

C  
IP6nJ  
ser14#57

UNIVERSITY OF ILLINOIS

lib

# THE NORMAL SCHOOL QUARTERLY

UNIVERSITY OF ILLINOIS LIBRARY

Series 14

January, 1916

Number 58

JAN 23 1917

## Standards Employd in the Determination of Teach- ing Efficiency

THE LIBRARY OF THE  
DEC 9 - 1938  
UNIVERSITY OF ILLINOIS

By

**EDWIN A. TURNER**

**PUBLISHT JANUARY, APRIL, JULY, AND OCTOBER OF EACH  
YEAR BY THE ILLINOIS STATE NORMAL UNIVERSITY,  
NORMAL, ILLINOIS**

**Enterd August, 1902, at Normal, Illinois, as second-class mail matter  
under Act of Congress of July 16, 1894**

N. B.—Any teacher in Illinois may get the *Normal School Quarterly* regularly by sending exact name and address, and by giving prompt notis of any change of address.

Simplified spellings ar used in the official publications of the Illinois State Normal University.



JAN 23 1917

# *Normal School Quarterly*

---

*Published by the Illinois State Normal University, Normal, Illinois*

---

Series 14

JANUARY, 1916

No. 58

---

## STANDARDS EMPLOYED IN THE DETERMINATION OF TEACHING EFFICIENCY

BY

EDWIN A. TURNER

At present our pedagogical literature bristles with the term efficiency. Even writers of ability use it extravagantly. The term itself seems to satisfy. It suggests the shop, the factory, and the salesroom, where performances are judged in terms of the concrete and where definite standards are blocked out in open competition. It apparently pacifies the longing for scientific accuracy and generates a feeling of confidence in him who sets it up for his goal.

Unfortunately the teaching profession in the main has adopted efficiency as its slogan without making adequate provision for determining when it is attained. Until the spokesmen for the profession can in a very simple and in a very practical way point out the meaning of efficiency as it relates to specific attainment and can give explicit directions for determining the degree of efficiency of this or that sort of teaching, the term efficiency must be considered more or less platitudinous.

In the industries the ability of the performer is easily measured, since the products of his labor are objective, concrete, and readily subjected to comparative tests. The efficiency of the blacksmith is measured by the length of time the shoe clings to the hoof and by the degree of comfort it gives the horse. The efficiency of a dentist is measured by the length of time the filling remains in order or by the permanency and comfort of the bridge he has made. The efficiency of a gardener is determined by the number and quality of vegetables produced per unit of area. In any case when the result is better than that ordinarily produced the performer is thought of as having superior ability and consequently he is considered efficient.

Subjectively considered, efficiency is the ability to produce superlativ results consistently. The median or average of a number of such abilities is a desirable standard to use in an endeavor to determine the merit of individual performances. In the industrial and scientific fields such standards are well known. In the teaching profession we have just begun to use them advantageously. We cannot hope to attain efficiency until we are able to determine when it is attained.

With the single exception of the minimum knowledge requirement, which is generally provided by law, there is no other legally accepted standard for judging the ability of teachers. The wide and varied use of standards employed in determining the ability of teachers is notorious.

The far-reaching significance of the conditions resulting from the application of dissimilar standards is beyond the comprehension of those who evaluate the teaching process in terms of local and personal standards. There is not a little evidence to substantiate the opinion that subnormality, retardation, disinterestedness, disobedience, and withdrawals from school are the direct result of the inadequate standards held by administrators and teachers. Until some of the standards now employed in measuring the results of the teaching process are discarded and others are materially modified, the proportion of abnormalities occurring in the schools will not be materially changed.

### STANDARDS OF MEASUREMENT

There are two distinct classes of standards now employed in determining the merit of teaching. These may well be called the *a priori standards* and the *objectiv standards*. The former are deductions based upon definitions formed, principles assumed, or inferences drawn from known causes. The latter are based upon the measured abilities of pupils.

#### I. A PRIORI STANDARDS

This class of standards is in the main the outgrowth of an attempt on the part of those who have been responsible for the direction of educational agencies to account for the character of the services rendered by teachers, on the basis of some real or imaginary principle either directly or indirectly related to the art of teaching. The quality and relative value of each standard in

this class depends upon the educational ideals and insight of those who have established it.

The standards employed in the early stages of educational development and those still employed by persons unfamiliar with the essentials of the teaching process are crude and often ludicrous. On the other hand the standards which have been established by educational experts, in the light of recent research, are exceedingly valuable in that they stimulate an analysis of the process and give valuable direction to teaching.

#### **The Attitude of Pupils and the Community Towards the Teacher**

This standard is too frequently used by school officials in determining the efficiency of their teachers. If the children and the community are fond of a teacher it is assumed that he is giving splendid service in the classroom. If he is not generally popular it is taken for granted that he is giving poor service. Doubtless this standard was developed in and about the private school, and especially the subscription school where the teacher "boarded around". Under such conditions adaptability was the prime requisite of survival. In spite of the wonderful growth in the science of teaching there still exists in some communities the notion that popularity is an index of efficiency.

It is reasonably certain that a teacher of character and of fine teaching ability will win the respect and usually the admiration of his pupils and patrons. It is quite as reasonably certain that a relatively inferior teacher may and not infrequently does win the esteem and hearty support of the entire community in which he teaches. This esteem may result from local political activity, church connections, participation in club activities, or it may be in response to a wholesome attitude of the teacher towards the life of the community, all of which may be excellent supplementary qualities for a teacher to possess. Certainly they should not be the main consideration in the selection of a teacher.

Being a "good fellow" is an enviable human trait, but it has no legitimate place among the basal standards which are employed in determining the worth of teachers. The social and personal qualities of the officers of a bank do not become an incentive to me as a depositor until the standing of the bank and the integrity of the officials have been ascertained. The hearty greeting and the talkative propensities of a barber do not become an induce-

ment to me to patronize his shop until I have determined the fine quality of his razor and the sanitary practices of his establishment. No thoughtful parent will let church connections, social prestige, political affiliations, or friendship of long standing be the predominating factor in the choice of a physician for his dangerously sick child. Certainly there are stronger reasons why these supplemental and most desirable qualities should not be considered basic in the selection of a teacher.

#### Character of Grades and Number of Promotions

Another common and widely used standard of judging teaching efficiency, closely related to the above, is that of *grades and promotions*. It is passing strange that this standard of measurement should be relied upon so extensively. A parent usually thinks his children well taught if they receive high grades. He is quite as strongly convinced of the teacher's inferiority if his children fail of promotion. In view of recent investigations in respect to the reliability of grades, as an index of actual achievement, this standard is a travesty upon the science of education. A grade as ordinarily determined is, to say the least, the expression of a conglomerate impression which may be colored by a single performance of the pupil, by his general attitude toward the school, by the emotional controls of the teacher, or by the personal relations which exist between teacher and pupil or between the teacher and the family of the pupil.

Grades vary in proportion to the variation of personal standards. It is reasonable to suppose that an easy-going teacher is more likely to give high grades than is the teacher who is excessively conscientious and diligent in an endeavor to improve the standing of his pupils. It not infrequently happens that the grades of two chums, or of two children whose families are intimate, are adjusted from month to month so that first one pupil and then the other has the higher grade. It is notorious that good children receive higher grades in proportion to their ability than do mischievous children. Other influences well known to the profession are factors in determining grades. The multiplicity of factors involved in grade making is a strong indictment of the practice of judging teachers exclusively or even partially on the basis of the promotion list.

### Classroom Technique

The value of this standard rests on the assumption that there is a close correlation between the character of the stimuli employed by the teacher and the character of the child's controls which result from the use of such stimuli.

On the basis of this assumption one proceeds to determine a teacher's efficiency by an examination of her classroom technique. The following items are usually considered in such procedure: (1) forms of presenting subject-matter, such as the lecture method, the textbook method, the developing method, including a combination of one or more of these methods; (2) the character of the question employed—the direct question, indirect question, elliptical question, leading question, etc.; (3) the sort of other devices used—illustrations, drawings, field trips, concrete materials for science work, pictures, maps, etc.; (4) the language of the teacher, his intonation, the board work, the general appearance of the classroom, and especially the spiritual atmosphere of the room.

This standard is decidedly more reliable than either of those previously considered. It finds justification in the common agreement that the majority of teachers who get splendid results employ a good technique. In fact, teachers of this type find technique indispensable. It is in harmony also with certain generally accepted psychological principles. However, technique is not of itself a sufficient guarantee of adequate results, because of the large number of variables introduced in its application. The value of a device depends in large measure upon the experiences, judgment, temperament, zest, clearness of vision, physical energy, and high ideals of the teacher. Without these attributes in their proper proportion, technique in operation resolves itself into the lifeless movement of school machinery; with them it insures accuracy, effectiveness, consistency, and the proper distribution of time and energy.

### The Reactive Attitude of the Child

In discussing the relative merit of this standard with that of the preceding one, F. M. McMurry says: "Teachers, supervisors of teachers, and authors of books on teaching, have been so intently observant of the procedure of the teacher that they have overlooked that of the pupil. Yet the center of gravity of the

school lies in the pupil, and what he himself finally does determines the value of the teacher's efforts. He, therefore, should be the primary object of consideration rather than the teacher, and the quality of the instruction should be judged mainly in terms of his activity."

In conformity to this notion McMurry formulated the following criteria for the measuring of teaching efficiency: (1) Motive on the part of the child; (2) Consideration of values by the pupils; (3) Attention to organization by the pupils; (4) Initiative by the pupils.

The superiority of this standard over those previously mentioned is at once evident. It strikes right at the heart of the learning process, or as Tompkins would put it, at the spiritual unity within the child. The author of the above criteria not only believes in the theory that "the center of gravity of the school lies in the pupil", but he applies this theory daily in his classroom. Those who have attended his classes know that he practices all that he preaches.

If the pupil's reactive attitude is the key to educational direction and the goal of educational effort, as we believe it to be, it is fair to assume that it should be of paramount consideration in any attempt to determine the quality of teaching.

As principles of direction the above criteria are all that is desired. They force analysis of the teaching process, and suggest the proper distribution and emphasis of the teaching agencies. They are basic to our whole scheme of pedagogy. To abandon the principles underlying these criteria would be to ignore teaching as a profession.

Though indispensable as an agency for the improvement of teaching these criteria are decidedly inadequate as a means of determining the relative merit of teaching. Their inadequacy is due to the fact that the character of their application depends entirely upon the judgments of those who attempt to determine the merit of teaching. The necessity of interpretation introduces a decided variable.

The decisions of several judges as to the merits of a certain recitation will vary in proportion to the variation in their experiences and insight. What may seem to be "motive on the part of the child" to one observer may appear as excessive emotion to another. Indications of a "consideration of values" to one judge



may appear as a wanton neglect of essentials to another. "Attention to organization" to another observer may impress his associates as being a mere juggling of facts. Indeed, what may seem to one critic as "initiative of the pupils" may appear to another as rampant individualism. Just as the jury is an uncontrollable variable in the machinery of justice, so the supervisor as a personal judge of teaching efficiency is a variable which is exceedingly difficult to reckon with in the application of the McMurry criteria.

### Subjectiv Guides and Scales

Numerous guides and scales have been developed of recent years for estimating the work of teachers. These are valuable to the supervisor in that they force analysis of the teaching act and thereby make it possible for him to point out definitely the strong and weak points in the recitation, and afford an opportunity for him to give the teacher some practical suggestions as to the improvement of his methods.

The following "Ten-point scale" is somewhat typical of helps of this sort:

#### TEN-POINT SCALE FOR ESTIMATING CLASSROOM WORK IN HIGH SCHOOLS<sup>1</sup>

- I. "Setting" of class topics in the course.
- II. Mastery of intellectual content and effective logical organization of materials.
- III. The mechanics of classroom management. Economy of time and grasp of pedagogical technique.
- IV. Effective emphasis upon the mental processes and values peculiar and essential to the subject.
- V. Independence of teacher and class as a growth toward their material.
- VI. Suitability to the pupil of the type of recitation employed.
- VII. The "common sense" factor.
- VIII. Evidence of culture versus mere erudition.
- IX. Class participation and class sense of responsibility.
- X. Class respect for learning.

Scales of this sort do not, however, materially assist the supervisor in judging the relative results of teaching. In the

---

<sup>1</sup>A tentative scale now being prepared by Professor Charles Hughes Johnston of the University of Illinois.

application of this scale as in the application of the McMurry standards a mark variable is introduced in the judge who applies it. Furthermore, the points are not of equal significance. Some of these points are several times more significant than others. Two teachers of widely different abilities when measured by this scale may receive the same numerical mark. One may be stronger in the essentials; the other stronger in the non-essentials.

## II. OBJECTIV STANDARDS

Objectiv standards may be divided roughly into two classes: (1) *standardized tests*; (2) *standardized scales*. The former is a graded series of problems accompanied by the number of correct answers obtained by a median pupil of a widely selected group. The Curtis Standard Tests, The Kansas Silent Reading Test, and The Thorndike Reading Scale are standards of this type. The latter is an arrangement of the carefully prepared work of pupils into an evenly graded system which has been determined and evaluated by a number of competent judges. Thorndike's and Ayres' Handwriting Scales, The Harvard-Newton Composition Scales, and Thorndike's Drawing Scale are standards of this type.

A historical survey of the objectiv standards, accompanied by a discussion of their relative merit, is perhaps the easiest and doubtless the most pedagogical way of showing the relative educational value of these standards as agencies in *determining* the quality of teaching and in paving the way for placing teaching upon a scientific basis, a distinction which it does not as yet merit.

### Origin of Objectiv Standards in America

So far as I can ascertain, Dr. J. B. Rice is the father of the objectiv standard in America. Zealous for better opportunities for the child, enthused by his recent psychological studies at Jena and Leipsic, free from prejudices which sometimes result from inferior teaching, he set for his task the exposition of certain evils which he conceived to exist in the public schools. Consequently from 1891 to 1896 he became a critical student of education. He visited and examined the schools of one hundred American cities. He pointed out in the columns of the "Forum" what seemed to him remedial measures for these schools. After four years of constant investigation he came to the very decided conviction that concerted effort towards obtaining satisfactory re-

sults in public education is impossible until we know *what satisfactory results are*. "If we do not know", he wrote in the "Forum", December, '96, "what we mean by satisfactory results, how shall we be able with any degree of intelligence to judge when our task has been satisfactorily performed? Until we come to a definite understanding in regard to this matter, our entire educational work will lack direction and we shall continue as heretofore, to grope our way along passages completely enveloped in darkness in an endeavor to land we know not where.

"If we might have a standard which would enable us to tell when our task has been completed, our attention might be earnestly directed towards the discovery of short cuts in educational processes. For want of such a standard each individual teacher has thus far been a law unto himself; permitted to experiment on his pupils in accordance with his own individual educational notions, whether inherited from his grandfather or the result of his study and reflection, entirely regardless of what was being done by others. So long as this condition is possible, pedagogy cannot lay claims to recognition as a science. Until an accurate *standard of measurement* (my italics) is recognized by which such truths may be discovered, ward politicians will continue to wield the baton and educational anarchy will continue to prevail."

#### The First Objectiv Standard

Dr. Rice was not a faddist. Indeed, he was exceedingly practical. In his characteristic way he set out in 1896 to establish a standard of measurement for spelling. He undertook personally the herculean task of examining 13,000 children in spelling. This investigation extended over a period of sixteen months and included sixteen American cities.

The children were tested on *a list of words*, on *words given to them in sentences*, and on the *words used in their compositions*. The tabulated results in the "Forum" for April, 1897, is, so far as I have discovered, the first objectiv standard in spelling or in any other subject. The list of words standardized by him consists of too few words to be of service in judging the spelling abilities of children. The list of words presented in sentences is subject to the same criticism. This objection does not hold for his composition test. Had he estimated the percent of words correctly spelled in the compositions on the basis of the number of *different*

words used, instead of upon the basis of the *entire number* of words used, he would have established the first practical *objectiv standard*. As it is his percents of words correctly spelled are entirely too high.

#### Rice's Arithmetic Test

In the October number of the "Forum", 1902, Dr. Rice reported the results of an arithmetic test which he had conducted in seven different cities, including eighteen buildings and 8,000 children. As Stone pointed out, later, Dr. Rice's results were not satisfactory as a standard, due to certain limitations in the problems used and the character of the methods employed in gathering and scoring these.

#### Rice's Language Test

One year later Dr. Rice gave a detailed report of the test he made in language. This test extended to nine cities, and included twenty-two schools, containing 8,300 children. The compositions were arranged in five groups on the basis of relative merit. The papers of each group were graded 100%, 75%, 50%, 25%, 0% respectively. The results show conclusively that there was a wide variation in the English abilities tested by him, but owing to the strong probability of error in his results they have not been employed as a standard for determining English ability.

Though Dr. Rice's results are of little value as standards, his experiments have stimulated two lines of research in education which are fraught with wonderful possibilities. I refer on the one hand to the investigations which have had for their goal the establishment of *objectiv standards* of measurement, and on the other to the investigations to determine *minimum essentials*. Both of these problems were raised by Dr. Rice and he has lived to see some partial solutions of both.

#### The Cornman Spelling Standard

Dr. O. P. Cornman, of Philadelphia, stimulated by the work of Rice, carried on a series of tests in spelling by the composition method, extending from June, '96, to June, '98. In 1903 he published the results of this investigation in a volume entitled *Spelling in the Elementary School*. Cornman's data were carefully gathered and the results methodically tabulated. He substituted the median for the average employed by Rice. In his composition test Rice counted all the words which were spelled

correctly, including all recurring words when properly speld. When a misspeld word recurd he counted it but once. On this basis of counting he determind the spelling abilities of the children in terms of the percentage of words speld correctly. This accounts for the high percentages which he reported. Cornman counted all words in the composition and determind the ratio of the speld words and misspeld words in terms of percent. He not only counted the recurring words which were speld correctly but the recurring misspeld words as well. This accounts for his percentages being lower than those reported by Rice.

The work of Rice and Cornman stimulated many young men in the large educational centers. Edward L. Thorndike, who has since become the wizard of the objectiv standard, wrote in the "Forum" in 1905 as follows: "The study of education is begining to be quantitativ, we are becoming properly disgusted with the one-sided booking which only takes account of dollars spent and neglects the debit side, the income in knowledge, habits, power, zeal and ideals. This ambition toward an exact objectiv measurement of the results of educational endeavor is a symptom of helthy scientific fervor and also of common sense wisdom. No one possesst of science or sense will deny the value of successful quantitativ study of school work."

#### **Arithmetic Abilities of Children in the Sixth Grade (Stone)**

In 1908 C. W. Stone publisht in the Columbia University *Contributions to Education* a report on the arithmetical abilities of children in the six-A grade. Mr. Stone personally conducted the examinations in twenty-six school systems, including seventy-nine schools and 6,000 children. He gave one test in the fundamentals and one in the reasoning processes.

Stone's method of gathering data and of tabulating results was superior. He set a standard in this particular which has been emulated by later investigators. The exercises in the tests proved, as Courtis pointed out, too complex for practical mesurements. The results were a mesure of a combination of abilities in the fundamentals and in the reasoning processes, and consequently were difficult of interpretation and application. Because of the difficulty of applying his results they have not been used extensivly in determining arithmetical abilities.

### The Thorndike Handwriting Scale

The first satisfactory result from a practical point of view of all the agitation for quantitative standards of measurement occurred in 1910. The Thorndike Scale for Judging Handwriting appeared in the *Teachers College Record* of that year. Referring to this scale, Ayres says, "The credit of developing the first measuring scale for handwriting belongs to Professor Edward L. Thorndike of Teachers College, Columbia University. The publication, in March, 1910, of his Handwriting Scale constituted a most important contribution not only to experimental pedagogy, but to the entire movement for the scientific study of education."

In reference to the need of such a scale Thorndike said, "At present we can do no better than estimate a handwriting as very bad, good, very good, or extremely good, knowing only vaguely what we mean thereby, running a risk of shifting our standards with time, and only by chance meaning the same by a word as some other student of the facts means by it. We are in the condition in which the students of temperature were before the discovery of the thermometer, or any other scale for measuring temperature beyond the very hot, hot, warm, lukewarm, and the like, of subjective opinion."

Altho, as Ayres pointed out, this Handwriting Scale constituted a most important contribution not only to experimental pedagogy but to the entire movement for the scientific study of education, Professor Thorndike in his presentation was sensitive of its imperfections. He says: "The scale is presented now in spite of its imperfections, for these reasons. It is the result of some twenty ratings, and ensures measurements far more accurate than anyone could make without it. For the present needs of school practice and educational research, a very precise instrument for measuring handwriting is not required. The best way to get a more perfect scale is by the use of this one as a starting point."

The Thorndike scale represents types of the handwriting of children of grades five to eight inclusively. The writing from these grades was grouped into eleven groups on the basis of quality. The quality of the groups is represented by figures 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17 respectively. Quality 7 represents the poorest samples taken from grade five and quality 17 represents the best samples taken from grade eight. The steps

of difference between the qualities were equal in the sense of being cald equal by from twenty-three to fifty-five competent judges. This means "that 14 is as much better than 13 as 13 is than 12; that 13 is as much better than 12, as 12 is better than 11, and so on; that quality 14 is two times as far above zero merit in handwriting as quality 7.

The scale includes quality 18, which was taken from a copy book, and qualities 4, 5, and 6. Samples 5 and 6 were taken from the fourth grade and sample 4 was manufactured for the purpose of extending the scale below the merit of fourth-grade children.

The Thorndike Handwriting Scale is easily applied in testing the quality of handwriting. After a little experience a teacher can scale the writing of her entire room in a very short time. By means of such a scale we have often mesured the writing of an entire room in less time than a forty minute period. The several samples supplied for each of the qualities 16, 15, 14, 13, 12, 11, 9, and 8 make it especially easy to apply this scale.

Teachers who habitually think of quality in terms of grades can, for all practical purposes, easily transfer the qualities of the scale into grades by multiplying the numbers of the scale by 5.8. Those who have mesured the merit of handwriting with this or the Ayres' scale will not be content to judge the merit of writing in terms of personal experience.

#### **The Ayres Handwriting Scale**

In November, 1911, Leonard P. Ayres of the Russel Sage Foundation began a preliminary experiment to determin the relativ legibility of different samples of handwriting. He early concluded that the scheme was feasible and proceded to perfect a writing scale on that basis. His first printed scale appeard in February, 1912. In discussing the merits of this scale he says: "The method by which the present scale has been produced, and the criterion on which it rests as a basis differ radically from those adopted by Professor Thorndike. The difference in the basis is that in the present case legibility has been adopted as a criterion for rating the different samples in place of 'general merit' used as the basis of Thorndike's scale. The change substitutes function for appearance as a criterion for judging handwriting."

Ayres gathered 1,578 samples of writing from forty school systems. The samples were read by ten readers, each of whom by means of a stop watch recorded the exact number of seconds required to read each sample. The samples were then placed in eight groups on the basis of the time required to read them. The following table shows the rating of a type sample of each group.

*Table I*

Point on scale	Rating in words read per minute, of sample found at each point
90% .....	209.2
80% .....	202.7
70% .....	195.1
60% .....	186.2
50% .....	175.7
40% .....	163.4
30% .....	149.1
20% .....	132.2

The scale was divided into three longitudinal divisions on the basis of slant. The top, or *A* division, contains the vertical samples. The middle or *B* division, contains the samples of medium slant, and the lower, or *C* division, contains the samples of extreme slant. As implied in the above table the scale is divided into eight vertical divisions, each of which contains a sample of each slant. The three samples in the right column are marked 90%, those in the next column to the left 80%, etc.

Because of its inclusion of samples representing the three main types of slant, this scale is easily applied. The application of this scale to the handwriting of most school systems at once reveals wide variation in writing abilities, which implies either widely different methods of teaching, widely different ideals as to the sort of writing which should obtain, or widely different degrees of zeal towards securing good writing. The following graph (Figure I) of the writing abilities of the children of the Training School of the Illinois State Normal University, as shown by the first application of the Ayres scale, reveals the sort of variation which frequently exists when subjective standards alone are relied upon.



The application of the Handwriting Scale not only revealed wide variation within each grade, but it revealed wide variation between the grades as well. This first application of the scale showed that there were two children in the sixth grade who wrote better than any of the children of the seventh and eighth grades. It showed also that there were six children in the sixth grade who made a grade of 70 while there were but four children in both the seventh and eighth grades who reached the 70 mark. This test made it perfectly evident on the one hand that grades five and six needed no extra consideration relative to drill in writing, while on the other hand it showed that grades seven and eight needed a writing revival.

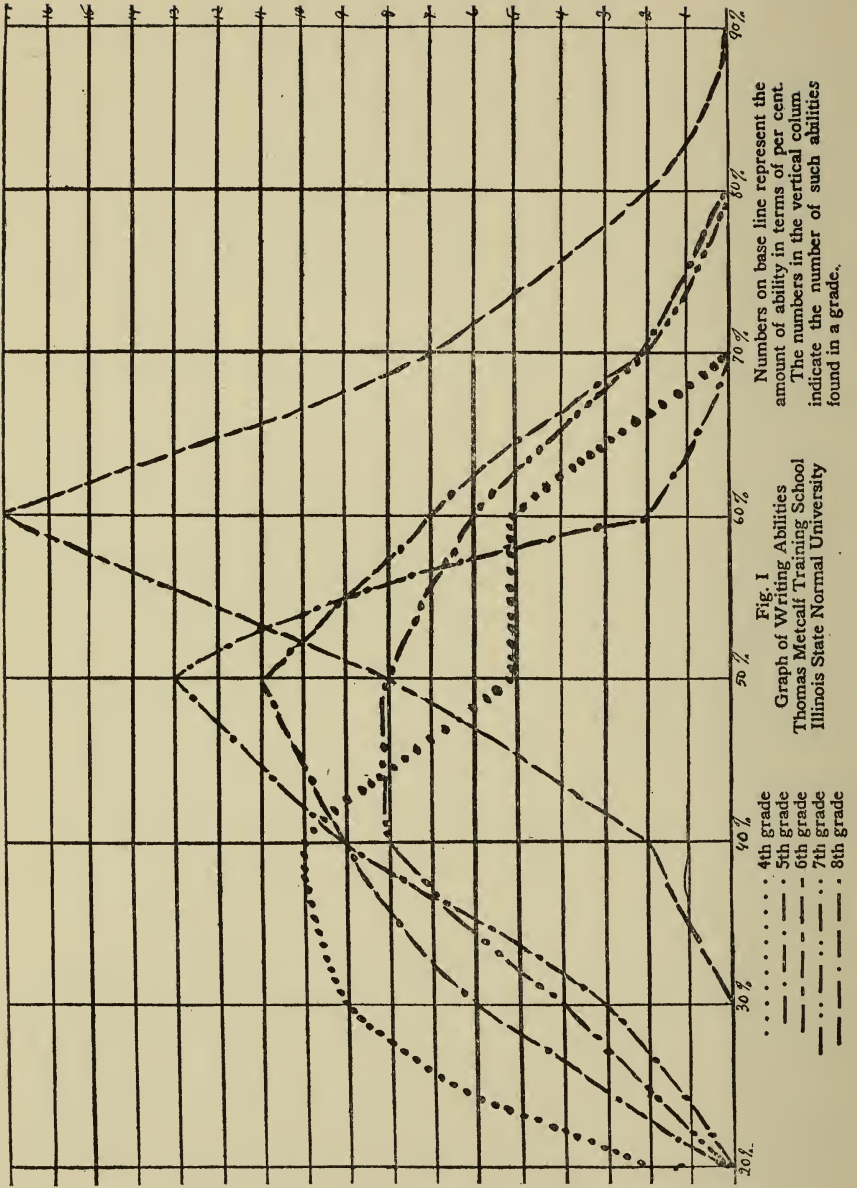
The graph (Figure II) shows what was accomplished by the eighth-grade teacher after he became conscious of the relative needs of his pupils. In the November test fifteen pupils made grades of 40% or less. In the May test none made a grade less than 50%. In the November test only two pupils made grades of 70% while in the May test ten pupils made grades of 70%, ten pupils made grades of 80%, and three pupils made grades of 90%. A careful examination of Figure II will reveal other marked changes which resulted from an application of the Handwriting Scale.

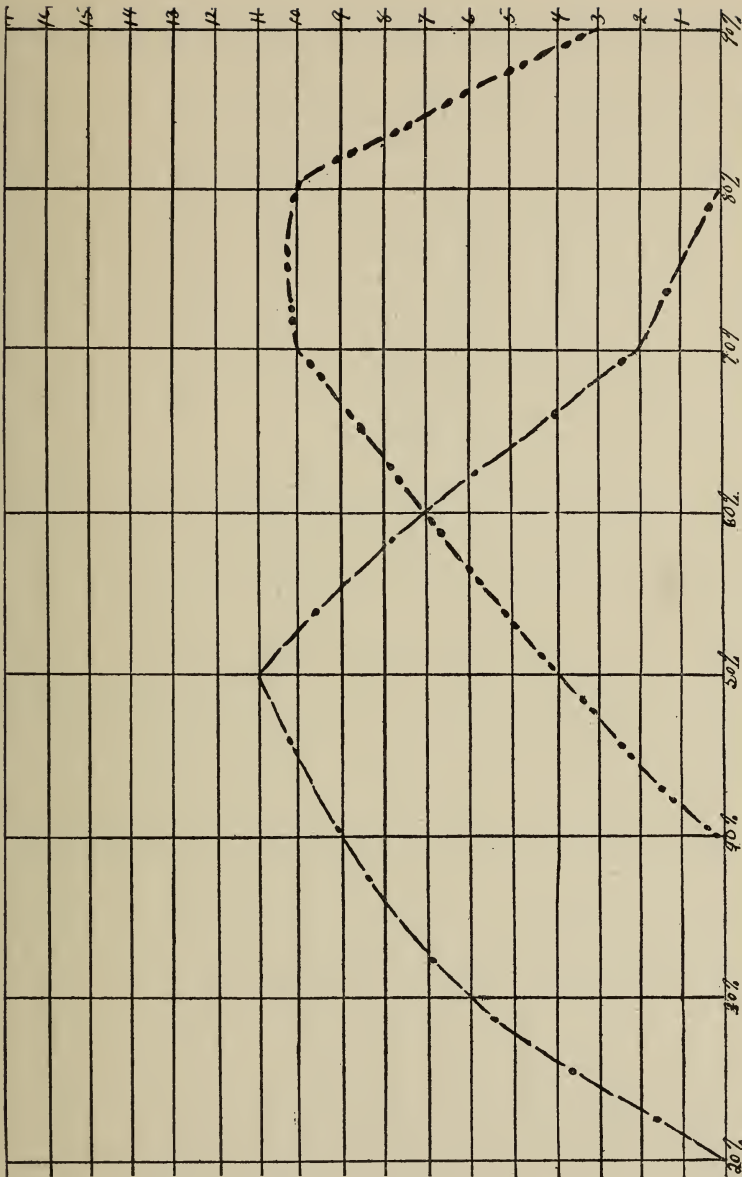
A record of the writing ability of the children of grades four to eight inclusive, taken in November and May, and filed for reference becomes a definite and valuable guide for any school. It makes it possible to determine at any time whether a sufficient amount of time and energy is being given to this subject. Such a record protects the children from the excessive zeal or the indifference of the teacher, and indicates to the teacher the relative merit of his endeavor.

#### **Starch's Letter-Exposure Handwriting Test**

Professor Daniel Starch of the University of Wisconsin, reported his handwriting test in the *Journal of Educational Psychology* for October, 1913. He pointed out that the Thorndike and Ayres scales were measures only of form and legibility respectively. He argued that a simple analysis of handwriting shows that its three chief elements are *legibility*, *producibility*, and *form*.

Starch held that legibility can best be determined by reading





Figures on the base line represent abilities in per cent. The figures in the vertical column indicate how many possess such abilities.

Fig. II

—•—•— Eighth grade abilities in handwriting October 10, 1913.  
 - - - - - Abilities of the same eighth grade, May 8, 1914.

exposed areas of handwriting and thereby determining the average rate per letter of such reading. In conformity with this theory he prepared a device for measuring handwriting as follows: In a piece of cardboard were cut three circular openings in a straight row 1.5 cm. apart. The openings were each 2.5 cm. in diameter. By shifting the cardboard about over the writing to be measured, he was able to test its legibility at several places. The number of letters exposed and the time required to read them were recorded after each trial. From the records of several exposed areas the average reading per letter was computed.

Starch's experiments proved that there is a remarkably close correlation in the results obtained by the Letter-Exposure Test and those secured by the Thorndike and Ayres scales.

It is doubtful if the Letter-Exposure Test is as convenient for testing the handwriting of large numbers of children as is either the Thorndike or Ayres scales.

After testing the efficiency of writing scales Starch says: "We may conclude that after some practice in the use of a scale the measurements with either scale are from three to four times as accurate as the valuations made by the usual percentil marking system."

#### **The Courtis Standard Tests**

In December, 1910, W. S. Courtis, of Detroit, reported in *The Elementary School Teacher* his Standard Test (Series A) in Arithmetic. This test developed as a result of applying the Stone test in the Detroit Home and Day School, in which Mr. Courtis was head of the Department of Science and Mathematics. After a free use of his Series A Test, which consisted of testing the pupils' ability to use the four fundamental processes when employed in tables ordinarily used in schoolrooms, and of testing the pupils' ability to employ the reasoning processes involved in the solution of problems suitable to the grammar grades, Mr. Courtis concluded that "The work done with Series A has proved that the basic problem in education to-day is that of ministering adequately to individual needs. The first step towards this end is the formation of definite objective standards." The standards derived from the use of Series A, however, are either complex or of questionable value, owing to the uncertainty of their meaning.

This is particularly true of the reasoning tests in which mere ability to read is a large factor.

Series B is the result of an attempt to secure definite objective standards for each of the four fundamental operations with whole numbers. With the establishment of this standard it is possible to set for each grade just the degree of skill in each of the fundamental processes that is within reach of the average, or median, child of the grade.

The following table shows the median skills of three distinct groups of children in the fundamentals of arithmetic provided in the Curtis test. The approximation of the series reveals the universal character of the results.

Table II

		5th grade			6th grade		
		D.	B.	G.	D.	B.	G.
Addition .....	A	6.7	7.2	7.1	8.4	8.3	8.
	R	3.9	3.7	3.7	4.6	4.9	4.4
Subtraction .....	A	8.	7.6	6.5	8.8	9.	8.9
	R	5.5	4.9	4.9	6.2	6.3	6.1
Multiplication .....	A	6.	5.8	6.	7.4	6.9	7.2
	R	3.8	3.3	2.6	4.8	4.8	4.5
Division .....	A	4.9	4.5	4.5	6.4	5.5	5.8
	R	2.7	2.	2.3	4.4	3.3	4.3
		7th grade			8th grade		
		D.	B.	G.	D.	B.	G.
Addition .....	A	9.2	9.2	8.9	10.2	11.	9.7
	R	3.4	5.6	4.7	6.7	7.5	5.6
Subtraction .....	A	9.8	10.	10.2	12.3	11.4	11.7
	R	7.3	6.9	7.8	9.5	8.6	8.4
Multiplication .....	A	9.6	8.	8.4	10.5	9.5	9.7
	R	6.	5.1	5.2	7.	6.5	6.4
Division .....	A	8.6	6.9	7.6	10.6	6.9	7.6
	R	7.1	5.1	5.1	8.8	6.9	6.3

D = Detroit (1,315 children tested)  
 B = Boston (20,441 children tested)  
 G = General (3,618 children tested)  
 A = Number of problems attempted  
 R = Number of problems right

Courtis early discovered the value of the objective standard in determining individual variation. He says: "The results of the tests disclosed the usual wide range of individual variation in every grade." After a use of the objective standard for some time Professor Courtis writes: "Not only did the variabilities decrease, but un hoped degrees of accuracy were attained."

The following graphs of the abilities of intermediate pupils in multiplication and oral reading as determined by the Courtis and Gray scales show conclusively how variability is easily detected by the application of objective standards.

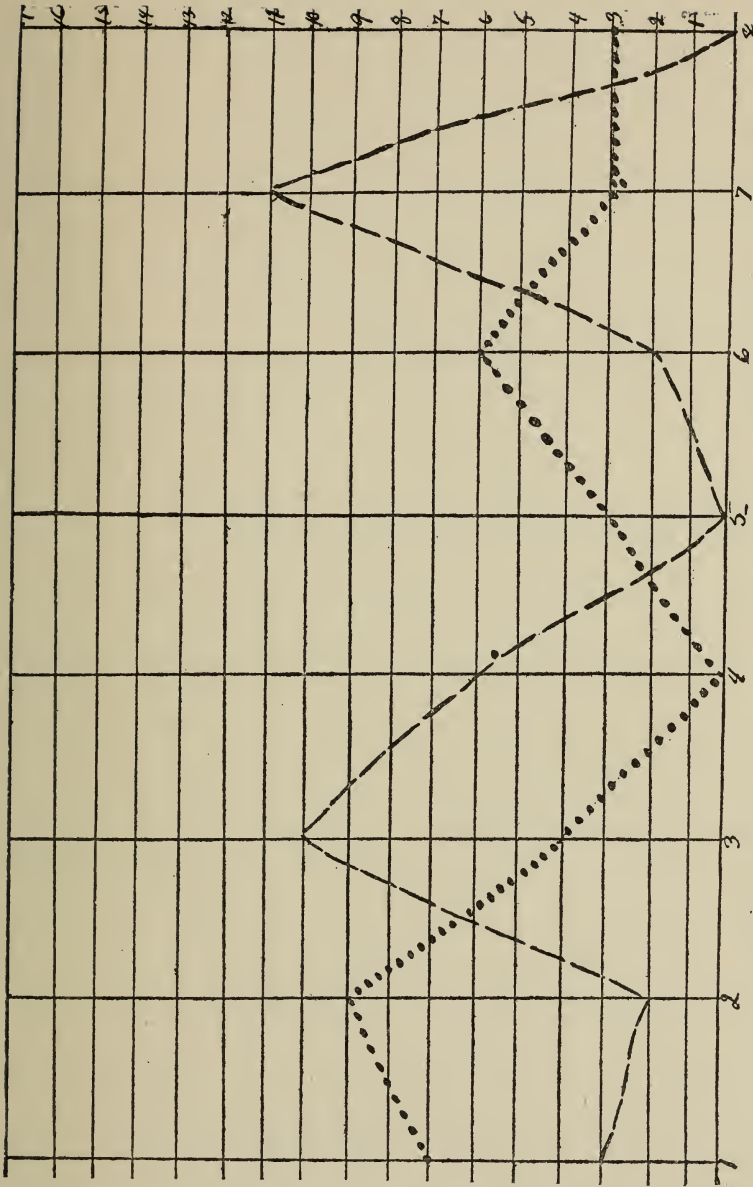
The graphs shown in Figure III reveal two distinct groups of abilities in each subject. This may mean that little care has been given to promotions. It is more likely to indicate a lack of sufficient drill under proper conditions. After the abilities are once revealed there is every reason to believe that a conscientious teacher will raise the abilities of the lower group and thereby reduce the degree of variability.

Just as a proper diagnosis in medicine is a prerequisite to effective medical treatment, so a proper diagnosis of the specific abilities of pupils is a prerequisite to the application of proper methods.

#### **The Hillegas Scale for the Measurement of Quality in English Composition**

In September, 1912, Professor M. B. Hillegas published his composition scale in *The Teachers College Record*. In the introduction to this scale Professor Hillegas refers to the previous efforts at quantitative standards by Cornman, Rice, Stone, and Thorndike. He does not, however, refer to Rice's pioneer effort to establish a standard in English composition in 1902.

Hillegas used a method similar to the one Thorndike used in determining quality in handwriting. He, aided by one other person, graded about 7,000 compositions into ten classes. From these ten classes seventy-five samples were chosen. Artificial samples were employed at the extremes of his scale, as they were in Thorndike's writing scale, in order to produce a scale of wide range of measurement. In all there were eighty-three samples employed. These eighty-three samples were given to more than one hundred persons, who were requested to rank them 1, 2, 3, etc., in the order of their merit.



Figures on base line represent units of ability.  
 Figures in vertical column indicate the number of persons having such abilities.

Fig. III

..... A graph of abilities in oral reading.  
 --- A graph of abilities in multiplication.

Owing to misunderstandings and errors, only seventy-three records were used. On the basis of like characteristics these records were reduced to twenty-three. This reduced number of samples contained all the important steps in quality from the poorest to the best. Six other samples, including two artificial ones, were finally added, making a total of twenty-nine samples.

The twenty-nine samples were ranked by 234 judges. On the basis of this ranking the number of samples was reduced to ten. The difference between the merit of the first and second samples in the scale is not identical with the difference in merit of any other two successive samples. These differences, however, are sufficiently equal for practical purposes.

The Hillegas scale is a meritorious piece of work. It is a decided step in the right direction. The brevity of the samples and the gradual gradation from one quality to another makes its application from this point of view quite easy. The Hillegas scale, tho a meritorious piece of work, has many defects. Commenting upon the Hillegas scale, Frank W. Ballou of the Department of Educational Investigation and Measurement of the Boston Schools says: "An experiment with the Hillegas scale showed that the use of such an objective measure did unify the grades given to compositions by teachers. It was also found, however, that the Hillegas scale was not satisfactory to the teachers of Newton, owing to what seemed to them to be inherent faults. These faults may be stated briefly as follows: first, the scale aims to measure too varied a product; second, the compositions in it are not typical of good school work—(a) some are artificial, (b) others are 'bookish', really reproductions, and (c) no conversation is contained in any of them." As Curtis's practical tests in arithmetic grew out of an attempt to use the conclusions of Stone, so an attempt on the part of the teachers of Newton, Mass., to use the Hillegas scale led directly to the practical Harvard-Newton Scales for the Measurement of English Composition.

#### **Report of Superintendent Bliss on English Composition**

While at Elmira, N. Y., Superintendent Bliss reported in the Psychological Clinic for March, 1912, a series of tests he had carried on in composition. He had the children reproduce stories read to them. These reproductions were taken to the central office and grouped, on the plan practiced by Rice, into five groups. He



determind the median ability for all of the children in each of the grades above the third. He then reported the median ability for all of the children of that grade in the city with the median for the particular grade in the school. He also publisht sample compositions of each group of compositions in the scale.

The results obtaind from the use of this scheme were little less than marvelous. He says: "In a Massachusetts school system, with 33 third-grade teachers the initial test showd a city average of 8.5 points, with twenty-three classes below the requirement and eight classes above. One year later the city average was 19.2 points, with thirteen classes below the requirement and nineteen classes above. This represented an increase of 126% in the level of efficiency in the third grade." Mr. Bliss cites other cases where even greater percents of increase were made by the use of this method.

#### **The Harvard-Newton Scales**

These scales are the product of the work of the eighth-grade teachers and the elementary-school principals in the public schools of Newton Mass., assisted by the teachers of English in the high schools of Newton, and by teachers and principals in Arlington, Mass., and Boston, under the direction of Frank W. Ballou and with the co-operation of the Joseph Lee Fellow for Research in Education.

The compositions were written by the eighth-grade pupils of Newton. All of the compositions of the eleven grade schools were groupd into five groups. Each group included specimens of a given type of composition (narration, description, etc.). Each eighth-grade teacher selected 25% of the compositions of her grade on the basis of their representativ merit. These selected compositions from the eleven schools were then arranged into four groups. Twenty-four readers were instructed to arrange the themes in each group in the order of their merit and to arbitrarily rate the best theme 95% and each of the remaining themes with reference to this standard. These ratings were tabulated and the median grade for each composition was workt out. For example, the highest grade for composition number one was 95%, the lowest grade was 68%, and the median grade was 83%. In like manner tabulation was made of the distribution of the

ranks given each composition. They were then arranged in serial order according to the median ranks, beginning with the highest. By means of this latter method it was discovered that 25% of the judges were radical in their judgment. Consequently the 25% of radical readers were cut off. The scale was then built on the median percentil basis. Out of the twenty-five compositions which were chosen to represent each form of discourse, six typical compositions were finally chosen for the scale. The difference in degree of quality was carefully worked out and the samples were arbitrarily marked 95%, 85%, 75%, 65%, 55%, and 45%, respectively.

The Harvard-Newton Scales<sup>1</sup> commend themselves to the practical school man on the following points: first, there is a scale for each form of discourse; second, the compositions in the scale are the real productions of children and not "built up" compositions for purposes of securing gradation in the scale; third, each scale consists only of six types. This makes it an easy matter for the person doing the grading to familiarize himself with the scales. The greatest weakness in these scales lies in the fact that they are best suited for eighth-grade pupils.

An application of these scales reveals the fact that there is but slight variation in the grades of two or more judges. Indeed, the variation is so slight that a single investigator can feel reasonably certain that his grades will not vary widely from the median of several judges.

In our opinion, the Harvard-Newton Scale ranks for practicability alongside the Thorndike and Ayres handwriting scales, and the Curtis Tests in Arithmetic (Series B). It has the real ring to it and will doubtless have a wide use.

#### **The Curtis Test in English**

Professor S. A. Curtis has five different tests in English: I, Handwriting Test; II, English Composition Test; III, Spelling, Punctuation, and Grammar Test; IV, Rates of Reading and Writing Test; and V, Rates of Reproduction Test. In his writing test Mr. Curtis uses four groups of letters with five in a group in each of ten lines. Pupils are required to copy these as rapidly as they can and maintain a good quality. The speed of each child is recorded and the quality of the writing is measured by the

---

<sup>1</sup>The Harvard Press, 50 cents

Thorndike and Ayres scales. Thru the co-operation of teachers Mr. Courtis hopes to establish a standard test in both speed and quality for each grade.

Mr. Courtis bases his English composition standard on an original story, "Bessie's Adventures", parts of which are real while other parts are imagined. His method of determining the relative merit of compositions is the same as that used by Dr. Rice. Teachers are requested to group these original stories into five groups, on the basis of merit. From each of these groups they are requested to select a sample and return such samples to him. In this way he hopes finally to establish a standard of English abilities in the several grades, similar to those he has determined in arithmetic. His other English investigations follow a similar procedure. All of his tests may be had in his "Manual of Instructions for Giving and Scoring the Courtis Standard Tests."

Mr. Courtis has not presented the exercises in his English tests so clearly and attractively as he presented those of his arithmetic tests.

#### **The Thorndike Scale for Measuring Achievement in Drawing**

In the *Teachers College Record* for November, 1913, Professor Thorndike presented a scale for "The Measurement of Achievement in Drawing". In reference to the purpose of the scale he says: "It is the purpose to present a provisional scale by which achievement and improvement in drawing can be measured with somewhat the same clearness, exactness, and commensurability as achievement and improvement in lifting weights."

The same general method which was used in determining the Thorndike Handwriting Scale and the Hillegas Composition Scale was employed in the making of this drawing scale. Forty-five drawings of children were first submitted to a number of critics whose ratings reduced the number to a series of fifteen drawings graded from zero up.

This series of fifteen drawings was rated by 376 persons, of whom sixty were artists of distinction, eighty were supervisors of art, and 236 were students of education and psychology.

The unit of the scale was *one merit*. This unit is "The difference of merit in children's drawings which 75% of artists, teachers of art, and intelligent judges generally can distinguish,

and which 25% of them fail to distinguish." The drawing lowest in the scale was judged of zero merit. The difference of merit between two drawings is not necessarily a unit merit. It depends upon the relative number of judges who considered one drawing better than the other. If 75% of the judges considered one drawing superior to another the difference in quality is called a unit merit. If less than 75% of the judges distinguished a difference in merit between two drawings the difference between the two is less than one merit. If more than 75% of the judges discerned a difference in merit the difference in quality was marked more than one merit. The following is the determined rating:

Table III

Drawing 1=0 merit	Drawing 8=10.5 merit
Drawing 2=2.4 merit	Drawing 9=11.8 merit
Drawing 3=3.9 merit	Drawing 10=12.6 merit
Drawing 4=5.7 merit	Drawing 11=13.5 merit
Drawing 5=6.5 merit	Drawing 12=14.4 merit
Drawing 6=7.8 merit	Drawing 13=16 merit
Drawing 7=8.6 merit	Drawing 14=17 merit

The reader should see the drawings in the *Teachers College Record*, which accompany these merit values.

No one is more conscious of the limitations of this scale than is Professor Thorndike. In spite of its limitations it is a valuable contribution to experimental education. The method of attack, the care employed in determining differences in merit, and the scientific attitude of the author in the whole procedure will have a wholesome effect upon investigators. It is as practical in determining the qualities of children's drawing as are the writing scales in determining the quality of handwriting. It would better meet the needs of the schools if it attempted to measure the various aspects of children's art instead of a single aspect. It is to be hoped that it will be followed by other "drawing scales" which are adapted to measure the various aspects of children's drawings.

#### The Thorndike Reading Scale A: Visual Vocabulary

Thorndike's Reading Scale A for visual vocabulary appeared in the *Teachers College Record* for September, 1914. In presenting this scale Professor Thorndike states that there are four phases of reading ability which should be measured: "(1) A pupil's

ability to pronounce words and sentences seen; (2) a pupil's ability to understand the meaning of words and sentences seen; (3) a pupil's ability to appreciate and enjoy what we roughly call 'good literature'; and (4) a pupil's ability to read orally, clearly, and effectively."

The following scale in conjunction with the silent reading tests perfected by both Kelly and Gray, given later in this report, is an adequate measurement of number (2) above. Gray's scale for the measurement of oral reading provides for number (1) above. Professor Thorndike says that he is working on scales to measure (3) and (4). It is hoped that these scales will soon be developed.

*Thorndike Reading Scale A: Visual Vocabulary*

Write your name here .....

Write your age here ..... years ..... months

Look at each word and write the letter F under every word that means a *flower*.

Then look at each word again and write the letter A under each word that means an *animal*.

Then look at each word again and write the letter N under each word that means a *boy's name*.

Then look at each word again and write the letter G under each word that means a *game*.

Then look at each word again and write the letter B under each word that means a *book*.

Then look at each word again and write the letter T under each word like *now* or *then* that means something to do with *time*.

Then look at each word again and write the word GOOD under every word that means something *good to be* or *do*.

Then look at each word again and write the word BAD under every word that means something *bad to be* or *do*.

4. camel, samuel, kind, lily, cruel
5. cowardly, dominoes, kangaroo, pansy, tennis
6. during, generous, later, modest, rhinoceros
7. claude, courteous, isaiah, merciful, reasonable
8. chrysanthemum, considerate, lynx, prevaricate, reuben
9. ezra, ichabod, ledger, parchesi, preceding
10. crocus, dahlia, jonquil, opossum, poltroon
- 10.5 begonia, equitable, pretentious, renegade, reprobate
11. armadillo, iguana, philanthropic

### The Kansas Silent Reading Test

Dean F. J. Kelly of the School of Education, University of Kansas, while director of the Training School in the State Normal at Emporia, developed and standardized The Kansas Silent Reading Test. This test will appeal to practical school men. It is definite, simple, and easily presented. The results can be quickly and definitely determined. In practicability it ranks with the Thorndike and Ayres Handwriting Scales, and Curtis Arithmetic Tests (series B), The Harvard-Newton Composition Scales, Thorndike's Reading Scale and the Ayres Spelling Scale.

The entire test consists of carefully graded groups of exercises; one for the primary grades, one for the grammar grades, and one for the high school. The following exercises are chosen from the sixteen exercises listed in the test for grades three, four, and five.

	.....	
	.....	
		No. 1
Value		Mary is older than Nellie, and Nellie is older than
2.1		Kate. which girl is older, Mary or Kate?
		.....
		.....
		No. 9
		It was a quiet, snowy day. The train was late.
		The ladies' waiting room was dark, smoky and close,
		and the dozen women, old and young, who sat wait-
		ing impatiently, all looked cross, low spirited or
		stupid.
Value		
4.9		In this scene, the women probably kept their
		wraps on, because they wished to be ready to take the
		train. Pretty soon the station agent came and put
		more coal in the stove, which was already redhot in
		spots. Do you think this made the women happier?
		.....
		.....

No. 10

Below are three lines. If the first is the shortest, place a dot above it. If the last line is shorter than the first but longer than the middle line, put a cross above the longest. If each of the other lines is longer than the last line, put a cross above the shortest line.

---

### The Gray Reading Tests

These tests were developed by Professor William S. Gray, now in the School of Education, University of Chicago, while a graduate student at Columbia and Chicago. In an endeavor to determine certain facts concerning reading achievement, rather than in an attempt to devise a test *per se*, this scale was worked out by Mr. Gray. The exercises employed consist of carefully graded selections. Those for the oral reading test increase in difficulty of interpretation. This test is not so easily operated as is the Kansas Silent Reading Test.

The oral test is designed to measure abilities in pronunciation, omissions, insertions, substitutions, and repetitions. The silent test is intended to measure the pupil's ability to determine the thought essentials in a series of reading exercises.

Already a sufficiently large number of children have been tested to determine a pretty safe standard of the median abilities of the children in grades three to eight inclusive. It is to be hoped that this scale will be put in a suitable form and soon be made accessible to teachers.

### The Ayres Spelling Scale<sup>1</sup>

A scale for measuring ability in spelling prepared by Dr. Ayres was determined from data consisting of 1,400,000 spellings by 7,000 children in 84 cities throughout the country. The words in the scale are 1,000 in number. These words are arranged in columns on the basis of their difficulty. All the words in each column

<sup>1</sup>Single copies of the Ayres Spelling Scale and of the Ayres Handwriting Scale may be had for 5 cents each, by addressing the Russell Sage Foundation, New York City.

have practically the same difficulty. The scale shows the percent that the median child of each grade should make on each column of words. For example, the median child in the third grade should spell correctly 58% of the words in column 14. The median child in a fourth grade should spell correctly 79% of the words in the same column. Median abilities are indicated in like manner for the other grades.

(The practicability of this scale is characteristic of Dr. Ayres' contributions to the science of education.) It is very satisfactory for determining the spelling abilities of children. Indeed, it is quite doubtful if there will be any improvement upon this scale for the measurement of spelling abilities in the near future.

#### **The Composition Method of Testing the Spelling Abilities of Children**

It will be remembered that both Rice and Cornman used the composition method of determining the spelling abilities of children. The abilities as shown by these investigations were so high that practical school men considered them worthless as standards.

The high grades reported by both Rice and Cornman were due to the methods employed. Rice found the ratio between all of the words spelled correctly (including duplicate words) and the misspelled words (duplicate misspelled words not counted). This method produced a low percentage of error. Cornman attempted to correct this error by counting all duplicate misspelled words as well as duplicate words which were spelled correctly. As is evident this method slightly increased the percentage of error in spelling.

The error in both methods resulted from the fact that both Rice and Cornman did not recognize that children duplicate a larger proportion of words which they can spell correctly than of words which they misspell. There are at least two reasons for this: first, there is a natural tendency to use freely words which one is confident he can spell and to avoid the use of words difficult to spell; second, there are a number of easily spelled words such as in, on, and, the, so, for, is, etc., which make up the major portion of the duplicated words.

If the above reasons are sound it is evident that one's spelling grade is raised by increasing the number of repetitions when measured by the Rice and Cornman plans. Since children necessarily repeat a large number of simple words it follows that the spelling grades of children will be too high when tested by the Rice-Cornman methods.



Because I believed that a spelling standard based upon the composition method is the only standard that is reliable for daily use in the school room, I began to gather data in the spring of 1915, for the purpose of determining a composition standard of spelling which is free from the manifest errors in Rice's and Cornman's conclusions. Instructions were sent to a number of superintendents and principals who had previously manifested a willingness to assist in this investigation. So far thirteen schools have reported. These instructions were to the effect, first, that all duplicate words and the words *I* and *a* in the compositions should be crossed out; second, that of the words not crossed out the ratio of the words spelled correctly to those misspelled should be expressed in percent.

Thirteen schools returned papers properly marked. The results from eleven of these schools have been tabulated, and the median ability for each grade determined as follows:

*Table IV*

Median Spelling Abilities of Eleven Schools as  
Determined by this Composition Method:

3rd grade	4th grade	5th grade	6th grade	7th grade	8th grade
91%	93.6%	95.5%	96.6%	96.9%	98.2%

The Median Spelling Abilities Reported by Cornman:

3rd grade	4th grade	5th grade	6th grade	7th grade	8th grade
94.6%	96.5%	97%	98.1%	98.9%	99.5%

A comparison of the two tables reveals a decided difference in the two results. This is greater than the tables indicate. Our instructions were to give the test to the best school in the city. These instructions were given with the thought that a standard to be of real value should represent abilities determined under most favorable circumstances rather than under mediocre circumstances. It is quite probable that the median abilities shown in our report (Table IV) are decidedly higher than medians which would be obtained from testing all of the children in the cities where these schools were located.

Table IV is but a tentative report of this investigation. Additional data and a more critical examination of the various papers reported are necessary before the reliability of these results can be depended upon. It is very probable, however, that additional data will show but slightly changed median abilities of the several

grades with the single exception of the third grade. There is evidence that this mark is too low.

There is a prevailing notion abroad in educational circles that objectiv standards can be used only in mesuring the skills of pupils. Persons who hold this notion argue that since these standards mesure skill only, the results of such measurements are of little value in determining the relativ merit of teachers. They further argue that since the objectiv standards mesure form and not content, any markt attention given to this sort of measurement will result in an over emfasis of form at the expense of content.

These arguments are based upon two fallacies: (1) It is fallacious to assume that only skill can be mesured by the objectiv standard. It is true that standards for the mesurement of skill were determind first. Standards for the mesurement of abilities to reason, to enjoy, and to appreciate are following. The Kansas Silent Reading Test and the Gray Silent Reading Test are both standards of the latter type. (2) It is fallacious to assume that attention to the mesurement of such abilities as the fundamentals in arithmetic, handwriting, spelling, form in reading, etc., will result in an over emfasis of the formal subjects to the detriment of the content subjects. This would not be fallacious were it not true that grades far above the median indicate an undue emfasis upon the subject taught and consequently are a mark of poor teaching.

It must be rememberd that an application of a standard test will detect an undue emfasis of some particular subject-matter as well as an insufficient emfasis of it.

It is excedingly important that the interest of the school men of the State of Illinois be elicited in support of a movement to apply the objectiv standard more generally. We should have Illinois standards for the various abilities which can now be definitely mesured.

I would suggest that a bureau be establishd by the State Teachers Association, or in connection with the Department of Public Instruction, the State Normal Schools, or the School of Education at the University of Illinois, for the direct purpose of preparing and distributing these tests and for the purpose of tabulating and distributing the results. Any one of these branches of the public school system of the state should be and, I believ, is willing to undertake this work if it is the wish of the school men of the state to have it done.





3 0112 105727298