries are available in a special edition designed for use with separate answer sheets which may be scored on the IBM Test Scoring Machine (or by hand). The tests in this special edition have the same content as the corresponding tests in the regular edition. These machine-scorable forms are designated Forms JM, KM, etc. The Social Studies, Science, and Study Skills tests are also available as separate tests, in the separate-answer-sheet version only.

The grades indicated for the several batteries are the grades for which the batteries are most appropriate in the great majority of school systems. There are, however, schools, or individual classes, in which achievement is so markedly superior (or inferior) to the typical level as to make it advisable to use in a given grade the battery designated for a higher (or lower) grade.

The five forms J, K, L, M, and N of the various tests are matched for content and difficulty, represent equally good measures of the respective subjects, and yield directly comparable results.

In the case of several tests, there is some overlapping or identity of content between adjacent batteries — that is, a question which appears, for example, in the Elementary level of a test may also appear in the Intermediate level of the same test. In almost all instances of such repetition, the question appears in the same form (J, K, etc.) of the two batteries. Because of this overlapping, it is prudent to use a different form of the test when pupils are retested at an interval of a year or less, even though different batteries are used.

The availability of five equivalent forms of *Stanford* is a decided advantage for systems which are conducting or plan to conduct a continuing program of annual testing. Annual testing, or even twice-yearly testing at the beginning and end of the school year, can be conducted without the necessity for re-administering the same form of a given battery to any pupils.

Although almost all tests in the several batteries are time-limit tests, the time limits are provided as matters of administrative convenience rather than for the purpose of placing any premium upon speed of work. The time limits in all cases are generous and calculated to give practically all pupils sufficient time to attempt all questions which they are capable of answering correctly. The tests, therefore, are fundamentally power tests and not speed tests.

2 Content of the Elementary Battery

The Elementary Battery includes tests of readin arithmetic, language, and spelling, the abilities to whic greatest attention is devoted in the instructional pr gram of Grades 3 and 4. Reading is measured b means of two tests, a Paragraph Meaning and a Wor Meaning test. Arithmetic is measured by means o two tests, Arithmetic Reasoning and Arithmetic Com putation. Language and Spelling are each measure by means of a single test. The nature and content o these several tests are briefly described below.

If slow-learning groups are to be tested early in th third grade, it will be preferable to use the Primar rather than the Elementary Battery.

Stanford Achievement Test is highly analytical amon subjects, but it makes no claim for accurate diagnosi within a given subject. Part scores based upon se lected items from a test or upon unit sections of a tes may be satisfactory for group diagnosis even thoug not sufficiently reliable for individual diagnosis.

Teachers should not attach undue importance to pupil's error on any single item or small groups o items. This is not to say that a teacher may not gai insight into a pupil's needs by going over his test recor item by item, or a group's needs by analysis of a fe items, and following by discussing with the childre the reasons for the incorrect responses.

Paragraph Meaning

The Paragraph Meaning Test consists of a series o paragraphs, graduated in difficulty, from each of whic one or more words have been omitted. The pupil' task is to demonstrate his comprehension of the para graph by selecting the proper word for each omissio from four choices that are afforded him. The tes thus provides a functional measure of the child' ability to comprehend connected discourse ranging i length from single sentences to paragraphs of six sen tences, and involving levels of comprehension varyin from extremely simple recognition to the making o inferences from several related sentences. Specia efforts have been made to devise paragraphs interest ing to young children, and to make certain that th level of vocabulary is such that the test does no become one of word knowledge rather than of com prehension of connected discourse. The authors hav attempted to emphasize the notion of "reading a reasoning" and, accordingly, have constructed exer cises that place a premium on genuine comprehensio of the material read.

Use of the multiple-choice type of item employed i this test has been found to be entirely satisfactory a these grade levels, and it eliminates the element o subjectivity to be found in tests of a completion or fill-i