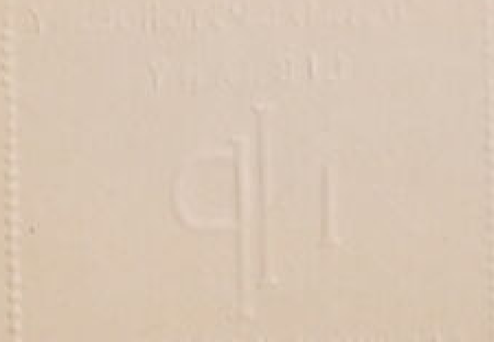


DIFFERENTIAL DIAGNOSIS OF ABILITY IN SCHOOL CHILDREN

Differential Diagnosis of Ability in School Children

BY

DAVID SEGEL

Educational Consultant and Specialist in Tests and Measurements

United States Office of Education

BALTIMORE

WARWICK & YORK, Inc.

1934

PRINTED IN THE UNITED STATES OF AMERICA

THE MAPLE PRESS COMPANY, YORK, PA.

TABLE OF CONTENTS

LIST OF TABLES

DIFFERENTIAL DIAGNOSIS

CHAPTER I

GENERAL OUTLINE OF PROBLEM

INTRODUCTION

It has been accepted for some time by most educators that an individual may be strong in one ability or skill and weak in another. One person may be particularly good at computation while at the same time very poor in reading. A child in school may be very good in one subject while very poor in another. For instance, many a child in the sixth grade is doing work of an eighth-grade character in arithmetic while doing work of a fifth-grade character in reading. However, any attempt to recognize such differences for use in instruction and guidance has been in the main casual and sporadic. Tests have been given with the view of discovering the strengths and weaknesses of individuals. The results of such testing have been interpreted merely by inspection. To get the full benefit of such measures as we now have, scientific methods of interpretation which have been developed and are available should be used.

This phase of measurement, together with other phases, is becoming more and more important because of changed social conditions and methods of instruction. On account of compulsory school attendance and the advance in the standard of living, children now attend school for longer and longer periods of time. This gives the school the opportunity as well as the duty to educate all children as individuals. Therefore, the discovery of strengths and weaknesses in the individual child becomes one of the principal aims of measurement. It is only through a knowledge of a child's strengths and weaknesses that remedial instruction can be planned and that interests and exceptional abilities can be conserved.

The discovery of strengths and weaknesses in an individual is called the differential diagnosis of the ability of an individual. This exposition will be concerned with the techniques and possibilities for the present determination of these individual strengths and weaknesses. It will not consider some of the more technical problems of *predicting* strengths and weaknesses.[1] It will be recognized that the deter-

[1] For the methods of *predicting* differences see the following: Segel, David. "Differential Prediction of Ability as Repre-

mination of a present condition is often in and of itself a prediction, although no exact statement of the probability of success can always be made without further investigation. This is because with many abilities, results of testing at one time have been found to be substantially correlated with the results of testing the same ability at a later time. It is possible to make a differential diagnosis of a class as well as of an individual but this study will confine itself to the consideration of the individual child.

The general introductory chapter will present first some of the evidence that traits are differentiable and will show some of the relationships between school subjects and skills and the more fundamental traits. It will present next a brief discussion of the need for differential diagnosis. The second chapter will outline the procedures by which the efficiency of tests may be evaluated. These methods may be used in determining the value of tests already constructed and aid in the process of constructing new tests. The third

---

sented by College Subject Groups." Journal of Educational Research. Vol. XXV, Nos. 1 and 2. (January-February, 1932) pp. 14–26, 93–98.

Lee, J. Murray and Segel, David. "The Utilization of Data for Simple or Direct Prediction in the Development of Differential Prediction Regression Equations." Journal of Educational Psychology. Vol. No. XXIV, No. 7 (October, 1933) pp. 550–554.

chapter will give the methods which may be applied to tests in order to make a differential diagnosis.

EVIDENCE OF THE EXISTENCE OF DIFFERENTIABLE TRAITS AND ABILITIES

Although the old theoretical faculty psychology with its emphasis on will power, reasoning, memory, etc., has been discredited, there is a new psychology based on the application of mathematical principles to test results which is interested in the discovery of independent unit traits. Evidence has been found regarding unit traits when the correlation of different available measures have been treated by newly developed statistical methods. Such unit traits are not, however, in general, of the type propounded by the faculty psychologists.

From investigations made, these unit traits seem to be more specific than those of the older faculty psychology although they are at the same time functional in nature. Kelley[2] and Spearman[3] working independently, using methods that were only in part similar, have arrived at conclusions in regard to the existence of unit traits which to a great degree corroborate each other. When investigators are

[2] Kelley, T. L. Crossroads in the Mind of Man. Stanford University; Stanford University Press, 1928.

[3] Spearman, C. The Abilities of Man. New York: The Macmillan Company, 1927.

able to analyze from their data the traits which are identified, as unitary, i.e., independent, then it will be possible to construct tests which will more closely measure these traits than hitherto. When unit traits can be identified accurately our curriculum and our guidance procedures will become much more scientific. Kelley gives evidence of the existence of the following traits: (1) Facility with verbal material; (2) manipulation of spatial relationships; and (3) memory. These traits were found with kindergarten, third-grade, and seventh-grade children. This is evidence that they are very stable and probably arise from birth or at an early age.

Other investigations besides those of Kelley and Spearman have not attempted to get at the fundamental psychological traits but have used the results of tests directly to identify variations in abilities and skills without seeking to identify the underlying ability or group of abilities. For example, reading and grammar may be measured and the results compared without attempting to get at the more fundamental underlying abilities.

This discussion will be confined to the abilities represented by subject achievement in the schools because it is there that the development of testing gives the greatest promise for the first attack with the differential diagnostic technique. It will not be concerned with abilities and traits that are very closely related to school activities but which are not perhaps standardized or described well enough to be used with the techniques given here. In such class, perhaps, may be put those traits measured by tests of personality.[4] However, the techniques to be described may be used in such a field when a difference between any two traits as measured by these tests of personality is thought to be of significance. This field will probably require the use of these techniques very soon.

An ability expressed by a school subject or a large division of a school subject is quite complex and is probably rooted in native tendencies, whereas the detailed items of knowledge or skills within a school subject or large division of a school subject are largely affected by the accidental school environmental variations. There are also no doubt abilities and skills intermediate between these two extremes.

To date there have been few studies made which throw light upon the variation in individual pupil ability. Kelley[5] in establishing a more accurate method of differential diagnosis also described the

[4] See Symonds, P. M. Diagnosing Personality and Conduct. New York: The Century Company, 1931.

Watson, Goodwin. "Tests of Personality and Character." Review of Educational Research. Vol. II, No. 3, (June, 1932).

[5] Kelley, T. L. "A New Method for Determining the Significance of Differences in Intelligence and Achievement Scores." Journal of Educational Psychology, Vol. XIV, No. 6, (September, 1923) pages 321–333.

method for discovering the percentage of differences between abilities as shown on results of tests in excess of chance. This method of judging the effi-

TABLE I.—PERCENTAGE OF DIFFERENCES IN INDIVIDUAL TEST SCORES IN EXCESS OF THE CHANCE PERCENTAGES
(*From Kelley*)

	Computation	Arithmetic reasoning	Arithmetic total	Word meaning	Sentence meaning	Paragraph meaning	Reading total	Language usage	Spelling	Science information
Arithmetic reasoning..	26									
Word meaning........	37	38								
Sentence meaning.....	33	33	..	18						
Paragraph meaning....	28	27	..	22	17					
Reading total.........	..	..	36							
Language usage.......	28	28	29	26	18	14	20			
Spelling..............	27	41	37	44	37	32	44	29		
Science information....	26	18	20	28	22	20	26	24	33	
History and literature information.........	33	31	33	35	24	27	31	27	33	10

ciency of a test for differential diagnosis is described in the second chapter. It may be used with tests already constructed but if possible should be used while the test or tests are being constructed. Kelley, using the Stanford Achievement Test battery (original edition), found the percentage of differences in individual test scores in excess of the chance percentage. The percentages are given in Table I.

The data of this table indicate that 44 per cent of eighth-grade children show measurable differences in relative ability between measures of the meaning of words and spelling while only 10 per cent show differences in excess of chance between history and literature information and science information. The other percentages of differences in excess of chance are between 10 per cent and 44 per cent. These results show that some school subjects are quite a bit alike while others are quite different in respect to measurable qualities. The possibilities for instruction and guidance can be appreciated from an examination of this table.

The Lees,[6] using Kelley's method, calculated the percentage of differences occurring in excess of chance between algebra and geometry. They found a percentage of 41 in the case of one school and a percentage of 34 in another. These percentages show that there are possibilities of differentiation among high-school subjects. Segel,[7] using another

[6] Lee, Dorris M., *and* Lee, J. Murray. "Some relationships between algebra and geometry." Journal of Educational Psychology, Vol. XXII, No. 7, (October, 1931) pp. 551–560.

[7] Segel, David. "Differential prediction of ability as repre-

method, found evidence of differences existing between the following college subject groups: languages, economics, history, biological science, physical science, and English. The differences between some of the subject pairs were very marked while between other subject pairs no difference at all was discernible.

Several correlational studies have been made which give indications of the relationship between different school subjects. Among such studies are those of Meier and Seashore[8] and of Paterson, Elliott, et al.[9] Meier and Seashore found correlations between their art tests and tests of general intelligence of from −.28 to −.14. Paterson, Elliott, et al., found correlations between certain tests of mechanical ability and tests of general intelligence to be −.13. Studies of this sort show great promise for differentiation between various abilities.

---

sented by college subject groups." Journal of Educational Research, January, February, 1932. pp. 14–26, 93–98. The method described in this article, as well as other methods mentioned here, are given in the chapter on "Technical Methods for Use in Determining the Efficiency of Tests to Make a Differential Diagnosis."

[8] Meier, N. C., *and* Seashore, C. E. The Meier-Seashore Art Judgment Test: Examiner's Manual. Iowa City: State University of Iowa, 1930.

[9] Paterson, D. G., Elliott, R. M., et al. Minnesota Tests of Mechanical Ability. Minneapolis: University of Minnesota Press, 1930.

NEED FOR DIFFERENTIAL DIAGNOSIS

Differential diagnosis makes possible a more accurate program of remedial instruction and guidance of pupils into channels in which they will be successful. This is a natural and logical direction for education to take because of the increasing emphasis on individual instruction. The differential emphasis in individual instruction brought about as a consequence of the application of the differential diagnostic technique has in it the righting of a serious defect in so far as education is supposed to be a democratic affair. This defect is that education *en masse* tends to level a child's gifts and deficiencies into a monotonous plane. This defect is pointed out in a research by Kelley[10] who studied the relative influence of nature (heredity) and nurture on individual differences. By reducing measures of reading, spelling, computation, arithmetic reasoning, language usage, etc., to common measures of growth he was able to compare differences between the abilities expressed by these measures for different age groups. He found some tendency for the differences between these abilities to decrease as the pupil progressed through the elementary school. The first measurements were taken at the age of eight and the last one at

[10] Kelley, T. L. The influence of nurture upon native differences. New York: The Macmillan Company, 1926.

fourteen. Suppose a boy of eight who has been in school a comparatively short time shows a considerable difference in computation and arithmetical reasoning, according to Kelley, by the age of eleven, one-third of this difference between the two would have disappeared and by the age of fourteen almost two-thirds of the difference would have disappeared. This result shows the existence of the lock-step in education because the educational process tended to eliminate outstanding abilities which were apparent at the beginning.

What makes this worse is that this leveling influence affects more those abilities between which it would seem from the standpoint of the welfare of the race that it would be very valuable to have differences, while with those abilities in which a common knowledge would seem desirable differences are in some cases increased. The differences between arithmetical reasoning and arithmetical computation for example is decreased very much while differences between history and literature information, language usage and spelling—subjects in which a common knowledge would be of great value—are somewhat increased.

Is it proper to increase a difference over and above that given by nature in a subject field which is concerned with the common understandings of people? The schools have not consciously brought about this result. It is probably brought about by the practice of schools in continually trying to eradicate errors and paying scant attention to subjects in which a student is particularly superior. This constant attention has brought results in such a field as arithmetic but has not brought results in the field of language usage, history and literature information, etc. The tendency for schools to direct teaching towards those abilities in which an individual pupil is poorest is no doubt the same tendency as has been observed in regard to the teaching of a class as a whole. There has been evidence[11] offered for some time to show that children with high IQ's tend to have AQ's (Accomplishment Quotients) below 100 and that pupils with low IQ's tend to have AQ's above 100. Brown and Lind[12] have now given evidence to show that in a class where *all the pupils are considerably below* 100 *IQ* the lower IQ group of pupils develops higher AQ's than the higher IQ group of pupils. This evidence shows that teaching is directed more towards defects than towards strengths. This experimental work of Brown and Lind's corroborates Kelley's conclusion that schools tend to eradicate defects

[11] Pintner, R. *and* Marshall, H. "Results of the combined-mental educational survey test." Journal of Educational Psychology, Vol. XII, No. 2, February, 1921, pp. 82–91.

[12] Brown, Andrew W., and Lind, Christine. "School achievement in relation to mental age—A comparative study." Journal of Educational Psychology, Vol. XXII, No. 8, November, 1931, pp. 561–576.

when they can be recognized as such by the school and thus in some cases tend to decrease differences which may be desirable not to have decreased.

This tendency for schools to eradicate defects in the individual is again shown in the description of the uses of diagnostic tests. It is almost universally the case that the weaknesses or defects in a child's performance and its remedy are emphasized in the description of these tests. This observation is not of course a criticism of the tests themselves but only of their usage.

In practice, the need for discovering the strength and weakness in children, that is, in making differential diagnoses of children's abilities, is two-fold. One need is for immediate instructional purposes, i.e., to know what remedial instruction to apply and in what fields to enrich a child's experience in the particular line in which he is best fitted. The other need of differential diagnosis is not quite so immediate—it is for accurate educational guidance. While pupils are in the elementary school instruction in both the subjects in which a child is weak and those in which he is strong should be prescribed until a certain minimum is reached. This minimum may depend in a small measure upon how long the child will remain in school. In any case, however, the strengths of the child should have attention so that in attempting to bring up a weakness he may have also a chance to grow in the things in which he has special aptitudes. At about the junior high-school period when differentiation of courses begins to be possible more special attention should be given to the strengths of an individual. Special attention should be given to deficiencies only if such deficiencies are in a subject which contains essentials of the common body of knowledge or which is related in such a way as to be a fundamental necessity to the vocation or vocations towards which it seems the child is pointed.

The interest of a child in a subject in which he is strong may be used in motivating the study of other essential subjects. Kelley[13] says: "Beginning with the junior high school and increasingly into higher grades, idiosyncracies should be fixed upon as clues for educational and vocational guidance. Not uncommonly an idiosyncrasy of a backward child, catered to, developed still further, and attended to in the choice of a vocation is the only opportunity of the individual leading a life of average social usefulness and economic return. A child of chronological age 14, and of 12 in general scholastic accomplishment in reading, spelling, language usage, and geography, but of accomplishment of average 13-year olds in arithmetic has had to face certain peculiar trials in his school life. He has been made aware in many ways, when reading,

[13] Kelley, T. L. Interpretation of Educational Measurements. Yonkers: World Book Company, 1927, pp. 99–100.

spelling, reciting history, etc., that he is inferior to classmates of his own age. In arithmetic this is not so true, and it has been so impressed upon him, in fact, he rather likes arithmetic. This is his opportunity. By special effort he can do average or possibly superior work in arithmetic, and he can get the satisfaction that comes from success, which satisfaction every child, no matter how dull, should receive somewhere in his life. Just as soon as this child has a tolerable knowledge of reading, writing, spelling, and history (which will be at about age 13 in the sixth grade), it is well to let these capacities grow as fast as may be possible, but not let them result in general scholastic retardation. In other words, let the child advance as far as possible with as much satisfaction as possible in mathematics that he may direct his steps to as important a vocation involving it as is within his power.

"A child who is generally superior, but vary markedly superior in some one trait, may become one of the great leaders of the race. This will not be accomplished by his being a general all-round good man."

Differential diagnosis may be considered in connection with the larger subdivisions of human knowledge and abilities or in connection with the minor skills or aspects of a single school subject. The statistical techniques for all such differentiations are the same regardless of whether they are in large subject fields or in the more specific fields.

The need for evaluating tests in school subjects for their efficiency in differentiating between abilities, and in urging teachers and research departments to use the technique of differential diagnosis is very great at this time. This is because of the great growth in the construction and use of educational tests which are diagnostic in character. Most test batteries and almost all diagnostic tests have been issued without adequate evidence of their efficiency as instruments of differential diagnosis and without adequate explanation of their use as instruments of differential diagnosis. Because of this increasing interest in diagnosis and remedial instruction the attention of test makers and test users should be directed towards making a diagnosis by an exact rule rather than by inspection.

CHAPTER II

TECHNICAL METHODS FOR USE IN DETERMINING THE EFFICIENCY OF TESTS AND MEASUREMENTS TO MAKE A DIFFERENTIAL DIAGNOSIS

GENERAL DISCUSSION

This chapter is concerned with the description of the techniques which may be used in *determining the efficiency of tests to make a differential diagnosis.* These techniques should not be confused with the differential diagnostic techniques themselves which are described in the next section. The necessity for using some technique in the evaluation of tests already constructed and in the construction of new tests designed to make a differential diagnosis is obvious when the importance of the decisions based on test results is appreciated.

An accurate knowledge of the skills and abilities of a pupil will aid materially in instruction and guidance. For example, the batteries of tests covering the subjects of the elementary school, such as the New Stanford Achievement Tests, the Public School

Achievement Tests, the Metropolitan Achievement Tests, the Modern School Achievement Tests, and the Unit Achievement Scales,[14] and combinations of other single tests, are now used to a great extent in both guidance and instruction in the elementary school and junior high school. A student who is found to be several points higher on a test in one subject than in another often receives different treatment both in instruction and guidance. For instructional purposes a pupil may be placed in different ability sections; receive remedial instruction; be relieved from studying a subject; etc., on the basis of this variation in score. Similarly in guidance, particularly in the junior high school, the future educational and vocational plans of a pupil may be prepared upon the basis of the variations in test score. Therefore it becomes important to know to what extent test batteries or combinations of tests can be relied upon to differentiate between abilities in a pupil.

The methods described in this chapter evaluate the validity of the tests as agents of differentiation. Validity in testing, in general, means the correspondence of a measure with the ability it is supposed to measure. In the case of the use of tests for differential diagnosis we have two elements determining the

[14] The various test batteries are described in more detail in the last chapter.

validity. Assume two actual different skills in reading *A*, and *B*. Suppose tests have been constructed to measure more or less efficiently each of these skills. Call these tests *a*, and *b*, to correspond with the skills *A*, and *B*, to be tested. To make a differential diagnosis one must find the difference between *a* and *b*. The two elements of validity to consider in the situation are: first, the correlation of tests *a* and *b* with the actual skills *A* and *B*; second, the efficiency (validity) with which the difference between tests *a* and *b* actually means a difference between the skills represented in the tests. The first sort of validity is obtained through correlations between various tests, through studies of errors, and various analytical studies. Many such studies and analyses are listed and described by Brueckner and Melby. For instance in the case of reading of the silent type they give an inventory of skills, knowledges, attitudes which begins as follows:[15]

I. In which silent reading predominates:
- *A*. Ability to locate material quickly required:
  1. Knowledge of and an ability to use an index, which suggests:
     - *a*. Learning the alphabet
     - *b*. Finding words in alphabetical arrangement
     - *c*. Arranging words alphabetically

[15] Brueckner, Leo. J., *and* Melby, Ernest O. Diagnostic and Remedial Teaching. Boston: Houghton, Mifflin and Co., 1931, pp. 250–51.

   d. Finding answers to questions by use of an index
   e. Making an index for a book which has none, etc.
2. Ability to use a table of contents, which suggests:
   a. Finding lessons in the table of contents
   b. Finding authors in a table of contents
   c. Finding all the stories written by a single author
   d. Finding all the stories on a certain topic
3. Ability to use the dictionary, which suggests:
   a. Practice in locating words rapidly
   b. Practice in selecting definitions pertinent to the material being read.
4. Ability to use a library file, which suggests:
   a. Responsibility for the care of classroom files
   b. Excursions to a near-by library
5. Ability to use reference material, which suggests:
   a. Practice lessons in using reference sets of informational readers in school
   b. Stimulate interest—locating material at home
6. Ability to use maps, tables, graphs, which suggests:
   a. Practice in interpreting maps, tables, graphs, and diagrams
   b. Practice in making simple graphs to illustrate a topic or point
   c. Answering questions based on maps, graphs, etc.
7. Ability to skim, which suggests:
   a. Skimming to find material on a certain topic
   b. Skimming to find the answer to a given question

c. Skimming to find a sentence which proves or disproves a certain point

d. Skimming to find suitable material for a school program

e. Skimming to find different types of information in the daily paper.

The second aspect of validity which measures the power of differentiation of two tests are those described in this section. The first type of validity is that common to all uses of tests. The second type is concerned with the use of instruments in differential diagnosis.

Test makers in general have neglected to establish the value of their tests in this regard. Diagnostic tests have been investigated even less than test batteries in regard to the differentiation of skills and abilities. Diagnostic tests are used in large numbers for remedial instruction or for a basis for individual instruction. In remedial instruction, or in individual instruction, teachers vary their teaching of single skills or abilities according to test results. This practice has grown up without adequate evaluation of the tests as instruments of diagnosis in practically all cases. Test makers should make it their duty to investigate the value of their tests for diagnostic purposes and make public the results of their investigations so that school systems can choose their tests intelligently.

One of the factors needed in finding the validity of a differential diagnosis using a diagnostic test is the reliability of each part of the test. These reliabilities are not usually available. In many cases it has apparently been assumed that the reliability of each of the different parts is 1.00. It has been found by Brueckner and Elwell[16] that diagnostic tests in arithmetic do not necessarily measure the different skills reliably. They show that pupils may solve one or two problems out of a series of four problems representing a particular skill while making

TABLE II.—VARIATIONS IN NUMBER OF FAULTY SOLUTIONS OF EACH TYPE OF ARITHMETIC PROBLEM (*After Brueckner and Elwell*)

Number of faulty solutions	Type						Total number	Total percentage
	A	*B*	*C*	*D*	*E*	*F*		
One	63	64	88	80	77	71	443	40.2
Two	45	36	53	45	34	48	261	23.7
Three	26	22	26	34	21	30	159	14.4
Four	34	26	57	39	30	52	238	21.6
Total	168	148	224	198	162	201	1101	99.9

[16] Brueckner, Leo J. *and* Elwell, Mary. "Reliability of Diagnosis of Error in Multiplication of Fractions." Journal of Educational Research Vol. XXVI, No. 2, November, 1932. pp. 175–185.

errors in the others. This shows plainly that there is a question regarding the validity of a differential diagnosis when the skills are so inadequately measured. The variations in the solution of six types of problems are given in Table II.

Brueckner and Elwell state in this connection: "One may conclude from the data in Table II that an error of a single example of a given type is not at all a reliable index of what a pupil is likely to do on another example of the four of a given type correctly missed from one to three of the remaining three."

KELLEY'S METHOD FOR EVALUATING THE EFFICIENCY OF TESTS IN MAKING A DIFFERENTIAL DIAGNOSIS

The most important method which gives the efficiency of tests in making a differential diagnosis is the one developed by Kelley in connection with his study of methods for individual diagnosis.[17] The standard deviation of a difference for a given pupil using standard scores is given by the formula

$$\sigma_{d \cdot \alpha\omega} = \sqrt{2 - r_{1\mathrm{I}} - r_{2\mathrm{II}}}$$

where $r_{1\mathrm{I}}$ and $r_{2\mathrm{II}}$ are respectively the reliabilities of the two measures under consideration. The standard deviation of differences in a distribution

[17] Kelley, T. L. "A new method for determining the significance of differences in intelligence and achievement scores." Journal of educational psychology, Vol. XIV, No. 6, September, 1923.

is expressed by the usual formula for the standard deviation of differences

$$\sigma_d = \sqrt{\sigma_1{}^2 + \sigma_1{}^2 - 2r_{12}\sigma_1\sigma_2}$$

which, when standard scores are used, will reduce the standard deviation to 1 and the formula becomes

$$\sigma_d = \sqrt{2 - 2r_{12}}$$

where r_{12} is the correlation between two measures.

The efficiency in differentiating between abilities or traits in an individual will depend upon the relation between these two standard deviations. If the standard deviation of the difference between two measures of an individual is the same as that of the distribution of differences as a whole, the difference in one individual will not express any knowledge of value about the individual above that of the whole distribution. If the standard deviation of the difference between two measures in an individual varies less than the standard deviation of the distribution of differences itself the difference will have meaning, because there will be some differences at the extremes from the mean of the distribution of differences which will not vary enough by chance to bring them past the mean. Some of the measures may be relied upon to mean that the student is better in one ability than another. The ratio between these two standard deviations and the corresponding proportions of the differences which will be in excess of chance are given in Table III.

The proportion of differences in excess of the chance proportion may be worked out for any two tests which have the reliabilities and the intercorrelation calculated for a particular grade or group. Upon the basis of the results it may be decided if the tests are efficient enough for the purposes outlined.

TABLE III.—PROPORTION OF DIFFERENCES IN EXCESS OF THE CHANCE PROPORTION (*From Kelley*)

$\frac{\sigma_{d \cdot \alpha\omega}}{\sigma_d}$	Proportion of differences in excess of the chance proportion	$\frac{\sigma_{d \cdot \alpha\omega}}{\sigma_d}$	Proportion of differences in excess of the chance proportion	$\frac{\sigma_{d \cdot \alpha\omega}}{\sigma_d}$	Proportion of differences in excess of the chance proportion
.02	.950	.35	.467	.70	.171
.05	.888	.40	.415	.75	.138
.10	.798	.45	.367	.80	.108
.15	.719	.50	.323	.85	.078
.20	.647	.55	.281	.90	.051
.25	.582	.60	.242	.95	.025
.30	.522	.65	.205	.99	.005

ILLUSTRATION OF THE USE OF KELLEY'S METHOD OF EVALUATING THE EFFICIENCY OF TESTS IN MAKING A DIFFERENTIAL DIAGNOSIS

The method was illustrated by Kelley with data from the Stanford Achievement Test (original edition).

No other use of this method with test batteries or diagnostic tests has come to the attention of the writer. The necessary data on very few batteries or diagnostic tests are available. Gates[18] has given the necessary data for the analysis of his reading tests. These tests are diagnostic and therefore it is proper that they be used in differential diagnosis if the subtests differentiate the abilities of an individual pupil. Gates' data for his primary tests for a 2*B* class is given in Table IV.

TABLE IV.—CORRELATIONS BETWEEN THE GATES PRIMARY READING TESTS (*Adapted from Gates*)

Grade	Number of pupils	Test I, words with Test II phrases and sentences	Test I, words with Test III directions	Test II, phrases etc., with Test III directions	Self-correlation (reliabilities), tests		
					I	II	III
2*B*	38	.55	.60	.58	.88	.81	.88

There are three possibilities for a differential diagnosis, i.e., as between Test I and II, Test I and III, and Test II and III. Considering the efficiency

[18] Gates, A. I. "The Gates Primary Reading Tests." Teachers College Record, Vol. XXVIII, No. 2, October, 1926, page 177.

———. "Methods of Constructing and Validating the Gates Reading Tests." Teachers College Record, Vol. XXIX, No. 1, November, 1927, pp. 152–53.

of Test I and II as instruments of differential diagnosis
$\sigma_d(\text{Test I and II}) = \sqrt{2 - 2r_{12}} = \sqrt{2 - 2(.55)} = .95$
whereas

$$\sigma_{d\cdot \alpha\omega}\ (\text{Test I and II}) = \sqrt{2 - r_{1\text{I}} - r_{2\text{II}}} = \sqrt{2 - .88 - .81} = .56$$

Then

$$\frac{\sigma_{d\cdot\alpha\omega}}{\sigma_d} = \frac{.56}{.95} = .59$$

A ratio of .59 according to Table III indicates a proportion of differences in excess of chance of approximately .25 (25 percent). The other two possibilities of differentiation have been treated similarly. The results for all three possibilities are as follows:

				Proportion of differences in excess of the chance proportion
Test I and II......	.95	.56	.59	.25
Test I and III.....	.89	.49	.55	.28
Test II and III....	.92	.56	.61	.23

The differentiating power of this diagnostic battery as a whole is therefore about 25 percent since the three differentiations showing its proportions of differences in excess of the chance proportions are .25, .28, and .23. Since these percentages are based on only 38 cases it is not certain that they are stable. It will be assumed for the purposes of this discussion that this sampling is sufficient. In like manner the percentages of differences in individual test scores in excess of the chance percentage for the Gates Silent Reading tests were calculated for 78 pupils in the sixth grade. The results are given in Table V.

TABLE V.—PERCENTAGES OF DIFFERENCES IN INDIVIDUAL TEST SCORES IN EXCESS OF THE CHANCE PERCENTAGE FOR THE GATES SILENT READING TESTS

Test *A* and *B*	.28
Test *A* and *C*	.28
Test *A* and *D*	.34
Test *B* and *C*	.26
Test *B* and *D*	.28
Test *C* and *D*	.23
Average	.28

The differentiating power of the Gates' Silent Reading Battery is slightly higher than that of the Gates' Primary Reading Tests. In like manner other batteries of tests may be investigated and the results on one compared with those on another. Experience with the practical situation will determine in the long run what percentage is necessary for the various types of differential diagnosis for instructional purposes and for educational guidance purposes. One thing that can be done immediately is to compare tests that are being constructed with those already published so that tests may be improved in this regard. The writer ventures to make a tentative judgment that any two tests that have differentiating power of less than 25 percent should not be called diagnostic in relation to each other or, in other words, that tests having a differentiating power of less than 25 percent should not be used in differential diagnosis.

DETAILED SUGGESTIONS FOR USES OF KELLEY'S METHOD IN TEST CONSTRUCTION

The aim in the construction of tests which are to have as one of their main requisites the power to discover idiosyncrasy is to make the ratio

$$\frac{\sigma_{d \cdot \alpha\omega}}{\sigma_d} \text{ or } \frac{\sqrt{2 - r_{1\text{I}} - r_{2\text{II}}}}{\sqrt{2 - 2r_{12}}}$$

as small a fraction as possible since the smaller the ratio the greater will be the proportion of differences above that of the chance distribution of differences. This will be seen by referring to Table III. This ratio can be affected by either the correlation between the tests (r_{12}) or the reliabilities of the tests ($r_{1\text{I}}$, $r_{2\text{II}}$). By decreasing the correlation between the tests (r_{12}) or by increasing either reliability coefficient, ($r_{1\text{I}}$ or $r_{2\text{II}}$) or both, the ratio is decreased.

Means of making tests reliable have been discussed by Ruch[19] and others in connection with the construc-

[19] Ruch, G. M. The Objective or New Type Examination. Chicago: Scott, Foresman and Co., 1929.

tion and use of new-type tests. Only one means of increasing the reliability of a test will be mentioned here. It has been found that the reliability of a test depends upon its length.

The relationship between the length of a test and its reliability is expressed by the Spearman-Brown formula. If, upon investigation, it seems necessary to increase the reliability, this formula may be used to determine either the number of items by which the test needs to be increased in length when a certain reliability is desired or to determine the reliability for any particular length of test. The Spearman-Brown formula is as follows:[20]

$$r_{af \cdot AF} = \frac{ar_{\text{II}}}{1 + (a - 1)r_{\text{II}}}$$

in which r_{II} is the reliability of the test, "a" is the number of similar forms and $r_{af \cdot AF}$ is the resulting reliability when a number of forms of the test is used.

For instance, if the reliability of a test is known to be .81 and the desired reliability is .90 the number of times the test must be lengthened will be determined by solving a Spearman-Brown formula for "a." Substituting the values in the formula the result is

[20] Kelley, T. L. Statistical Method. New York: The Macmillan Company, 1924, page 205, formula 157.

$$90 = \frac{(a)(.81)}{1 + (a - 1)(.81)} \text{ or } a = 2.1$$

That is, by lengthening the test a little over two items, the reliability is increased from .81 to .90. If there were two forms of the test already constructed the use of both the forms would bring the reliability at almost the required figure. If, on the other hand, it is thought to be possible to lengthen a test two and one-half times its original length, what would be the resulting reliability if the original reliability was .75? Substituting this data in the formula the following is obtained:

$$r_{af \cdot AF} = \frac{.75(2\frac{1}{2})}{1 + .75(2\frac{1}{2} - 1)} \text{ or } r_{af \cdot AF} = .88$$

In general, referring again to the formula

$$\frac{\sigma_{d \cdot a\omega}}{\sigma_d} = \frac{\sqrt{2 - r_{1I} - r_{2II}}}{\sqrt{2 - 2r_{12}}}$$

the method for decreasing r_{12} is to restrict the materials for the tests so that they do not overlap in regard to (*a*) content; or in regard to the (*b*) processes used by pupils in learning the content. In the case of large subdivisions of subject matter, such as the school subjects themselves, differences in content are the most important to observe. The word content in this connection includes the principles and even the attitudes accompanying the learning of a subject as well as the base informational material. Within a subject the processes or skills used in learning are probably very important in getting or setting up differentiable material. For instance, Gates[21] investigated the possibilities of differentiating reading into the following components by constructing tests (*a*) reading to get the general significance of a passage, (*b*) reading to get the main idea, (*c*) reading to predict the outcome of given events, (*d*) reading to understand precise directions, (*e*) reading to outline a paragraph, and (*f*) reading to note details. These tests differ mainly in the process they evoke in attacking the problem and not in the difference in information called for, as in any of these tests the pupils are not supposed to have previous information of the subject matter of these tests. Through a comparison of the inter-correlations and the reliabilities of the tests, Gates discards two of the tests and offers the remaining four tests as his battery of silent reading tests. As far as is known he did not calculate the percentage of differences in individual test scores in excess of chance for either the six original tests of the final four tests. In Table V were presented the percentages

[21] Gates, A. L. "Methods of Constructing and Validating the Gates Reading Tests." Teachers College Record, Vol. 29, November, 1927, pp. 148–59.

of differences in individual test scores in excess of chance for the four tests finally included in his battery. The average percentage for this table was 28. The percentages for the two tests he discarded were also calculated. The results are given in Table VI. They are the same sixth-grade children from which the results of Table V were obtained.

TABLE VI.—PERCENTAGE OF DIFFERENCES IN INDIVIDUAL TEST SCORES IN EXCESS OF THE CHANCE PERCENTAGE FOR THE TWO TESTS WHICH WERE DISCARDED BY GATES*

Test *X* and *A*	.10
Test *X* and *B*	.01
Test *X* and *C*	.08
Test *X* and *D*	.08
Test *Y* and *A*	.38
Test *Y* and *B*	.33
Test *Y* and *C*	.13
Test *Y* and *D*	.14
Test *X* and *Y*	.12
Average	.15

* Test *X* is the Gates' "Reading to Get the Main Idea," and Test *Y* is his "Reading to Outline a Paragraph." These are the two tests which were discarded by Gates.

These results show the effect on the percentage of differences in individual test scores in excess of the chance percentage when certain tests have lower reliabilities and higher inter-correlations than certain other tests.

SECOND METHOD FOR EVALUATING THE EFFICIENCY OF TESTS IN MAKING A DIFFERENTIAL DIAGNOSIS

Another technique used in deciding what the possibilities of differentiation are between two tests is to correct the correlation for attenuation. The formula for the correction for attenuation is

$$r_{\alpha\omega} = \frac{r_{12}}{\sqrt{r_{1\mathrm{I}}}\sqrt{r_{2\mathrm{II}}}}$$

where r_{12} is the correlation of any two measures, $r_{1\mathrm{I}}$ is the reliability of the first measure, $r_{2\mathrm{II}}$ is the reliability of the second measure, and $r_{\alpha\omega}$ is the corrected coefficient. $r_{\alpha\omega}$ therefore represents the true correlation of a relationship existing between the two measures when the errors of measurement have been allowed for. This method was used by Segel[22] in establishing possibilities of differentiation in a study of differential prediction. He found the correlations between marks in different subjects corrected for attenuation as given in Table VII for a group of college students.

[22] Segel, David. "Differential Prediction of Ability as Represented by College Subject Groups." Journal of Educational Research, Vol. XXV, Nos. 1 and 2, January and February, 1932, pp. 14–26, 93–98.

TABLE VII.—INTERCORRELATION COEFFICIENTS FOR MARKS IN CERTAIN COLLEGE SUBJECTS CORRECTED FOR ATTENUATION

	Economics	History	Biol. science	Physical science	English
Languages	.352	1.054	.618	.475	.996
Economics	....	.643	.741	1.004	.388
History	....	.....	.734	.377	.845
Biological science	....	.....	....	.741	.395
Physical science	....	.....	....	.....	.697

As it will be noted, some of these correlations are approximately unity (1.00), such as languages—history, economics—physical science, and languages—English; whereas some correlations are low, such as languages—economics, history—physical science, and biological science—English. The variations in this group of subjects is very great, no doubt in part because of various sources of error. The method used gave very valuable clues, however, as to what pairs or groups of subjects would give the best chance for differentiation. The more accurate method using the ratio $\frac{\sigma_{d \cdot \alpha\omega}}{\sigma_d}$ is to be preferred since it can be easily interpreted on the basis of chances of differences occurring. The correlations corrected for attenuation should not be used in place of the original correlations in calculating $\sqrt{2 - 2r_{12}}$ of the ratio

$$\frac{\sqrt{2 - r_{1I} - r_{2II}}}{\sqrt{2 - 2r_{12}}} \text{ or } \frac{\sigma_{d \cdot \alpha\omega}}{\sigma_d}.$$

In this section the methods for determining the efficiency of tests for marking a differential diagnosis have been given. It is necessary to subject batteries of tests and "diagnostic" tests to the technique described in order to be certain that the tests actually do differentiate between the abilities and skills for which scores are given. The direction of instruction and guidance based upon tests, the differential diagnostic value of which are unknown, may be in many cases wrong. It cannot be emphasized too strongly that tests should be evaluated before being chosen for use as instruments of differential diagnosis.

CHAPTER III

METHODS WHICH MAY BE APPLIED TO TESTS AND MEASUREMENTS TO MAKE A DIFFERENTIAL DIAGNOSIS

GENERAL DISCUSSION

This section will give two methods for treating test results to make a differential diagnosis. These methods are usable no matter what kind of abilities or skills are in question. Both methods involve the calculation of differences between scores on tests and the application of a probable error technique to estimate the significance of the differences. So far as the differences are concerned, either method may be considered sound, but so far as the probable errors of the differences are concerned, the second is the more reliable because it allows for differences in variables of a group in the two or more measures involved by assuming standard deviations equivalent, whereas the first method does not do this. The probable errors of the differences here given are only approximate which ever method is used, but they are probably rather close approximations.

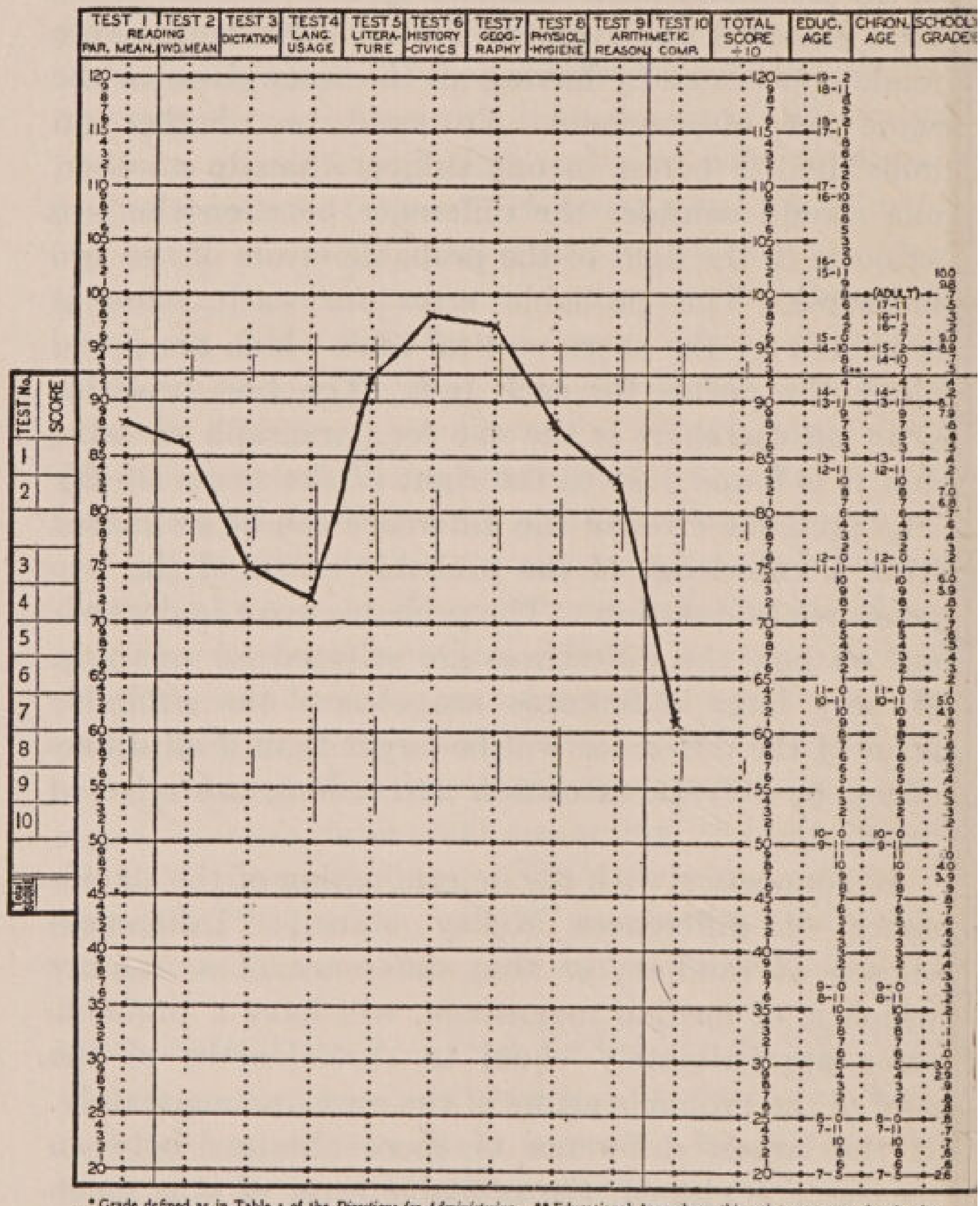
By the use of these methods, teachers and research departments can accurately determine for individual children the significance of differences between scores. The importance of knowing the meaning of differences between results in different tests has been discussed in the previous sections. If tests, which have been shown by the techniques described in the previous section to be valueless for the purposes of differential diagnosis, are nevertheless used, the application of the techniques given in this section will be a waste of time. This is true because the techniques here given cannot establish a difference between test results on an individual child if tests have been found by the analysis of results on large numbers of children to be incapable of showing such differences. Also, if it has been shown that a fairly large proportion of differences between two tests are in excess of the proportion of chance differences, it is advisable for most purposes to establish the meaning of the differences in each individual according to one of the methods presented here. The accurate planning of individual instruction and of guidance requires constant attention to differences within an individual.

FIRST METHOD OF MAKING A DIFFERENTIAL DIAGNOSIS

This method has been used in a few instances when test batteries have been given. One illustration of it will be given in the elementary field and one in the secondary field. Chart I gives the norms in the New Stanford Achievement Test. These norms were made equivalent by having all the tests given to the same set of students. To estimate whether an individual is better in one subject than in another, one should consider the difference between the two subjects in the light of the probable errors of the two measures. The probable error for each score is indicated by the short vertical lines which are found along the norms for each test. The first probable error on the chart is the one for paragraph meaning and it is found just to the right of the scores 90–95. The probable error of the difference can be estimated from a knowledge of the probable errors of the two measures in question. The probable error is dependent on how the differences are selected for analysis. If only large differences are chosen the probable error of the difference will be larger than if all differences, or differences chosen at random, are selected for study.

In connection with the determination of the significance of differences Kelley states: "Differences chosen at random, or that difference of a number which is of median magnitude, will have a probable error approximately equal to three-fourths of the sum of the probable errors of the measures separately. If the largest difference of those obtained between measures is chosen, the probable error of it is much

EDUCATIONAL PROFILE CHART: NEW STANFORD ACHIEVEMENT TEST, ADVANCED EXAMINATION*



* Grade defined as in Table 1 of the *Directions for Administering*. ** Educational Ages above this point are extrapolated values. See *Guide for Interpreting* for explanation of vertical lines.

☞ *This Profile Chart is the table of norms for the Advanced Examination.*

Chart I.

larger than this. The following table should give an approximate idea of the probable error of this largest difference."[23]

TABLE VIII.—PROBABLE ERROR VALUES FOR VARYING CONDITIONS OF SELECTION OF DIFFERENCES (*After Kelley*)

Number of measures	Number of differences	Approximate value of the PE of the median difference	Approximate value of the PE of the largest difference	
2	1	.75 (PE_1–PE_2)	Preceding value times	1.00
3	3	do	do	1.4
4	6	do	do	1.7
5	10	do	do	1.9
6	15	do	do	2.1
7	21	do	do	2.3
8	28	do	do	2.4
9	36	do	do	2.5
10	45	do	do	2.5
15	105	do	do	2.8
20	190	do	do	3.0

The significance of various sized differences measured in terms of probable errors is given in Table IX.

The chance that the real difference is of opposite sign to that found is represented by the area in one

[23] The statement here given and the table following were given by Dr. T. L. Kelley to the writer in correspondence regarding this particular part of the study of differential diagnosis.

tail of the distribution and not the sum of the areas in two tails. For instance, taking the first few

Table IX.—Significance of Differences[24]

1.0 × PE	3 to 1 chance that difference is significant
1.5 × PE	5 to 1 chance that difference is significant
2.0 × PE	10 to 1 chance that difference is significant
2.5 × PE	19 to 1 chance that difference is significant
3.0 × PE	48 to 1 chance that difference is significant
3.5 × PE	100 to 1 chance that difference is significant
4.0 × PE	284 to 1 chance that difference is significant

differences beginning on the left of Chart I, the following differences are found:

Paragraph meaning (88)—word meaning (86)............. 2
Paragraph meaning (88)—dictation (spelling) (75)........ 13
Paragraph meaning (88)—language usage (72)............ 16

The probable error of the paragraph meaning score at this level is about 5.4 and the probable error of the word meaning score is about 2.3. To get the probable error of the difference between paragraph meaning and word meaning, the probable errors of the two socres, i.e., 5.4 and 2.3, are added and multiplied by .75 since this is the best approximation when the difference is a random sampling. Therefore, the probable error of this difference is equal to 5.8. Since the actual difference 2 is less than its

[24] Adapted from Kelley, T. L. Statistical Method. The Macmillan Company, New York, 1923. Table XXVI, page 103.

probable error the significance of the difference may be considered nil. Similarly the probable errors of the differences between paragraph meaning and spelling and paragraph meaning and language usage would be $.75(5.4 + 2.0)$ or 5.5 and $.75(5.4 + 8.5)$ or 10.4 respectively. The differences for these subjects as given above, 13 and 16 respectively, divided by these probable errors, give ratios of 2.4 and 1.5. What is the significance of these ratios? Referring to Table IX and interpolating it can be seen that the chance of the difference between paragraph meaning and spelling (2.4 times its probable error) being significant would be about 17 to 1. Similarly it can be seen that the chance of the difference between paragraph meaning and language usage (1.5 times its probable error) being significant would be about 5 to 1.

If instead of taking random differences the largest difference in the chart is taken the probable error of the difference would not be $.75\ (PE_1 + PE_2)$ but $2.5 \times .75\ (PE_1 + PE_2)$ (See Table VIII). Since there are 10 measures in this chart, the largest difference is that between history-civics and arithmetic computation. The difference is $98 - 61$ or 37, and the probable error of the difference is $2.5 \times .75\ (4.8 + 3.0)$ or 14.6. The ratio of the difference to this probable error is 2.5. The chance of its being significant is therefore approximately 19 to 1.

Sometimes several differences may be considered together. In the case of the pupil whose abilities are pictured in the chart it can be seen that he is definitely superior in subjects requiring wide reading and extensive interests, i.e., in literature, history-civics, and geography, as opposed to those subjects requiring drill and attention to detail, i.e., in dictation (spelling), language usage, and arithmetic computation. In coming to any conclusions regarding differences between two groups of subjects there should be, of course, some clearly discernible lines of relationship within each of the groups.

As a further illustration of this method of differential diagnosis Chart II[25] is presented. This chart was developed for use in placement and guidance through comparing scores on the various tests of the Iowa High School Content Examination,[26] the total score on that test, the score on the American Council of Education Psychological Examination,[27] and other measures.

This chart was obtained by getting equivalent scores on the different tests which were all given

[25] Devised by Shirley L. Brintle and the writer for use in the Long Beach Junior College.

[26] Published by the Bureau of Educational Research and Service, University of Iowa, Iowa City.

[27] Published by the American Council on Education, 744 Jackson Place, Washington, D. C.

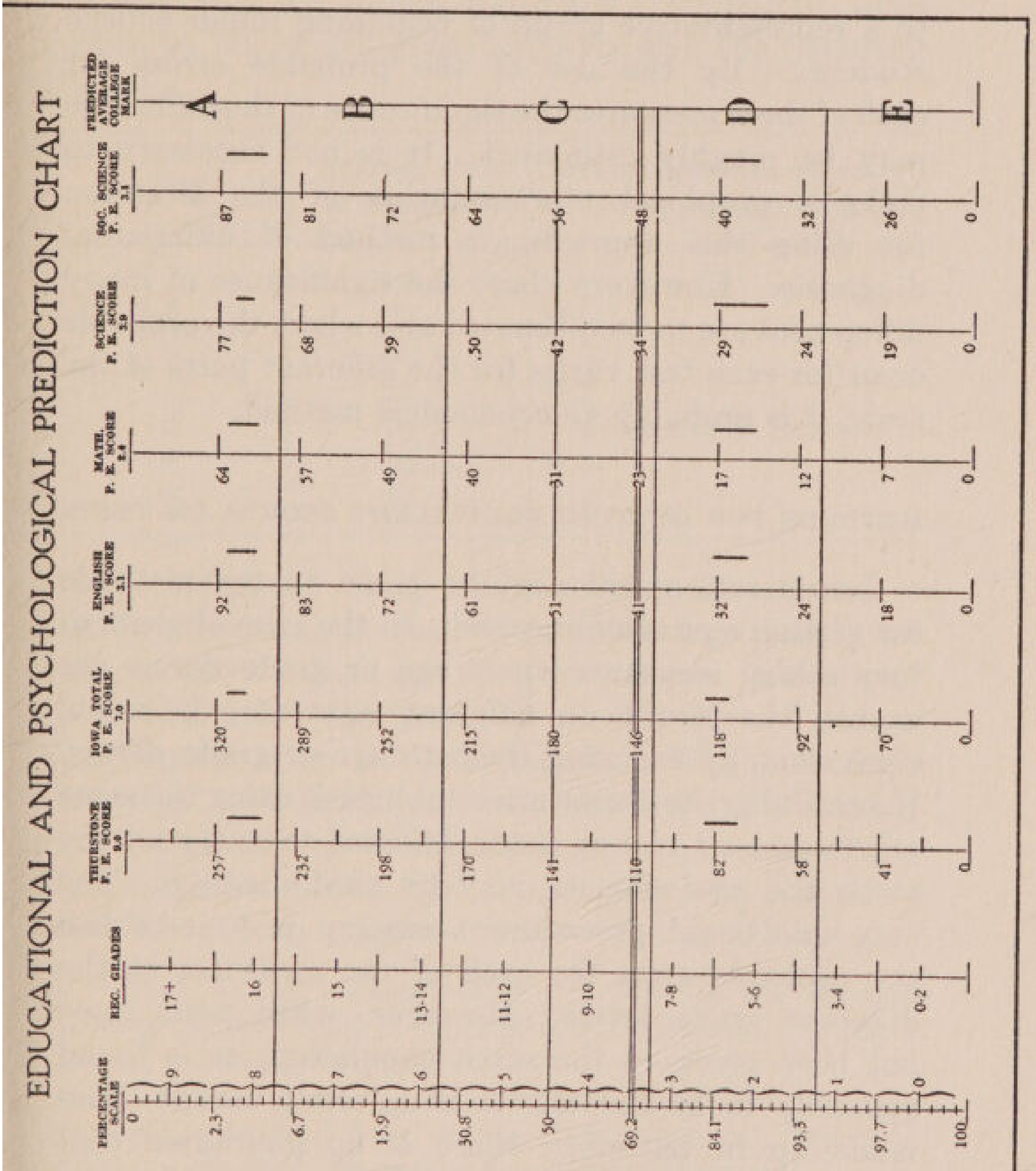

EDUCATIONAL AND PSYCHOLOGICAL PREDICTION CHART
PERCENTAGE SCALE
REC. GRADES
THURSTONE P. E. SCORE
IOWA TOTAL P. E. SCORE
ENGLISH P. E. SCORE
MATH. P. E. SCORE
SCIENCE P. E. SCORE
SOC. SCIENCE P. E. SCORE
PREDICTED AVERAGE COLLEGE MARK
A
B
C
D
E

CHART II.

to a representative group of beginning junior college students. By the use of the probable errors for each of these measures the significance of the difference may be roughly estimated. It is not necessary to make a graphical representation of the situation for using this approximate method of differential diagnosis. However, where the significance of many differences are to be estimated and where the probable error for each test varies for the different parts of the scale, it is probably an economical method.

METHODS FOR GETTING EQUIVALENT SCORES ON TESTS

Consideration will next be given to the methods for getting equivalent scores. In the case of elementary school measures where age or grade norms are usable, the results on different tests may be made equivalent by reducing them to age or grade norms. If age and grade norms are established, using the same population on two or more different tests for all the tests, the equivalences may be used directly. The only additional procedure necessary is to establish the probable error for each of the measures at the different grade levels. However, when tests have not been given to the same population, as is found generally to be the case with all tests except those made up in batteries, there is no guarantee that the norms are comparable. This has been shown by

Ruch[28] who had seven well-known arithmetic tests given to 152 eighth-grade children with the following result:

Test	Average score of 152 pupils	Average score in terms of grade equivalents (norms)
A	224.3	8.5
B	70.8	H 6
C	32.0	H 7
D	26.8	H 5
E	30.1	9 (estimated; above H 8)
F	38.0	H 6
G	56.1	11

If arithmetic tests vary among themselves in grade placement it would seem that no reliance can be placed on either age or grade norms. Kefauver[29] has shown that age norms on certain tests of general intelligence were not comparable. What has been stated of tests in subjects can be said equally forcibly as to subtests or so-called diagnostic tests. If the subtests of a subject test or the parts of a diagnostic test have been standardized on the same group of

[28] Ruch, G. M. Twenty-ninth Yearbook of the National Society for the Study of Education, Bloomington, Ill., Public School Publishing Co., p. 699.

[29] Kefauver, Grayson N. Need of Equating Intelligence Quotients obtained from Group Tests. Vol. 19, February, 1929, pp. 92–101. Journal of Educational Research.

children, the age or grade norms resulting may be considered sufficient for equating purposes. There is a caution to be added here, namely, that even if the tests have been given to the same population

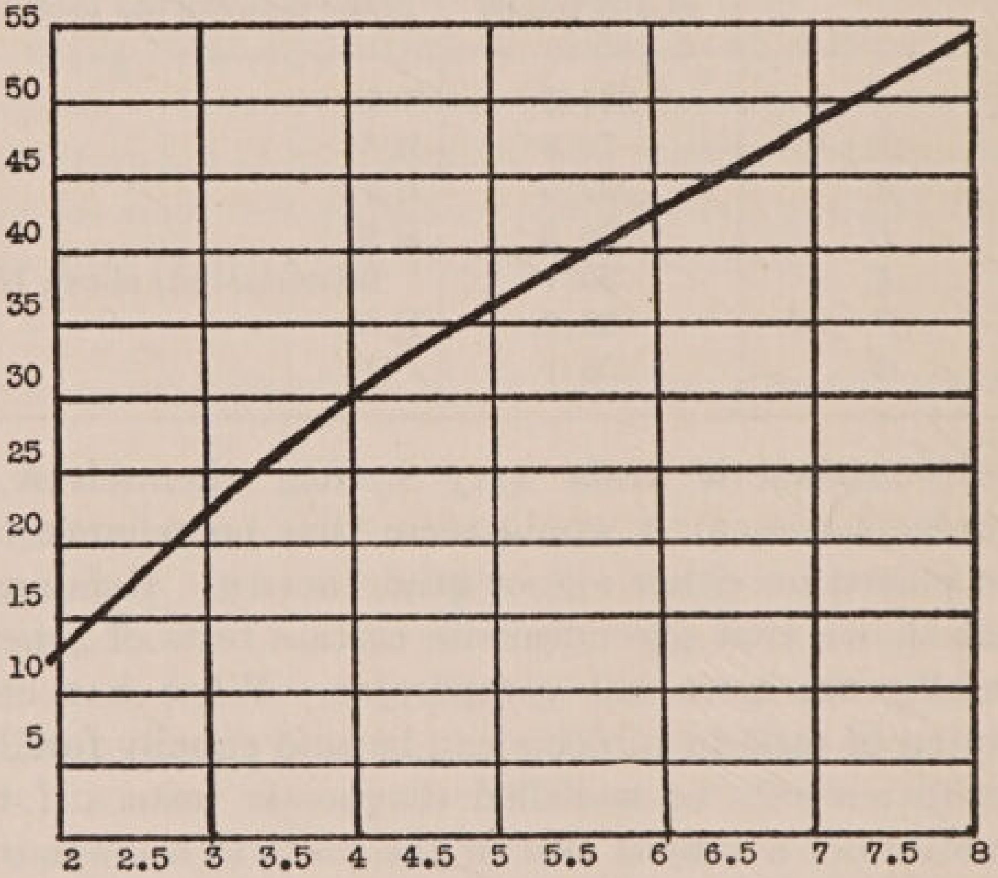


Fig. 1.

the relationship is not necessarily exactly established for groups of children in different environments.

The more representative the population is upon which this relationship is established the more stable it is. It is thought that this relationship of scores is not likely to vary on account of different school systems and differing environments. No further reference to the unreliability due to this particular sampling error will be made.

The use of the following methods for obtaining equivalent scores is recommended. With elementary school subjects, such as arithmetic, reading, etc., or important subdivisions of such subjects, where there is presumably a growth through the grades, equivalence may be obtained by establishing age or grade norms. This may be done by averaging scores for different ages of grades, drawing a curve and interpolating for values between the original points. For example, a reading test had the following grade averages at the very beginning of the year:

Grade	Score
3.0	22
4.0	30
5.0	36
6.0	42
7.0	48
8.0	54

The graph on p. 48 is obtained by plotting these points and drawing a smooth curve to fit.

The results obtained from reading this graph are given in Table X.

When other tests which have been given to the same group of pupils are reduced to grade norms, equivalence of the different tests can be established. This equivalence may be shown graphically or by tables. For convenience all scores may be translated into a common scale of grade placements.

In the case of tests in the high school field, equivalence must be obtained without resorting to age or grade norms because acquisition of material above

TABLE X.—GRADE EQUIVALENTS ON A READING TEST

Score	Grade	Score	Grade	Score	Grade	Score	Grade
13	2.0	24	3.2	35	4.8	46	6.7
14	2.1	28	3.4	36	5.0	47	6.8
15	2.2	26	3.5	37	5.2	48	7.0
16	2.3	27	3.6	38	5.4	49	7.2
17	2.5	28	3.8	39	5.5	50	7.4
18	2.6	29	3.9	40	5.7	51	7.5
19	2.7	30	4.0	41	5.8	52	7.7
20	2.8	31	4.2	42	6.0	53	7.8
21	2.9	32	4.4	43	6.2	54	8.0
22	3.0	33	4.5	44	6.4		
23	3.1	34	4.7	45	6.5		

the elementary grades is dependent on the particular subjects that have been studied. It is found often that the average ability of students enrolled in one subject will differ from the average ability[30] of students

[30] By ability in this connection is meant general scholastic ability such as would be ascertained from the results of general intelligence tests.

enrolled in another subject. Such differences in average ability between students in different subjects have not been found to be the same for different schools. Therefore, with any method of obtaining equivalences on the secondary school level certain general cautions may be stated. Equivalences should be calculated by using only scores for individuals who have taken both tests. Also equivalences established for one school should not be adopted bodily by another school. Equivalences established from the records in many typical schools may be used in any school, providing the school has been found from past experience to be like the typical schools in regard to test norms in different subjects.

The method most commonly used in equating scores on tests on the secondary school level is by applying the formula

$$X_1 = M_1 + \frac{\sigma_1}{\sigma_2}(X_2 - M_2)$$

which is a variation of the statement of equivalences between standard scores, i.e.,

$$\frac{X_1 - M_1}{\sigma_1} = \frac{X_2 - M_2}{\sigma_2}$$

The application of this formula brings about results which are more accurate from some standpoints than the results obtained through equating by getting grade placements. This method is quite accurate when the distributions involved are fairly normal and when the reliabilities of the two measures are not very different.

This formula

$$X_1 = M_1 + \frac{\sigma_1}{\sigma_2}(X_2 - M_2)$$

must not be confused with the regression equation which is equal to

$$X_1 = M_1 + r_{12}\frac{\sigma_1}{\sigma_2}(X_2 - M_2)$$

Test scores have been in some cases erroneously equated through the regression equation. The regression equation regresses the estimated scores toward the mean. This procedure does not produce equivalence of scores, since the range of distribution of such regressed scores on a test is less than that of the range of distribution of unregressed scores on the same test.

There are also other methods for getting approximate equivalent scores. One of the most commonly used methods is that called the equal percentile method. With large numbers this method is accurate enough for most purposes. Assume two distributions of test scores. The scores in each distribution are related to their proper percentile rank. Then the scores on the two distributions having the same percentile rank are considered to be equivalent.[32]

In these methods for obtaining equivalents it is necessary to transmute the scores on the different measures concerned to a common scale. For example, in the New Stanford Achievement Test, as shown in Chart I, all the measures have been reduced to a common 20–120 point measuring rod. The paragraph reading score of 89 was obtained from 49 correct answers. The dictation (spelling) score of 75 was obtained from 54 correct spellings. The other scores were obtained similarly. In each case the raw score has been changed into an equivalent score. The probable error for each of the scores should be calculated from the standard deviation expressed in the common measure. The technique described is a good method for obtaining equivalent scores. There is very little use being made of this method. Only a few tests constructed have made an attempt at arriving at scores that can be compared with scores on other tests. There is need for coöperative effort on the part of research departments of public schools and test constructors in this regard.

[32] For a detailed description of translating scores into percentiles, see Buros, Francis C. and Oscar K. Expressing Educational Measures as Percentile Rank. Test Method Help No. 3. World Book Company, Yonkers, New York.

SECOND METHOD OF MAKING A DIFFERENTIAL DIAGNOSIS

This method of discovering idiosyncrasies was developed by Kelley.[33] This method of making a differential diagnosis allows for the differences in variables of a group in the two or more measures involved by assuming standard deviations equivalent, whereas the first method does not do this. It involves the use of standard scores. Take the difference between any two test scores in different subjects, such as $X_1 - X_2$. These differences cannot be interpreted directly. They must first be expressed as deviations from the grade means divided by the standard deviation. This makes standard scores expressed as follow: $Z_1 = \frac{X_1 - M_1}{\sigma_1}$ $Z_2 = \frac{X_2 - M_2}{\sigma_2}$. Their standard deviations are then equal to 1. Therefore, in standard scores the difference between X_1 and X_2 will be $\frac{X_1 - M_1}{\sigma_1}$ minus $\frac{X_2 - M_2}{\sigma_2}$* or $Z_1 - Z_2$.

[33] Kelley, T. L. A new method for determining the significance of differences in intelligence and achievement scores. Journal of Educational Psychology, Vol. XIV, No. 6, September, 1923, pp. 321–33.

* It should be noted that the calculation of differences between standard scores, i.e. such as $\frac{X_1 - M_1}{\sigma_1} - \frac{X_2 - M_2}{\sigma_2}$, is a different process than finding equivalent scores from the equation $X_1 = M_1 + \frac{\sigma_1}{\sigma_2}(X_2 - M_2)$ even though they are both

The result of calculating this difference and comparing the result with its probable error will reveal the significance of it. The probable error of any such difference is equal to: $.6745\sqrt{2 - r_{11} - r_{2II}}$ where r_{1I} is the reliability of the X_1 scores and r_{2II} is the reliability of the X_2 scores. It can be seen that for this method the reliability, standard deviation and the grade mean for each of the measures for the grade in which the child under consideration is located is necessary. With this method, the probable errors in the case of random sampling is obtained without reference to the data given in Table VIII. However, where there is a selection of differences by the choice

---

evolved from the relationship between two standard scores. There has been some confusion regarding this point. To illustrate the calculation in each case the following data are given.

$$X_1 = 28 \qquad M_1 = 22 \qquad \sigma_1 = 5$$
$$X_2 = 33 \qquad M_2 = 43 \qquad \sigma_2 = 10$$

In this case the difference between two standard scores would be

$$\frac{28 - 22}{5} - \frac{33 - 43}{10} = 1.2 - (-1.0) = 2.2$$

whereas an equivalent score on one test, knowing the score on the other, would be solved as follows

$$X_1 = 22 + 5/10\ (33 - 43) = 27$$

The calculation of equivalent scores does not affect the standard deviations of the distribution of the scores as is the case in the calculation of standard scores themselves.

of the largest of a number of differences within an individual the values given in the last column of Table VIII, headed "Approximate value of the PE of the largest difference," applies to this method in the same manner as it was applied in the other method.

This method of differential diagnosis will be illustrated by using the same scores and differences

TABLE XI.—TABLE SHOWING THE CALCULATION OF STANDARD SCORES FOR A PUPIL

	Score X	Mean (M)	Sigma (σ)	$\frac{X - M}{\sigma}$ or Z
Test I, paragraph meaning....	88	91.3	12.5	$\frac{88 - 91.3}{12.5}$ or $-$ .26
Test II, word meaning........	86	88.9	13.5	$\frac{86 - 88.9}{13.5}$ or $-$ .21
Test III, dictation (spelling)	75	85.0	10.2	$\frac{75 - 85}{10.2}$ or $-$ .98
Test IV, language usage......	72	84.7	18.5	$\frac{72 - 84.7}{18.5}$ or $-$ .69
Test VI, history civics........	98	84.7	14.3	$\frac{98 - 84.7}{14.3}$ or $+$ .93
Text X, arithmetic computation	61	90.4	15.2	$\frac{61 - 90.4}{15.2}$ or -1.9

as were used in the first method. The means, standard deviations, and reliabilities[34] are for the seventh grade since the pupil whose record is being used was in the seventh grade. The standard scores

[34] Obtained from the Guide for Interpreting for the New Stanford Achievement Test. World Book Company, New York City, 1929.

for the five scores between which differences are to be calculated are given in Table XI.

The reliabilities for the five tests at the seventh-grade level are .85, .79, .83, .69, .71, and .73 respectively. The probable error of the difference between Test I (paragraph meaning) and Test II (word meaning) would therefore be $.6745\sqrt{2 - .85 - .79}$ or .40. Similarly the probable errors for the differences between Test I and Test III and Test I and IV would be .38 and .46 respectively. The calculation of the differences and the resulting ratios when the difference is divided by the probable errors are given in Table XII.

TABLE XII.—TABLE SHOWING THE CALCULATION OF D/PE_d RATIOS FROM STANDARD SCORES

Z scores	Difference	PE_d	D/PE_d
Test I minus Test II, (−.26) − (−.21)	−.05	.40	− .12
Test I minus Test III, (−.26) − (−.98)	+.72	.38	+1.90
Test I minus Test IV, (−.26) − (−.69)	+.43	.46	+ .93

These ratios may be referred to Table IX for interpretation of the significance of the differences.

The difference between history-civics and arithmetic computation was selected for study because it was the largest of the differences between test results on this particular pupil. The calculation of the difference and its probable error is as follows. The difference is $(+.93) - (-1.90)$ or 2.83. The probable error is 2.50 times $.6745\sqrt{2 - r_{1I} - r_{2II}}$ or 2.50 times $.6745\sqrt{2 - .71 - .73}$ or 1.26. The ratio of the probable error (1.28) to the difference 2.83 is 2.83/1.26 or 2.2. The chance that this difference is significant is about 14 to 1. It will be seen that the probable errors using this method are somewhat different from the probable errors obtained through the first method. Also the probable errors obtained by this method are more reliable than those obtained by the first method.

In the secondary field the calculation of differences between sigma scores will be illustrated by using scores on the Powers General Chemistry Test and the Ruch-Cossman Biology Test[35] and the data which are presented in the Manuals of Directions for these tests.[36] In using the data from the Manuals of Directions and in giving an illustration from these tests a certain assumption is made. This is that the

[35] These tests are published by the World Book Company.

[36] The medians given in the Manuals of Directions were used as means. The standard deviations were not given in Manuals of Directions. These were obtained by solving the equation $PE_{Score} = .6745\sigma_1\sqrt{1 - r_{1I}}$ for using the PE_{Score} and r_{1I} as given in the Manuals of Directions.

distribution of general ability in the samples used for obtaining the norms on these tests is the same. There is also an assumption that the student concerned has studied both chemistry and biology during the same year.

Suppose a student obtained a score of 26 on the Powers Test and a score of 47 on the Ruch-Cossman Test. The standard scores for these two test scores would therefore be:

$$\text{Powers} \quad Z_1 = \frac{X_1 - M_1}{\sigma_1} = \frac{26 - 36}{10} = -1.0$$

$$\text{Ruch-Cossman } Z_1 = \frac{X_1 - M_1}{\sigma_1} = \frac{47 - 37}{14.5} = +.7$$

The difference between the two scores will therefore be $-1.0 - (+.7) = -1.7$. The minus sign of this difference indicates that the difference is in favor of the second score over the first. The probable error of this difference, i.e.

$$PE_d = .6745\sqrt{2 - r_{1I} - r_{2II}} = .6745\sqrt{2 - .80 - .93} = .35$$

Since the difference -1.7 is almost five times its probable error (.35) it can be considered as significant.

Methods for determining significant differences between test scores on different subjects in an individual have been described. These methods may be used in both the elementary school and in the high school.

Where it has been shown by previous research by using such techniques as were described in the second chapter that differences between any two tests or parts of a test are very likely to be significant, the application of the methods described in this section might be to some extent superfluous. However, to date no two tests or parts of tests used in school have been shown to have a high efficiency in differentiation. Therefore, the methods described here should be used whenever the differences in results on two tests are to be used in altering materially the instruction or guidance of pupils.

CHAPTER IV

AVAILABLE TESTS AND MEASUREMENTS WHICH MAY HAVE VALUE IN MAKING A DIFFERENTIAL DIAGNOSIS

GENERAL DISCUSSION

All the means of measurement which may be investigated for possibilities of use in differential diagnosis cannot be mentioned here. This is because any two tests which are not measuring exactly the same function may have possibilities for differential diagnosis.

For the differential diagnosis of the larger groups of school abilities as represented by reading, arithmetic, English, history, art, music, industrial arts, etc., there are large numbers of possible tests to be used.[37] The batteries of tests covering several of these may be of special use in this regard because they have already been given to common populations

[37] For other educational and mental tests than those mentioned in this chapter see Hildreth, Gertrude, A Bibliography of Mental Tests and Rating Scales. The Psychological Corporation, New York, N. Y. 1933.

so that the underlying data necessary for both methods of differential diagnosis may be at hand. The manuals of directions for some batteries are much better in regard to the giving of necessary basic data than others. But since such basic data, if not in the manual of instructions, may in some cases be obtained from the authors, we shall not distinguish between batteries of tests in this regard. Where facilities are at hand to give tests and calculate standard deviations and reliabilities there is no reason why the results on any subject matter test cannot be compared with the results on another. This is not true only for standardized tests that are published and sold commercially, but with any test constructed for which it is possible to obtain the adequate data necessary to make a differential diagnosis. That is, objective tests constructed by the teacher or uniform tests constructed for a school or school system may be used in differential diagnosis. In all cases it must be remembered always that the value of the tests for this purpose depends upon their differentiating power.

If two tests distinguish between school subjects in only a very small proportion of cases when the differential diagnostic procedure is applied, this does not *necessarily* mean that the two abilities which the tests are presumed to measure are too dependent to allow a difference to be shown, but it may mean that the particular tests used were inadequate for the job.

In the high school, when using test batteries, it should be remembered that the most accurate differential diagnosis can be made as between subjects only if the student has studied the subjects. It is true in many cases that the reason the student has not studied a subject is because he attempted it and failed, or because he was not interested in the subject, or still again because he did not believe he would be able to pass in the subject. In so far as these reasons are the true ones, the differential diagnostic technique will indicate intrinsic differences in part. Most of the high school batteries which are published are usually divided into such broad subject divisions as to allow comparisons between such subject divisions when the tests are taken at the end of the high school course.

The tests mentioned in this section are given because they represent some of the possibilities put forth in this field at present. Each must be judged in the light of the techniques given in Chapters II and III. It is to be emphasized that in mentioning the names of batteries of tests we are not recommending them or implying that other subject matter tests are not as good or better for the purpose. The tests named are given because there seems to be more of a possibility to get the necessary data for them than for others.

DIFFERENTIAL DIAGNOSIS BETWEEN SUBJECTS WITH AVAILABLE TEST BATTERIES

A list of test batteries, together with the subjects found in each, between which a differential diagnosis may be possible, is given herewith.

Every Pupil Primary Achievement Test. Published by the Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kansas.
For grades 1–3.
Two forms.
Subjects tested: (*a*) arithmetic computation; (*b*) reading arithmetic; (*c*) sentence spelling; (*d*) word knowledge; (*e*) sentence understanding; and (*f*) paragraph meaning.

Illinois Examination. Published by the Public School Publishing Company, Bloomington, Illinois.
No. 1 for grades 3–5.
No. 2 for grades 6–8.
Two forms each.
Subjects tested: (*a*) arithmetic; and (*b*) silent reading. (Also general intelligence.)

Indiana Composite Achievement Test. Published by the Bureau of Coöperative Research, University of Indiana, Bloomington, Indiana.
For grades 7 and 8.
Subjects tested: (*a*) arithmetic; (*b*) American history; (*c*) Indiana history; (*d*) civics; (*e*) geography; (*f*) language; (*g*) reading; (*h*) physiology; and (*i*) spelling.

Iowa High School Content Examination. Published by the Bureau of Educational Research and Service, State

University of Iowa, Iowa City, Iowa.
For high school and college.
Two forms.
Subjects tested: (*a*) English; (*b*) social studies; (*c*) science; and (*d*) mathematics.

Iowa Silent Reading Test. Published by the World Book Company, Yonkers, New York.
For high school and first year of college.
Two forms.
Subjects tested: (*a*) Tests on paragraph meaning in (1) science and (2) literature; (*b*) word meaning tests in (1) social sciences, (2) science, (3) mathematics, and (4) English; (*c*) paragraph organization tests in (1) selection of the central idea and (2) outlining; (*d*) sentence meaning; (*e*) use of the index; (*f*) selection of key words; and (*g*) rate of reading.

Metropolitan Achievement Tests. Two editions. National and New York City. Published by the World Book Company, Yonkers, New York.
Primary I. For grade 1.
Three forms.
Subjects tested: (*a*) word picture; (*b*) word recognition; (*c*) word meaning; and (*d*) numbers.
Primary II. For grades 2 and 3.
Three forms.
Subjects tested: (*a*) reading completion; (*b*) paragraph meaning; (*c*) vocabulary; (*d*) arithmetic fundamentals; (*e*) arithmetic problems; (*f*) language; and (*g*) spelling.
Intermediate. For grades 4–6.
Five forms.
Subjects tested: (*a*) spelling; (*b*) reading; (*c*) vocabulary; (*d*) arithmetic problems; (*e*) arithmetic fundamentals; (*f*) language usage; (*g*) punctuation and capitalization; (*h*)

literature; (*i*) history and civics; (*j*) geography; and (*k*) spelling.

Advanced. For grades 7 and 8.

Five forms.

Subjects tested: As in the Intermediate battery and in addition, grammar.

New Stanford Achievement Test. Published by the World Book Company, Yonkers, New York.

Primary examination. For grades 2 and 3.

Five forms.

Subjects tested: (*a*) paragraph meaning; (*b*) word meaning; (*c*) computation; (*d*) arithmetic reasoning; and (*e*) dictation (spelling).

Advanced examination. For grades 4–9.

Five forms.

Subjects tested: (*a*) paragraph meaning; (*b*) word meaning; (*c*) dictation (spelling); (*d*) arithmetic reasoning; (*e*) arithmetic computation; (*f*) literature; (*g*) history and civics; (*h*) geography; (*i*) physiology and hygiene; and (*j*) language usage.

O'Rourke Attainment Tests. Published by the Educational and Personnel Publishing Company. Washington, D. C.

For the elementary grades.

Subject tested: (*a*) vocabulary; (*b*) arithmetic fundamentals; (*c*) arithmetic reasoning; and (*d*) language.

Pintner Educational Achievement Test. Published by the Bureau of Publications, Teachers College, Columbia University, New York, New York.

For grades 4–8.

Four forms.

Subjects tested: (*a*) information (geography, history, health, civics, and others); (*b*) spelling; (*c*) arithmetic (computation and problems); (*d*) vocabulary; and (*e*) reading.

Pressey Second Grade Attainment Scale. Published by the Public School Publishing Company, Bloomington, Illinois.
Two forms.
Subjects tested: (*a*) spelling; (*b*) reading (speed and recognition of words); and (*c*) arithmetic.

Pressey Third Grade Attainment Scale. Published by the Public School Publishing Company, Bloomington, Illinois.
Two forms.
Subjects tested: (*a*) spelling; (*b*) arithmetic (fundamental operations and problems); and (*c*) silent reading (rate and comprehension).

Progressive Achievement Tests. Published by the Southern California School Book Depository, Hollywood, California.
Primary. For grades 1–3.
Two forms.
Subjects tested: (*a*) reading vocabulary (three tests), (1) word form, (2) word recognition, and (3) meaning of opposites; (*b*) reading comprehension (three tests), (1) following directions, (2) directly stated facts, and (3) interpretations; (*c*) arithmetic reasoning (five tests), (1) number and sequence, (2) number and time, (3) signs and symbols, (4) money and (5) problems; (*d*) arithmetic fundamentals (four tests), (1) addition combinations, (2) subtraction combinations, (3) multiplication combinations and (4) problems; and (*e*) language (four tests), (1) capitalization, (2) punctuation, (3) spelling and (4) handwriting.
Tests for grades 4–9 are also being constructed.

Public School Achievement Tests. Published by the Public School Publishing Company, Bloomington, Illinois.
Battery A. For grades 3–8.
Four forms.
Subjects tested: (*a*) reading; (*b*) computation; (*c*) arithmetic reasoning; (*d*) language usage; and (*e*) spelling.

Battery B. For grades 6–8.
Four forms.
Subjects tested: (*a*) grammar; (*b*) history; and (*c*) geography.
Battery C. For grades 4–8.
Four forms.
Subjects tested: (*a*) nature study; and (*b*) health.

Sones-Harry High School Achievement Test. Published by the World Book Company, Yonkers, New York.
For high school and first year of college.
Two forms.
Subjects tested: (*a*) language and literature; (*b*) mathematics; (*c*) natural science; and (*d*) social studies.

Unit Scales of Attainment. Published by the Educational Test Bureau, Minneapolis, Minnesota.
Division I. For grades 3 and 4.
Division II. For grades 5 and 6.
Division III. For grades 7 and 8.
Three forms each.
Subjects tested: (*a*) reading; (*b*) geography; (*c*) literature; (*d*) elementary science; (*e*) American history; (*f*) arithmetic problems; (*g*) arithmetic—fundamental operations; (*h*) spelling; (*i*) English—capitalization; (*j*) English—punctuation; and (*k*) English usage.

Several States have testing programs covering the different school subjects, particularly in the high school field. The tests are usually given at the close of the year and the forms differ one year from another. For this reason such tests do not lend themselves so well for use in diagnosis. However as such programs are improved through the production of comparable forms and the keeping of cumulative records their value in this respect should be increased considerably.[38]

DIFFERENTIAL DIAGNOSIS OF SKILLS AND ABILITIES WITHIN SUBJECTS

In considering the tests which are presumed to be diagnostic in particular subjects, a great number and variety are found. Since it is probable that very few, if any, of these diagnostic tests in single subject fields have been subjected to the techniques given in the second chapter, it is proper that the results on these tests be used with caution. For school systems with facilities for research, it is recommended that all diagnostic tests be subjected to the general test of efficiency as described in the second section. Until further research along this line is accomplished one cannot say definitely with what degree of certainty we can accept differences by inspection between any skills no matter how independent they may seem to be. It seems at first glance, for instance, that differences between skills in arithmetic may be taken at full value. But until investigation proves that the differences exhibited by the raw test results are stable and significant, they cannot be accepted as

[38] For a description and analysis of State programs see Segel, David. National and State Coöperative High School Testing Programs. U. S. Office of Education, Bulletin No. 9, 1933. Government Printing Office, Washington, D. C.

unique skills. When school systems do not have facilities for evaluating their diagnostic tests the only other thing to do is to consider the difference found by testing in the light of their probable errors according to the techniques described for use in making a differential diagnosis.

In the case of groups of high school subjects, such as English, mathematics, social studies, etc., we can find important subdivisions of subject matter with each subject group. It may be true that differences between composition and literature in English and algebra and geometry in mathematics are more significant than differences between general subject groups. They may show differences in fundamental abilities whereas the larger subject divisions mentioned may be more general in nature. The Iowa High School Content Examination and the Sones-Harry High School Achievement batteries do not give so much promise in this direction. The Iowa Silent Reading Test attempts to test more elements in this field and therefore offers more of an opportunity for differential diagnosis. There are many standardized tests in specific high school subjects which may be used for this purpose.

These possibilities for diagnosis within special subject fields will be presented herewith.

Arithmetic.—There are many tests in arithmetic that have been constructed with the view to discovering the sources of error in working arithmetic problems. These tests are listed as follows:

Brueckner Diagnostic Test in Decimals. Educational Test Bureau, Minneapolis, Minnesota. For grades 5–7.

Brueckner Diagnostic Test in Fractions. Educational Test Bureau, Minneapolis, Minnesota. For grades 5–8.

Brueckner Diagnostic Test in Whole Numbers. Educational Test Bureau, Minneapolis, Minnesota. For grades 3–6.

Buswell-John Diagnostic Test for Fundamental Processes in Arithmetic. Public School Publishing Company, Bloomington, Illinois. For the elementary grades.

Clark-Otis-Hatton Instructional Tests in Arithmetic. World Book Company. Yonkers, New York. For beginners.

Compass Diagnostic Test in Arithmetic. Scott, Foresman & Company, Chicago, Illinois. For the elementary grades.

Curriculum Tests in Arithmetic Problem Solving. John C. Winston Company, Philadelphia, Pennsylvania. For grades 3–8.

Curriculum Tests in Arithmetic Processes. John C. Winston Company, Philadelphia, Pennsylvania. For grades 3–8.

Diagnostic Computation Scale. Catholic Education Press, Washington, D. C. For grades 2–8.

Lee Maintenance Drills and Tests in Arithmetic. Southern California School Book Depository, Hollywood, California. For grades 7 and 8.

Los Angeles Diagnostic Tests: Fundamentals of Arithmetic. Southern California School Book Depository, Hollywood, California. For grades 3–8.

Los Angeles Diagnostic Tests: Reasoning in Arithmetic. Southern California School Book Depository, Hollywood, California. For grades 3–9.

Los Angeles Diagnostic Tests: Signs, Symbols, and Vocabulary of Arithmetic. Southern California School Book Depository, Hollywood, California. For grades 3–9.

Lunceford Diagnostic Number Tests. Bureau of Educational Measurements and Standards, Kansas State Teachers College, Emporia, Kansas. For grades 1–4.

Monroe Diagnostic Arithmetic Tests. Public School Publishing Company, Bloomington, Illinois. For grades 4–8.

Objective Drills and Problems in Arithmetic. Benj. H. Sanborn & Co. New York, N. Y. For grades 2–8.

Sangren-Reidy Instructional Tests in Arithmetic. Public School Publishing Company, Bloomington, Illinois. For grades 2–8.

Schorling-Clark-Potter Instructional Tests in Arithmetic. World Book Company, Yonkers, New York. For grades 5–8.

Spencer Diagnostic Tests in Arithmetic. C. A. Gregory & Company, Cincinnati, Ohio. For grades 3–8.

Wilson Fact Inventory and Diagnostic Tests in Arithmetic. University Publishing Company, Lincoln, Nebraska. For grades 2 and 3.

Wisconsin Inventory Tests in Arithmetic. Public School Publishing Company, Bloomington, Illinois. For the elementary grades.

Woody Arithmetic Scale. Bureau of Publications, Teachers College, Columbia University, New York, N. Y. For grades 2–8.

See also arithmetic tests in the field of test batteries.)

Some of the skills in arithmetic are quite specifically due to variations in the immediate school environment whereas others seem to be more fundamental. The differentiation between the minor skills are important to guide the next step in instruction, while the differentiation between the larger fields of arithmetic are of importance in the long range guidance of the individual pupil.

Algebra.—Differential diagnosis in the case of algebra may proceed in about the same manner as in arithmetic. Most of the skills of algebra seem to be dependent upon the direction of the instruction. The same sort of division in algebra probably occurs as in arithmetic, i.e. a division into a manipulation of symbols on the one hand and the analysis of written problems on the other. It would be interesting to discover if the subdivisions as indicated in the Hotz Algebra Tests represent abilities in which individuals differ fundamentally. The possibilities as to diagnostic tests in algebra are as follows:

Douglas Standard Diagnostic Tests in First Year Algebra. C. A. Gregory Company, Cincinnati, Ohio. Two series.

Exercises and Tests in Algebra Through Quadratics. Ginn & Company, New York, N. Y.

First Year Algebra Test. Harlow Publishing Company, Oklahoma City, Oklahoma.

Hart Diagnostic Tests and Remedial Drills in First Course Algebra. D. C. Heath Company, New York, N. Y.

Hotz First Year Algebra Scales. Bureau of Publications, Teachers College, Columbia University, New York, N. Y. Two series. Five parts.

Iowa Unit-Achievement Tests in First Year Algebra. Bureau of Educational Research and Service, State University of Iowa, Iowa City, Iowa.

Nyberg, J. A. Tests and Drills in First Year Algebra. American Book Company, New York, N. Y.

Schorling-Clark-Lindell Instructional Tests in Algebra. World Book Company, Yonkers, New York.

Reading.—The skills in reading are so complex that tests of the simple fundamental processes cannot be made to the same extent as it seems possible in mathematics. For convenience the abilities in reading may be divided into (*a*) silent reading abilities, (*b*) oral reading abilities, and (*c*) auxiliary aids which facilitate the total reading process. The auxiliary skills may be represented by such skills as (*a*) knowledge of the structure of tables of content, and (*b*) skills in using an index.[34] Other auxiliary aids are sometimes found as parts of diagnostic reading test batteries. For the measurement of oral reading the reader is referred to:

Gray Oral Reading Check Tests *and* Gray Oral Reading Paragraphs. Both published by the Public School Publishing Company. Bloomington, Illinois.

There is one thing that can be measured in reading which may sometimes give a lead to the type of reading skills which are weak or strong but which in itself is not considered by the writer to be a reading skill. Speed of reading is referred to. Speed of

[34] For a discussion of some of these aids see Brueckner, L. J. and Melby, E. O. Diagnostic and Remedial Teaching. Houghton, Mifflin Company, New York, N. Y. 1931, page 255.

reading is more of a general measure of reading ability than anything else. Therefore, such tests will be omitted from our further discussion because their use in differential diagnosis is not as yet clear.

Although the relation of eye movements and vocalization to silent reading ability was shown by Miles and Segel[40] to be high, there is no agreement as to what particular types of reading these physical aspects are associated with, and until further research is accomplished they should be used as an instrument of general diagnosis in reading such as tests of speed of reading. The work of Marion Monroe might be mentioned in this connection as an attempt to solve the difficulties in reading from the physical side. Monroe's procedure has to do with substitutions, omissions, etc., in reading, and diagnosis of the facility in reading mirror writing. The Pressey Diagnostic Tests in Fundamental Reading Habits is a test of eye movements and vocalization.

The Gates Silent Reading Test has subdivisions in its primary series of (*a*) word recognition; (*b*) phrase and sentence reading; and (*c*) reading of paragraphs of directions; and for the upper elementary grades, (*a*) reading to appreciate the general significance of a paragraph; (*b*) reading to predict the out-

[40] Miles, W. R., *and* Segel, David. "Clinical Observation of Eye Movements in the Rating of Reading Ability." Journal of Educational Psychology, Vol. 20 (October, 1929), pp. 520–29.

come of given events; (*c*) reading to understand precise directions; and (*d*) reading to note details.

For more searching instruments for use in differential diagnosis Gates[41] recommends, in addition to the tests of the abilities measured by the Gates' tests already mentioned, the following: (*a*) word pronunciation; (*b*) phonetic abilities tests; (*c*) visual perception tests; (*d*) visual analysis and recognition tests; (*e*) visual memory span tests; (*f*) auditory functions tests; (*g*) auditory memory span tests; and (*h*) associative learning tests. Other tests measure various types of reading.

The Ingraham-Clark Diagnostic Reading Tests have nine subdivisions, measuring ability to recognize (*a*) word form; (*b*) likeness and differences in words; (*c*) identification by visual and auditory stimuli; (*d*) association of word meanings; (*e*) opposites and similarities; (*f*) following printed directions; (*g*) distinguishing relevant and irrelevant statements; (*h*) making inferences and deductions; and (*i*) organization. The Sangren-Woody Reading Test has the following subtests: (*a*) word meaning; (*b*) rate; (*c*) fact material; (*d*) total meaning; (*e*) central thought; (*f*) following directions; and (*g*) organization. The Greene-Noar Self-Diagnostic Reading Tests contains subdivisions testing (*a*) recreational reading; (*b*)

[41] As described in Gates, A. I. The Improvement of Reading. MacMillan & Company, New York, N. Y., 1927.

selecting facts; and (*c*) understanding directions. The Lee-Clark Reading Test—Primer and First Grade contains a unique approach to beginning reading by having sections covering (*a*) auditory stimuli; (*b*) visual stimuli; and (*c*) following directions in the Primer Test, and (*a*) auditory stimuli; (*b*) visual stimuli; (*c*) following directions; (*d*) completion; and (*e*) reference, in the First Grade Test. The Iowa Silent Reading Tests—Elementary Test contains subdivisions on (*a*) paragraph meaning in science and history; (*b*) word meaning in general vocabulary and subject-matter vocabulary; (*c*) selection of central idea of paragraph; (*d*) sentence meaning; (*e*) location of information; and (*f*) a division for the determination of the rate of silent reading. That reading may be divided into types somewhat independent of each other seems probable. Just what types are most important cannot be definitely stated without further research and possibly until other types of tests have been evolved.

The silent reading tests discussed in this section are as follows:

Gates Silent Reading Test. Bureau of Publications, Teachers College, Columbia University, New York, N. Y. Primary series (for grades 1 and 2), and grades 3–8 series.

Greene-Noar Self-Diagnostic Reading Tests. D. C. Heath & Company, New York, N. Y. For grades 3–8.

Ingraham-Clark Diagnostic Reading Tests. Southern California School Book Depository, Hollywood, California.

Primary series for grades 1–3. Intermediate series for grades 4–8.

Iowa Silent Reading Test—Elementary Test. World Book Company, Yonkers, New York. For grades 4–8.

Lee-Clark Reading Tests. Southern California School Book Depository, Hollywood, California. Primer Test covers the first 89 pages of the Primer "Billy and Terry," and the First Grade Test covers the first 111 pages of the First reader "Jack and Jane."

Monroe's Diagnostic Reading Examination. C. H. Stoelting Company, Chicago, Illinois.

Multiple Skill First Grade Reading Scale. Educational Test Bureau, Minneapolis, Minnesota.

Pressey Diagnostic Tests in Fundamental Reading Habits. Public School Publishing Company, Bloomington, Illinois. For grades 2–8.

Pressey Diagnostic Reading Tests. Public School Publishing Company. Bloomington, Illinois. For grades 3–9.

Sangren-Woody Reading Test. World Book Company, Yonkers, New York. For grades 4–8.

(Several of the test batteries on the elementary-school level also contain sections on various phases of reading.)

English.—We may classify the English field in high school into two subdivisions—grammar, language usage, and composition on the one hand, and literature (including various phases of reading) on the other. These rather large subdivisions of the English field will most properly be considered as being separate subjects. As concerns English, then, a differential diagnosis might be made as between English in general and science or mathematics or the social studies; and again as between the two subdivisions of the English field as mentioned, i.e. language usage and literature; and then again in each of these two subdivisions there are possibilities for diagnosis.

English usage may be measured by composition scales or by tests of the knowledge of good language usage made up in new type forms. Composition scales are as follows:

Hudelson English Composition Scale. World Book Company, Yonkers, New York. For grades 4–12.

Minimal Essentials Test in English Composition. Public School Publishing Company, Bloomington, Illinois. For grades 3–6.

Nassau County Supplement to the Hillegas Scale and Thorndike Extension of the Hillegas Scale. Bureau of Publications, Teachers College, Columbia University, New York, N. Y.

Van Wagenen English Composition Scale. World Book Company, Yonkers, New York.

Willing Scale for Measuring Composition. Public School Publishing Company, Bloomington, Illinois. For grades 4–9.

There are many tests of good usage, among which are:

English Minimum Essentials Test. Public School Publishing Company, Bloomington, Illinois. For grades 8–12.

Iowa Elementary Language Tests. Educational Test Bureau, Minneapolis, Minnesota. For grades 4–9.

Kirby Grammar Test. Bureau of Educational Research and Service, State University of Iowa, Iowa City, Iowa. For grades 7–12.

Modern School Achievement Tests—Language Usage. Bureau of Publications, Teachers College, Columbia University, New York, N. Y. For grades 3–9.

New Stanford Language Usage Test. World Book Company, Yonkers, New York. For grades 4–9.

O'Rourke Grammar Achievement Tests. Educational and Personnel Publishing Company, Washington, D. C.

Public School Achievement Tests—Grammar. Public School Publishing Company, Bloomington, Illinois. For grades 6–8.

Public School Achievement Tests—Language Usage. Public School Publishing Company, Bloomington, Illinois. For grades 3–8.

Test for Correct English. Houghton, Mifflin Company, Boston, Massachusetts.

Wilson Language Error Test. World Book Company, Yonkers, New York. For grades 3–12.

Literature tests may be made up for particular classics or for general purposes. It follows that the tests for particular classics would be of value only if they were studied by the pupils concerned. Literature tests are as follows:

For particular classics:

Accomplishment Tests in Literature. Lyons & Carnahan, New York, N. Y.

Exercises and Tests on English Classics. Ginn & Company, New York, N. Y.

Hadsell-Wells Objective Test on English and American Classics. Harlow Publishing Company, Oklahoma City, Oklahoma.

For general use:

Barrett-Ryon Literature Test. Bureau of Educational Measurement and Standards. Kansas State Teachers College, Emporia, Kansas.

English Literature Test. Center for Psychological Service, Washington, D. C.

Stanford American Literature Test. C. A. Gregory Company, Cincinnati, Ohio.

Stanford English Literature Test. C. A. Gregory Company, Cincinnati, Ohio.

There are other English tests which are not specifically language usage or literature tests. Such tests would be of value in comparing the pupil's work in English in general with other school subjects. The language usage ability may be separated into finer ability groups such as are represented by the various grammatical forms. There are many tests for attempting their measurement. Some of them are as follows.

Correct English Usage Test (Diagnostic). Harlow Publishing Company, Oklahoma City, Oklahoma. For high school.

Diagnostic Tests in English Composition. Public School Publishing Company, Bloomington, Illinois. For grades 7–12.

Franseen Diagnostic Tests in Language. C. A. Gregory Company, Cincinnati, Ohio. For grades 3–8.

Leonard Diagnostic Test in Punctuation and Capitalization. World Book Company, Yonkers, New York.

Pribble-McCrory Diagnostic Elementary Language Tests. Lyons & Carnahan, New York, N. Y. For grades 3–6.

Pribble-McCrory Diagnostic Tests in Practical English Grammar. Lyons & Carnahan, New York, N. Y. One test for grades 7 and 8 and another for high school and college.
Purdue Diagnostic English Test. Lafayette Printing Company, Lafayette, Indiana. For grades 6–12.
Tressler Grammar Minimum Essentials. Diagnostic. Practice and Mastery Tests. D. C. Heath & Company, New York, N. Y.
Wakefield Diagnostic English Test. C. A. Gregory Company, Cincinnati, Ohio. For grades 9–12.

Few tests for the diagnosis within the subject of literature appreciation have been developed. A possible valid test in this field is:

Tests for the Appreciation of Literature. Public School Publishing Company, Bloomington, Illinois. For high school and college.

Social Studies.—Brueckner and Melby[42] have suggested dividing tests in the social studies into (*a*) background tests, (*b*) information tests, (*c*) thought tests, (*d*) tests of understanding, and (*e*) attitude tests:

For background:
Kepner Background Test in Social Studies. Ginn & Company, Boston, Massachusetts.
For information:
Van Wagenen American History Scales. (Revised edition.) Bureau of Publications, Teachers College, Columbia University, New York, N. Y. For grades 5–12.

[42] Brueckner L. J. *and* Melby, E. O. Diagnostic and Remedial Teaching. Houghton Mifflin Company. 1931. pp. 449–69.

Burton Civics Test. World Book Company, Yonkers, New York. For grades 5–9.

Brown-Woody Civics Test. Part II. World Book Company, Yonkers, New York. For grades 7–12.

Hill Civics Tests. (Hill Tests in Civic Information and Attitudes and Hill-Wilson Civic Action Test.) Public School Publishing Company, Bloomington, Illinois. For junior and senior high school.

For thought:

Brown-Woody Civics Test. Part III. World Book Company, Yonkers, New York. For grades 7–12.

Barr Diagnostic Tests in American History. Part V. Public School Publishing Company, Bloomington, Illinois.

Tests of understanding:

Pressey-Richards American History Test. Public School Publishing Company, Bloomington, Illinois.

Tests of attitude:

Hill Civics Tests. (Hill Tests in Civic Information and Attitudes *and* Hill-Wilson Civic Action.) Public School Publishing Company, Bloomington, Illinois.

This field is undergoing experimentation at this time so that there may be soon a better analysis of the abilities in the social studies. The differential diagnostic technique should be a very important procedure in this field. Differential diagnosis in the social studies will aid in determining the very materials of the curriculum because of the differential emphasis which might be desired for the different abilities discovered. It seems to be necessary to check the differential diagnosis in this field with the techniques described in the previous chapters since the various abilities are quite broad and may or may not be very independent.

The increasing use of tests and measurements in individual instruction and guidance of pupils forces into use certain statistical methods. Among such methods are those which evaluate the efficiency of tests to make a differential diagnosis itself. There have been listed many tests which have possibilities in making a differential diagnosis. Some of these tests are undoubtedly of value in making an accurate differential diagnosis while for others the values are problematical. In all cases, however, the decision as to the value of the test should rest upon the evidence in the light of the application of such techniques as are described here.

INDEX