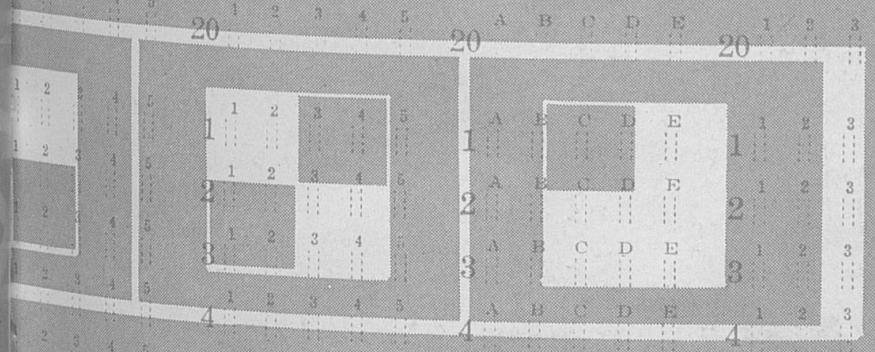
THE STANDARDIZED TESTING PROGRAM



THE STANDARDIZED TESTING PROGRAM

Planning — Administration — Use

ACKNOWLEDGEMENTS

This bulletin became a reality through the efforts of many people. The following deserve recognition for their contributions in planning and writing this bulletin.

Mrs. Justine Campbell, Counselor, Carroll County Schools
Mr. Ordie Davis, Asst., Director, School Relations, University of Kentucky
Mr. John R. Durham, Counselor, Middleburg High School
Mr. E. O. Eiken, Counselor, duPont Manual High School, Louisville City Schools
Mr. Ben X. Freeman, Director, Pupil Personnel Services, Louisville City Schools
Mrs. Ruth Fuller (the late), Western State College
Mr. William C. Foutch, Counselor, Campbell County High School
Mr. Cecil Haven, Counselor, Allen County High School
Mrs. Agnes Hopper, Counselor, Williamsburg High School
Mrs. Josephine King, Counselor, Carter County Schools
Dr. Curtis Phipps, Dean of Students, Georgetown College
Sister Lucile Raley, Counselor Educator, Catherine Spalding College
Dr. Travis Rawlings, Dean, Midway Junior College
Mr. Charles Sanders, Louisville City Schools
Mr. Paul Seyfrit, Dean of Men, Eastern State College
Dr. James Snowden, Department of Education, Eastern State College

For assistance from the Division of Publications and Information goes a word of appreciation.

For Butler County, Edmonson County, Warren County and Barren County schools appreciation is expressed for the use of pictures which were included in the bulletin.

Special recognition is given to Mr. Wayne Ashley, Division of Guidance Services, who gave generously of his time in the planning, writing and editing of this bulletin.

Finally, appreciation is expressed to Dr. Ralph Tesseneer, Murray State College, who did the final editing of this bulletin.

Kearney Campbell, Director
Division of Guidance Services

FOREWORD

Evaluation is an indispensable part of every sound educational program. While there are several tools of evaluation in education—the standardized test has become one of the most important.

To develop a good testing program and to make effective use of test results in guidance, instruction, and administration, all school personnel should have a basic understanding of the standardized test. Those responsible for administering and using standardized tests should become better acquainted with techniques and procedures for interpreting and using standardized tests.

It is hoped that the suggestions contained in this bulletin will be of assistance to school personnel throughout the State who use the results of standardized tests in guidance, instruction, and administration.

Harry M. Sparks
Superintendent of Public Instruction

TABLE OF CONTENTS

SECTION I

THE STANDARDIZED TESTING PROGRAM

The use of standardized tests as psychological tools for measurement of various components of behavior has become widely recognized and accepted. The majority of educators and psychologists will affirm that the schools can do a better job for their students by using the various test and measurement devices. They can serve many purposes in education. They have served many purposes and have changed to a great extent since the beginning of the last century. Noteworthy events from the historical background of the testing movement would include:

1. early survey of Boston School Children stimulated by criticisms of schools by Horace Mann in early 1800's
2. emphases placed on measurement by early pioneers, Galton and Cattell
3. contributions to the testing field by E. L. Thorndike and his students at the close of the 19th century
4. significant contributions in intelligence testing by the French Psychologist, Alfred Binet
5. development of group intelligence tests and rating scales for use by the Army in World War I which gave tremendous impetus to testing in schools
6. widespread use of tests by all branches of the armed forces during World War II
7. use of tests by Veterans Administration in their advisory centers
8. provisions of the National Defense Education Act for testing of secondary school students
9. continuous improvement and development in test construction from their earliest beginnings to today.

A statement from Chauncey and Dobbin, "Testing, It's Place in Education Today", will illustrate current thought for use of tests in educational measurement.

> When tests are taken as they are intended to be taken as interesting job samples on which the student can try out his hard-won skills, the process is stimulating for most students and enjoyable for many. Learning is a satisfying experience; and as testing is the proving part of that experience, it, too, can be satisfying and enjoyable. More important though, than the satisfactions it gives directly to students, is the guidance that

good testing gives to the teacher, the counselor, the admissions officer, and the student himself.[1]

NEED FOR TESTS IN EDUCATION

During the 1960's there is a continued and growing interest in testing and testing techniques at the national, state and local level. Many local districts are developing their own tests and testing programs. National and state testing progrms are being expanded and offered to local districts, in addition to, or as a part of their local programs. The expansion of the National Defense Education Act of 1958 has provisions which makes it possible for districts to test all pupils in the elementary, secondary schools and in junior colleges.

The tremendous increase in school enrollment at all levels in the past few years accompanied by an increase in the holding power of the schools and the demands of a highly technical age have made it increasingly difficult to develop and maintain an effective educational program for all boys and girls. School personnel are obligated more today than ever before to study their students thoroughly, and change the curricula accordingly, to meet the needs of the students.

INFORMATION TESTS CAN PROVIDE

An excellent discussion of testing program purposes and the contribution testing can make is found in the publication, "A Proposed 12 Year Testing Program."

"Tests can provide much of the information needed to:

—understand the pupil better
—guide his learning and achievement
—evaluate his achievement in terms of his ability
—discover and guide his interests
—adjust to and provide for individual differences
—help each student understand himself
—give the student background for course selection and career planning.[2]

---

[1] Henry Chauncey and John E. Dobbin, *Testing: It's Place in Education Today.* Harper & Row, Publishers, 1963 p. 184.

[2] Kent Area Guidance Council, *A Proposed 12 Year Testing Program.* Ohio Scholarship Tests, State Department of Education, 1959, p. 1.

Purposes for which tests are used are outlined in three broad categories by Hill.

1. For instruction
   —diagnostic, the study of pupils strengths and weaknesses in relation to learnings desired
   —placement, the study of pupil growth for the purpose of placing the child in the most helpful learning situations.

2. For guidance
   —for help in the selection of courses and subjects
   —for help in making choices related to vocational planning
   —for help in counseling and group guidance regarding educational and vocational planning and in assisting pupils to meet personal and social problems.

3. For administration
   —to provide data helpful in determining the efficiency of the school system as a whole
   —for help in determining promotion policies
   —for help in determining policies regarding the grouping of pupils for instruction[3]

Obviously tests have many and various uses, and if used for the purpose of making specific contributions to the total school program, the testing programs may differ with each school system. The same test may be used in different schools for the same or different purposes, and at the same time different tests may be used in different systems for the same purpose. This will point out the versatile nature of tests and the necessity for purposeful testing as a service to the system in achieving and improving the schools objectives.

The administrator may use a survey battery of achievement to study performance trends in subject matter areas of their student population for curricular revision. The principal may use the same results as an aid in the grouping of students for instruction. The counselor may use the results to aid the student in evaluating his strengths and weaknesses. Results from the same battery may be used in conference with parents to interpret student progress. Teachers may use the results to assess effectiveness of teaching methods or to diagnose student weaknesses. At the same time the testing director

---

[3] George E. Hill. *The Staff Evaluate the School's Testing Program.* Center for Educational Research, College of Education, Ohio University, Athens, Ohio, Pupil Services Series, 1960 No. 1, p. 4.

may use the results to furnish data for compilation of local norms, or for comparison with a larger population.

There are many other ways that tests and measurement devices can be helpful to the school system. Characteristics unique to each school or system, the necessity of testing for a purpose, readiness of faculty to use the results, and many other factors necessitate *a program* to fit *a system* designed to serve *a function* based upon sound educational objectives.

A BASIC TESTING PROGRAM

To define testing programs in terms that would describe the large school systems, as well as the small independent systems requires a broad definition in very general terms. Findley in the NSSE Yearbook, 1963, defines it as:

> . . . consisting of the systematic use, at more than one level and in more than one area, of one or more tests where the selection, administration, recording, and interpretation of the tests are all planned and conducted in close accord with the major objectives of the educational program.[4]

Attempting to outline a minimum or maximum testing program applicable to all systems would be impossible. From attempts to develop a testing program conforming to the definition above will come a framework which passes through a minimal state of development, upon which the development of an expanded program will be possible. A minimum testing program becomes a base beyond which the school system may go, as qualified personnel, facilities, and other criteria may determine.

Recognizing that there are certain needs common to all systems many authorities in educational measurement have listed a minimum program, not intended to identify *the program* for *the school,* but to serve as a guide to identification of needs common to all and to indicate successful testing patterns which have served to provide commonly needed information about each student.

The problems facing educators are complicated by a more informed and concerned public, pressured by the rapidity of change, and further complicated by the basic belief inherent in a democratic

---

[4] Warren G. Findley. *The Impact and Improvement of School Testing Programs.* NSSE Yearbook LXII, Part II, 1963. p. 3.

society, that each student has the right to make his own decisions or choices.

The following statements from an annual report by Henry Chauncey, President of the Educational Testing Service, illustrates this point and implies an ever increasing obligation on the part of educators to provide assistance to each student:

> "In an American educational system, the student is offered a number of choices about himself and his career as he progresses through school, choices that he is enccuraged to make, and, hopefully, is equipped to make. The variety and number of choices are characteristic of American Education, as is the freedom of the student to make the choices himself. Almost without exception, the educational systems of other nations in the western culture have only one of two choice points in the career of the student, and almost always the choices are made for him by someone else.
>
> Here as in so many human endeavors, the extent of free choice is sometimes limited by unequal school standards, by variable home and community environments, and by the lost development of potential in the early childhood and elementary school years. But the fact remains that the framework for free choice and a great variety of higher educational opportunities, are open to the American student."[5]

As they meet and cope with this challenge of change, educators are becoming more aware that purposeful action must be taken, individual needs must be better met, and that the individuals basic rights must be recognized and assured. Prerequisite to all of these is more and better information about each student, the curriculum, and the instructional program accompanied by continuous and systematic evaluation which affirms that sound educational objectives in harmony with sound educational philosophy are being reached. Psychological tests and inventories can make a valuable contribution in assisting the total educational program in achieving its objectives.

As pointed out by Willey and Andrews[6], schools should not administer any more tests than necessary. An extra test, administered, then filed and forgotten is a waste of time which could be most profitably used in developing ways and means to interpret

---

[5] Chauncey, Henry. *Annual Report, 1962-63.* Educational Testing Service.

[6] Willey and Andrews. *Modern Methods and Techniques in Guidance.* Harper & Bros., Publishers, N.E., 1955, p. 167.

available test scores and place them in the hands of individuals who can use them to help students help themselves.

The following minimum testing program for guidance follows a pattern of testing for ability and achievement at the elementary, junior and senior high school. It is designed to provide needed information for guidance purposes.

Grade	Time of Administration	Type of Test
First	Beginning of Year	Reading Readiness
Second		
Third	Beginning of Year	Mental Ability
	End of Year	Achievement tests
Fourth	Beginning of Year	Diagnostic reading test
Fifth		
Sixth	End of Year	Achievement tests
Seventh	End of Year	Mental ability
Eighth	End of Year	Achievement tests
Ninth	Beginning of Year	Mental ability
	End of Year	Achievement tests
Tenth	Beginning of Year	Interest test
Eleventh	End of Year	Achievement tests
Twelfth	End of Year	Achievement tests

Tests of mental ability and achievement are recommended at the levels usually considered as transition points in the school life of the student. As mental ability and tests of achievement normally have high positive correlation coefficients, duplicative testing is avoided by placing them at different levels in the minimum program. An interest inventory is recommended at the beginning of the tenth grade as the student reaches a point of greater specialization. Special aptitude and personality measurement devices are not included in the above program; however, many schools may want to include them if they can administer and use them.

As there is a greater need for a general ability score early in the student's school life, many authorities recommend general ability testing at the end of Kindergarten or at the beginning of the first grade in addition to the mental ability test recommended above at the beginning of the third grade. Others, because of low correlation coefficients of early ability scores with those obtained later recommend the use of Reading Readiness and Reading Achievement at

the first and second grade level. The mental ability test should be as culture free as possible and not loaded with reading.

Schools with broad achievement testing programs for guidance, instructional, and administrative purposes administer an achievement battery at the end of each grade level. In this case the need for group mental ability testing at the upper grade levels will decrease, and the addition of special aptitude and personality devices, supplemented by individual intelligence tests will change the format of the above minimum program. The suggested minimum program could be expanded considerably by shifting levels of mental ability tests, including an achievement battery at the fifth and seventh grade, and by substituing a multi-factor aptitude test battery at the ninth grade for the achievement battery. Aptitude testing could then be enriched by using the USES General Aptitude Test Battery at the twelfth grade. Problem checklists used as pupil's statements of their problems or concerns can be of value in individual work with students and are used by many schools. Although research data is consistently discouraging as to the value of personality tests or inventories in group wide assessment, they may be added if qualified personnel to use them are available.

Many authorities in measurement recommend a basic or minimum testing program similar to the one listed above. Determining the program needed in any particular school will require careful planning and coordination as it develops from a minimum to a complete testing program.

SECTION II

PLANNING THE TESTING PROGRAM

Testing, like many other aspects of education will not be adequate without proper organization and planning. To insure an effective program the school administrator must assume the central point of responsibility. Spence[1] has suggested that a good testing program should be supplementary not duplicative, usable not confusing, economical not burdensome, comprehensive not sporadic, suggestive not dogmatic, progressive not static. Several questions, answered affirmatively, will provide a start toward initiating or revising a testing program that could be described by the positive aspect of the above phrases. Some of these questions are:

1. Do we have a well defined philosophy of education understood by the staff?
2. Do we have a well defined philosophy of measurement understood by the staff?
3. Is the role of measurement as it relates to the attainment of educational goals and objectives clarified?
4. Do we need a standarized testing program?
5. Does the program need to be planned and coordinated to meet the needs of the system?

A negative answer to any or all of the above questions will pinpoint the necessity for a system-wide testing committee to determine the needs to be served, to assure full cooperation and participation of the entire staff, and to develop an organizational pattern or design.

Testing Committee. Composition of the committee may differ with each school system as it must function under the conditions existing in the system. Representatives of the entire staff are recommended participants. These may be representatives from the individual elementary and secondary school guidance or testing committees and the system-wide testing coordinator designated to coordinate the development and operation of the total program.

No blueprint or universal application can be given to those planning a testing program, but a committee contemplating such a

[1]Ralph B. Spence, *A Comprehensive Testing Program for Elementary Schools,* Teachers College Record, Volume 34, pp. 279-284, Jan. 1933.

"Planning the Standardized Testing Program"

program should give consideration to these factors; existing personnel for administering the testing program, experience of teachers with test and testing techniques, attitude of school and community toward standardized tests, and present and previous policies of the school or system.

Steps in the testing program. Inability or failure to answer the question, "Where do we start?" has resulted in many failures to keep testing in its proper perspective. The following steps or stages through which a complete program must pass listed by Ross and Stanley[2] are included to assist the committee in developing a functional program.

1. Determining the purpose of the program
2. Selecting the appropriate test or tests
3. Administering the tests
4. Scoring the tests

[2] C. C. Ross and Julian C. Stanley, *Measurement in Todays Schools,* Prentice Hall, Inc. N. J., 1960, p. 212.

5. Analyzing and interpreting the scores
6. Applying the results
7. Retesting to determine the success of the program
8. Making suitable records and reports.

Planning should begin with a study or review of the school's educational objectives and the translation of these objectives into specific behavior patterns. Vague generalizations should be changed to concrete and attainable educational goals. Standardized tests should be selected to provide the data which are best obtained through the testing programs and plans made to secure other data by more appropriate techniques. Planning should include preparation of a detailed outline of the ways in which tests results will be used, how scores will be made available for use, and the *assistance to be provided* for those who will be using them.[3]

Determining the purpose. After translation of educational goals into working hypotheses, relating of these to the system-wide philosophy of measurement, isolation of those which testing can assist in obtaining, and determining of those which other techniques can be more helpful, the committee will have a list of purposes for a testing program, common to systems who ascribe to identical goals, and unique to their own system.

The following is an example of how a general objective can be hypothesized and how purposeful testing can make a significant contribution toward its attainment.

Objective: Preparation of each student for effective American citizenship.

This is a general objective of courses in social studies in the instructional program.

Hypothesized this statement could read:

Hypothesis: The social studies program helps students to acquire knowledge and skills essential for good citizenship.

Purpose for Testing: Periodic assessment at selected grade levels of standardized achievement test to:

1. identify students strengths and weaknesses in the area

[3] State Department of Education, *The Testing Service,* Hartford, Connecticut, June 1959, p. 3.

2. provide information for curicular revision or modification
3. evaluate effectiveness of teaching methods being used
4. assist student in evaluation of their progress
5. provide comparative data with a larger population indicating the degree of understanding in the area being tested.

Command of the fundamental processes is a very broad generalization commonly found in the schools statement of goals and objectives. This encompasses all areas of the curriculum. For each area the same and many other purposes could be listed for which standardized tests of achievement can make a contribution toward realization of a desired goal.

To provide for the student such assistance as may be needed for the development of his aptitudes and the full utilization of his abilities is a general guidance objective for which assistance from the testing program is needed. Testing may serve other purposes providing group data needed for planning curricular and instructional programs.

After the purpose is determined, the other seven stages listed above must be planned. The succeeding chapters are concerned with selection, administration, scoring, interpreting and using test results in the school program. The major focus is upon testing in the guidance program. The State Department of Education, Hartford, Connecticut, in a bulletin entitled, "The Testing Service: A Design for Program Development listed the following principles:

Guiding Principles for a Testing Program

1. Since tests provide data about pupils and the school program, the testing program should be planned as an integral part of the educational process. Tests should be used as an aid to planning and carrying out instructional activities, but they should not determine the curriculum. They are its servants rather than its master.
2. Testing should be used as an aid to pupil self-understanding and to vocational and educational planning. But testing in itself is not guidance nor should decisions in guidance be made on the basis of tests alone.
3. Tests should be selected to meet the needs of the local school. The school staff should participate in considering these needs and the ways in which tests can contribute to the realization of educational objectives. The nature and

purposes of testing should be understood also by the student body and interested persons in the community, especially parents.

4. Testing should be planned to include measurement of several pertinent aspects of pupil ability and achievement. The testing program should not consist of isolated segments.
5. The quantity and type of testing should be limited by the time and ability of the staff to use the results. What counts is not how many tests are given but what use is made of the results. A testing program should not attempt to do so many things that none of them is done well. The purpose most important to the school should have priority. Other purposes may be left for later development.
6. Tests given to all pupils in a grade should be supplemented by additional tests for selected pupils and groups as needed.
7. Evaluation is a continuous process and tests should be administered at several grade levels. Test information should be available when it will be used, especially at critical and transition periods in the life of each pupil.
8. A test should be used only for those purposes for which it is valid. There must be an understanding of what a test can do and what it cannot do. Test scores should be interpreted in relation to other evaluative data. The significance of the results of any test is dependent on many factors and relationships.
9. Tests should be interpreted and used only by competent persons. The level of competence required depends on the test and the use to be made of it. Test scores should be made available only to persons who can make proper use of this information. Interpretations of the meaning of the test results can be made to others who are entitled to this information. The confidential nature of test information should be respected.
10. Provision should be made for recording test data cumulatively and in a way that will make such data useful. This takes time, clerical help, and professional knowledge of tests.
11. The testing program should be under the direction of a person who has the knowledge, skill, time, and the assistance needed to assure its proper functioning.
12. The testing program should be evaluated regularly to determine whether the tests are serving the purposes for which they were selected, and to determine whether these purposes continue to be important.
13. The perfect testing program has not yet been developed. Expectations must be kept as a realistic level. The best guarantee of a good program is a plan for continuous de-

velopment which includes modest objectives and sound purposes designed to meet specific needs.[4]

Careful consideration of the preceding principles individually and collectively by the committee will promote orderly progression through the remaining steps or stages in developing a complete testing program. The types of test to be given and the grade levels at which they will be used to meet the pre-determined needs of the local school must come before the actual test selection. A general discussion of the most common types of tests and inventories used in the complete testing program is included in the following chapter. Succeeding chapters deal with test selection, administration, scoring, interpreting and using test results in the school program.

---

[4] *Ibid.*, p. 10.

SECTION III

TYPES OF TESTS AND INVENTORIES

The multiplicity of test types make any attempt at classification difficult. Many ways of classifying standardized tests can be found in the educational literature. The following discussion of types of tests and inventories is based on the classification by Cronbach who lists two major classes; those which seek to measure the maximum or best performance of a student and those which attempt to measure the typical performance or what the student is likely to do in broad situations.[1] Many other classifications can be found. Representative of these are: verbal, non-verbal, group-individual, general-specific, rating scales, checklists and many others.

The above classification by Cronbach, although general, avoids sub-divisions which places the same test, for discussion purposes, in more than one sub-division. A standardized test is defined as one in which the procedure, apparatus, and scoring have been fixed so that precisely the same test can be given at different times and places.[2]

Test of Maximum Performance

Tests of maximum performance are referred to as tests of ability. These include group tests of mental ability, proficiency tests called achievement tests and special aptitude tests primarily used to predict success in some future undertaking. These test types have common elements and their accuracy is initially determined by how well they actually reflect the best the student can do on the samples included in the instrument.

Achievement Tests. Achievement tests are tests of specific skills such as reading, arithmetic, social studies, etc. At practically all grade levels and in most content areas, tests to measure the student's proficiency are available. Here again accuracy of the tested level of attainment is determined by the degree of best performance reflected and adequacy of the instrument in sampling the area tested. Two basic types are *survey* and *diagnostic* achievement tests, both designed to measure levels of attainment in specific areas. Survey

---

[1] Cronbach, Lee J., *Essentials of Psychological Testing,* Second edition, Harper and Bros., N. Y., 1960, p. 29-34.

[2] *Ibid.,* p. 22.

tests, as the name implies, are designed primarily to yield general information regarding achievement levels, school or system wide. The diagnostic tests are designed to give maximum information for assessing individual or group strengths and weaknesses. Survey instruments are easier to administer and less time consuming than the diagnostic tests which yields the same and more information. The appendix provides a list of frequently used tests of achievement and other types described in this section.

Mental Ability, Intelligence or Scholastic Aptitude Tests.—This type of maximum performance test is called by various names. They are commonly called intelligence, mental ability, general ability, scholastic aptitude and other names. A commonly accepted statement regarding tests of mental ability is that they are instruments yielding an estimate of learning capacity. It is important to remember that the score derived from any test of ability is only a numerical description of a sampling of behavior, at that time, and the capacity estimate is as valid as the tasks are actual samples which measure capacity. It must recognize, also, internal and external environmental conditions may affect the individuals performance at the time of taking the test causing the score to reflect less than maximum performance.

Tests of mental ability and achievement have many common elements, some more than others. Rummel, writing about differences in achievement and intelligence tests stated:

> The typical achievement test usually deals with a narrower range of academic knowledge than the typical intelligence test and includes more items for a comprehensive measurement of that range. On the other hand, many parts of the typical intelligence test involve novel situations not usually found in the typical achievement test. Thus the intelligence tests' novelty and apparent unrelatedness to ordinary school subjects may create greater interest and better testing morale on the part of the students who have antipathies toward one or more subjects than will the achievement tests.
>
> The intent of the intelligence test is not to provide a measure of behavior free of learning; but to measure behavior free of classroom learning. It should contain items which do not penalize the student who has fallen behind in school subjects. It should supplement the achievement testing, classroom and standardized, and provide an estimate of learning potential sampling those constructs recognized as indicative of learning ability.[3]

[3] Rummel, *The Role of Intelligence Testing, Education,* XXCI, October, 1960, p. 79.

Aptitude Tests. A number of multifactor tests have been developed to measure various kinds of developed or special abilities other than academic. There are many differences in two students having the same global general ability score. These include spatial, abstract reasoning, clerical speed and accuracy, motor coordination, mechanical and many other carefully defined special abilities. Multifactor batteries and single area aptitude tests have been developed to provide information descriptive of these abilities. Their use is primarily in educational and vocational guidance to predict future performance in higher educational attainment and occupations. Major out-of-school uses for these tests are for use in guiding persons seeking work. The United States Employment Service is one of the major users of maximum performance tests of this nature.

Tests of Typical Performance

The other class is described by Cronbach as tests of typical performance, and are designed to measure, not what a person can do, but what he does. Measures of typical performance are divided into behavior observations (attempts to study the student when "acting naturally) and self-report devices (asking the student to give a report of his own typical behavior.) In this category are personality tests, problem checklist, interest and many other types of inventories.

Personality Tests. Because of the definiteness of the term "test" these instruments are most often called scales, questionnaires, inventories, etc. They are designed to measure personal adjustment, to describe a person in terms of his personality traits, or to discover the nature of his problems. Accuracy of the obtained score is dependent upon the willingness and ability of the student to present an honest and accurate report of his normal behavior, and adequacy of the instrument to measure what it claims to measure.

Nunnally, discussing measurement of personality, states:

> On the face of it, one would think that self report should provide valuable measures of personality. The individual "accompanies himself" wherever he goes and is in a position to observe himself in many types of social interactions. However, people sometimes are blissfully unaware of their own personality traits. Also, when people are asked to describe themselves on personality inventories, they tend to distort their responses in such a way as to place themselves in the best light. In addition, personality inventories are beset by language

problems, and their results are influenced by numerous response sets.[4]

Data on the validity and usefulness of personality inventories are complex and confusing, but they suggest that such instruments are most helpful in conjunction with other data which may be available in the cumulative record.

Much care should be exercised in the use and interpretation of results, for much damage can be done when inadequately trained people interpret inventory scores.

Interest Inventories. An interest inventory is designed to permit a person to record in an organized manner his likes and dislikes, or preferences, in a number of different situations, such as liking for occupations, choice between activities, or reactions to peculiarities of people. A person's choices are then compared with the choices previously made by successful people in various occupations, or, an inventory may be scored to indicate a larger interest area, such as computational, persuasive, musical, or social service. An interest score is no indicator of the ability of the person, although a person may have a greater chance of success in work which he likes to do. Research conducted thus far has shown that there is considerable stability of interests after grade 10, which suggests a level after which this type of inventory might prove most useful in helping a student to make a vocational choice.

External Tests

A number of external testing programs are available to the schools on a nationwide basis, primarily from non-governmental agencies. An appropriate definition is stated by the North Central Association's Committee on Articulation.[5] External tests are defined as "those which are used primarily by some institution or organization over which the local school has or feels it has no real choice as to whether its students take these tests. Common purposes of external testing programs include, screening for college admissions, selecting scholarship awards, placement of college students, counseling aids, etc.

Many questions will arise as the selection committee attempts to select external tests to be included in the testing program. Tyler

---

[4] Nunnally, Jum C. *Educational Measurement and Evaluation,* McGraw-Hill, Inc., N. Y., 1964, p. 374.

[5] *North Central Association Today,* p. 1, Chicago: North Central Association of Colleges and Secondary Schools, August 1961.

discusses various problems as seen by school administrators which merit careful consideration. They have been severely criticized as time consuming, expensive, duplicative, exerting undue influence on the curriculum, inciting invidious comparison of schools, and exerting undue influence on schools to participate.[6]

There are many programs available which could be described as external. A general description of some of these commonly used in Kentucky schools may be helpful.

The American College Test. The test is an admissions, scholarship, guidance and placement test battery administered at designated test centers throughout the United States. The battery has two parts, consisting of four tests, averaging 45 minutes in length, and a student profile which takes about 20 minutes to complete. The test items are designed to measure ability to perform the kinds of intellectual tasks that college students typically perform. Initial cost of the test battery is $4.00 with an additional cost of $2.00 in event of necessity for re-registration, and $1.00 if additional score reports are desired.

The College Entrance Examination Board Test. The College Board Test most used is the three hour Scholastic Aptitude Test for seniors designed to provide reliable indications of a student's ability to do college work. Additional information is provided by another series of entrance examinations — the subject matter Achievement Tests. Scores are reported on a scale from 200 to 800 with 500 as the mean. Tests are administered at selected test centers throughout the United States at different times during the year. The PSAT is a shorter version of the regular SAT, which has been available since October, 1959. It is designed to assist in the guidance of high school juniors with regard to college.

College Qualification Tests. Test is designed for 11th, 12th, and 13th grade students. Tests are available from companies for use in regular school testing program and scoring services for this test are available from Kentucky Cooperative Counseling and Testing Service. Scores are provided in three (3) areas: *Verbal* (a synonym-antonym test designed to measure verbal ability); *Numerical* (test of basic skills using arithmetic, algebra, and geometric principles designed to measure ability to handle numerical concepts.)

Information Tests. Such tests are of general information in the fields of Science and Social Studies. Scores are reported in percentiles

---

[6] *NSSE Yearbook,* p. 193-210.

with various reference groups. Tests can be administered in the local schools.

General Aptitude Test Battery. Test is a battery of 12 tests using "paper and pencil" group test and some mechanical apparatus designed to measure potential abilities or aptitudes in the following nine areas: Intelligence, Verbal Aptitude, Numerical Aptitude, Spatial Aptitude, Form Perception, Clerical Perception, Motor Coordination, Finger and Manual Dexterity. Scores made on the tests constitute the students occupational pattern which can be compared with standards established for aptitude patterns covering over 500 different occupations.

This test is available through a cooperative program provided by the United States Employment Service. Schools desiring to use GATB should request it from the nearest local office of the State employment service.

National Merit Scholarship Qualifying Test. This test is designed for second semester juniors and first semester juniors. It is used primarily as an instrument for locating scholarship winners. The NMSQT is an initial screening test to select semi-finalists who are later required to take the SAT. Scholarship donors using the NMSQT are primarily colleges, industrial, philanthropic organizations and foundations.

The test is administered in the school and scores are obtained in English Usage, Math Usage, Social Studies, Reading, Natural Science Reading, Work Usage. All are expressed in percentiles and standard scores.

The general descriptions above are not an all inclusive list of external tests. There are many others which could be described as external. Guiding principles which apply to other standardized tests in use will also apply to the national programs. When the school or system can justify the use of such tests in terms of individual pupil growth and value to the overall instructional program they should be used.

SECTION IV

SELECTION OF TESTS

From the common test types, previously discussed, the test selection committee must choose the specific test that best meets their needs. There is an ever increasing number of standardized tests and inventories on the market. A systematic and business like procedure is recommended in the selection process.

Test Selection Criteria. George E. Hill lists the following important criteria which the tests selected should meet.[1]

1. They should provide truly usable information.
2. They should be as easy to administer and score as possible.
3. They should be selected with an eye to cost, although cost is a poor critrion of usefulness.
4. They should be appropriate to the range of abilities of the children to be tested.
5. They should be valid—truly measure what is hoped to measure.
6. They should be reliable—the results should be stable and dependable.
7. They should come in equivalent forms so that re-testing is possible.
8. The manual and other explanatory and interpretive material should be readable and usable.

Characteristics of a Satisfactory Measuring Instrument

Ross and Stanley[2] list validity, reliability and usability as indispensable qualities in any satsifactory measuring instrument. The above criteria are descriptive of these three qualities. As the committee relates the various instruments under consideration to these criteria, they will be studying various technical and non-technical features of a good test. Assuming that a test is generally good is an unwarranted and false assumption. A test is good only for a specific purpose or purposes with specific students. When they decide the test that is better for their students than any other test in terms of the above qualities, they will be saying, "This test will best measure

[1] George E. Hill, p. 5.
[2] Ross and Stanley, p. 106.

what we intend to measure and in our opinion will yield a dependable score.''

Test and measurement terminology would state this as a valid and reliable instrument; test manuals would report evidence supporting the statement as validity and reliability co-efficients of correlation, discussion of which would warrant many technical and statistical concepts beyond the scope of this bulletin.

Validity. Validity is the most important single characteristic of a test. Simply stated, a valid test actually measures what it was intended to measure. It is valid only when it accomplishes the purpose for which it was intended. Determining the validity of the intelligence test is considerably more difficult for the local committee than of the achievement test. In the achievement test, of primary concern to the committee is analysis of its content from the standpoint of agreement with curricular materials being used in the particular grade, subject or area being tested. It is possible for the committee to determine the statistical validity, provided an adequate criterion is available; however, a critical analysis made by principals, supervisors, counselors and teachers as they relate the test content to the materials commonly taught in the particular subject field is most often used to judge the adequacy of the achievement test under consideration.

The content of the many ability tests in use is quite similar, and the tests correlate rather highly with each other. Analysis of the intelligence test content will be of little value to the testing committee because of the general nature of the test tasks or constructs composing the typical test of mental ability. Final selection is most frequently based on practical considerations of test usability rather than highly technical and statistical data. Some of the considerations often determining the final selection of the mental ability test to be used are:

1. Mechanical and administrative details of the total program
2. The grade level at which the test will be used
3. Type of test desired for the level to be tested
4. Pupil population characteristics—racial, ethnic, socio-economic groups often penalized by environmental or cultural influences
5. Standardization population used to establish norms
6. Usability of the instruments under consideration—ease of administering, scoring, interpreting, using, etc.

These considerations will also apply to the other tests being selected. Content is of major concern in selection of the achievement battery as it is very necessary to test what is being taught in this area. As the content nature of ability, special aptitude, interest, and personality instruments are theoretically free of achievement with validity and reliability established by author and publisher, usability concerns are often the major determinants in the final selection by the local committee of these types. This does not relieve the author and publisher of the responsibility of providing a valid and dependable instrument in these areas, nor does it relieve the committee of the necessity of careful study of test reviews and accompanying manuals insuring them that these two main criteria are being met.

Reliability. The dependability of a test is a built-in feature of a good test. The responsibility for a reliable instrument must be assumed by the test author and publisher. Assurance of reliable results is a responsibility of the system using the instrument. Reputable test companies build instruments of sufficient length, standardize the conditions, tasks and scoring procedures, and supply information in form of reliability coefficients. Unless the tests are administered and scored properly and the results used wisely, a reliable test may prove to be highly unreliable.

Usability. Although validity and reliability are of primary concern, the testing committee must remain in the practability realm. They must select instruments which can be justified from the standpoint of funds available, time, ease of administration, scoring and using the results.

Sources of Aid in Selection

The committee can find many aids to assist in the selection process to determine the best instrument for their system meeting the criterion previously outlined by Hill.

Mental Measurements Yearbook. The mental measurement yearbooks[3] contain valuable information. Carefully edited descriptions and critical reviews by test specialists of practically every standardized test published are available in Buros' Mental Measurement Yearbook. The following example is a part of a review of one of the widely used multi-factor aptitude tests.

---

[3] Buros, *Mental Measurement Yearbook,* 5th edition, Entry 605.

Differential Aptitude Tests — VR + NA Score
(Entry 605 in Buros Fifth Yearbook)
Grades 8-12

> This score based on the DAT Verbal Reasoning and Numerical ability subtests, was developed to provide the same coverage as typical C-D level group tests of "mental ability". . . . The test items are well constructed and local reliabilities are usually quite satisfactory. . . . The VR-NA score often correlates quite well with cumulative teacher grade averages.

Authors: G. K. Bennett, H. G. Seashore, and A. G. Wesman
Publisher: The Psychological Corporation

Periodicals. Educational periodicals provide reviews of tests which can also be valuable aids. The following example of the same test used above was selected from the Personnel and Guidance Journal.[4]

> For the school system considering the use of a test battery covering a variety of aptitudes, the Differential Aptitude Test battery provides a series of relatively independent tests which are easily administered and simple to score (particularly the 1963 edition). An abundance of material has been prepared to aid in the interpretation of scores that can be readily profiled. . . . In addition to scores from eight different tests the battery yields a single measure which is probably as good a measure of general academic ability as any test they may now have in use.

The above two examples are not an endorsement of the particular test, but are used to illustrate aids available to the committee to supplement the information provided by the publisher in the manual accompanying the test.

Other Aids. The school system will find it necessary to maintain an up-to-date file of the most important tests. A complete specimen set for each member of the committee of the tests under consideration is recommended. Many committees find the use of standardized test rating scales helpful in selection, not necessarily because of the point value assigned but because of the attention called to significant features of a test which might be overlooked without them. At some of the committee meetings it is wise to bring in consultants. Consultative services are available through the State Department of Education, the Universities and State colleges and other responsible agencies.

---

[4] Jack C. Merwin, *"Review of the Differential Aptitude Test"*, Personnel and Guidance Journal, December 1964, Volume XLIII, No. 4, p. 401.

There are many other sources of aid such as test bulletins and other aids from the publishing company, published and unpublished research studies, checklists and appraisal devices, etc., not discussed in this bulletin which the committee may use.

Irrespective of the specific test selected, the full participation of the staff, the use of available aids, the tests serving specific purposes selected to fit the conditions existing in the system will provide a needed start toward development of a complete, functional program.

SECTION V

ADMINISTRATION OF TESTS

All of the test maker's efforts to construct precise well standardized tests and the testing committee's efforts to select the test best built to meet their educational objectives can be nullified by improper administration of the test. Valid and reliable instruments selected by the testing committee can produce invalid and unreliable results if principles of test administration are not followed closely. Numerous administrative details necessary to obtain a reliable or dependable score must be recognized and carefully coordinated.

Principles of Test Administration

1. Test Administration should be under the direction of a person who has knowledge, skill, time and the assistance needed for its operation.
2. Tests should be given at the time when there is a need for the result.
3. A testing schedule should be arranged to utilize the best personnel, conditions, and facilities available during the testing period.
4. Materials and supplies should be in readiness prior to the testing period.
5. All concerned, administrator, tester, proctor, teacher, parent, and pupil should understand the purpose of the test.
6. The tester should maintain rapport, control and an impartial attitude.

Test Administration. Tests should be given by persons trained to give them. The degree of training necessary is dependent upon the type of test to be given. Usually in a school a trained person has charge of the testing program. This person, dependent upon time and personnel available, either administers or supervises the administration of tests.

Individual tests require the skill of a highly trained psychometrist. Many of the group tests are designed so that with careful attention to the manual a teacher may successfully administer them. Most achievement tests are in the latter group.

Administering the Standardized Test

Ross and Stanley[1] list and discuss three requirements that anyone who administers group tests should meet; ability to read well, to keep time accurately, and the ability to follow directions.

Very few training sessions are necessary for preparation of the entire teaching staff to administer their own group tests. Apparent advantages of teacher administration are: existing teacher-pupil rapport and added teacher insight on pupil behavior. In the regular testing program using group tests of mental ability and achievement, the classroom teacher should usually administer the test.

Time of Testing. Tests should be given only when there is a need for the test results. The purpose for which is given and the time when the results will have the greatest utility should determine the time of year a test is scheduled. The end of the school year is the logical time to test if measuring end-of-course achievement or for grouping students during the summer for the following year. Fall testing might be necessary for evaluation of pupil readiness, planning learning activities for year or determining weaknesses for diagnostic

[1] Ross and Stanley, p. 226.

instructional purposes. Fall testing allows the teacher more time to study the significance of pupil performance on the test and to accurately record the results.

Test Schedules. Test schedules, of necessity, have to be tailored to best utilize the existing facilities and personnel within the school. Administrators seem to agree that there are advantages to scheduling the whole class or group at one time, whether in one large room or in several rooms, rather than one room at a time tested by one examiner. A large comfortable, well-lighted, well-ventilated room with tables is to be desired, especially if separate answer sheets are to be used. In the absence of regular testing rooms, a library or lunch room may offer the next best conditions. Sometimes it is necessary to place chairs with wide arms in the gymnasium and use a microphone and proctors. Sometimes regular classrooms are the only places available. Testees should be seated in such a way to minimize the temptation to cheat.

Test administrators will find it necessary to schedule an in-service training period prior to the test for those teachers or proctors who are involved, during which time the purpose of the test is explained; the manual, the room assignments, distribution and explanation of test schedule. Taking the test to be given or sections of the test is a recommended procedure in the training program.

The actual test administration has been approached in various ways. Some test administrators find it advantageous, especially in elementary schools, to schedule a day for testing in each school with each teacher responsible for the administration in the homeroom. The test coordinator or principal stands by for any assistance needed. In some schools, tests are scheduled at such times that teachers with the most training in testing can be free to administer the tests. In other schools a pre-trained teacher has charge of a room, and taped instructions are used. In other schools the test administrator moves from room to room giving all directions, leaving a designated proctor to record the time.

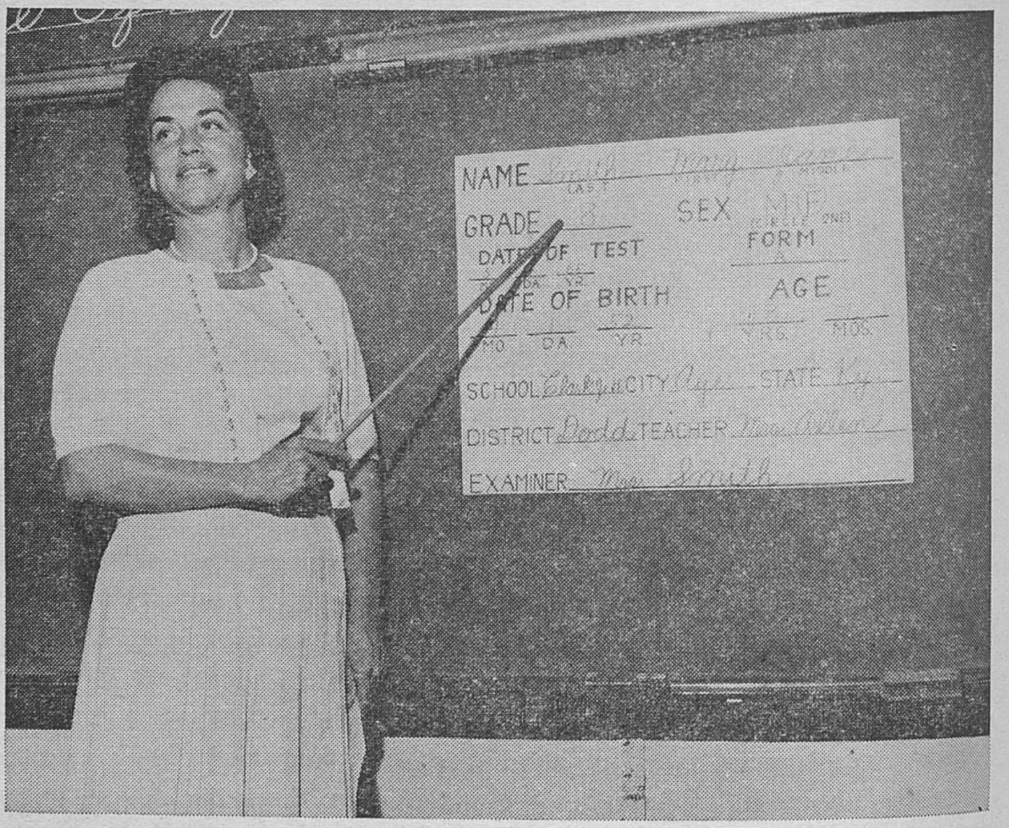
A test battery should be scheduled in more than one test session and at a time when the student body is rested and calm rather than before or after a holiday, or following some strenuous physical education activity. Appendix D. contains a complete list of items which should be carefully studied prior to test administration.

Achieving Test Readiness. Long before the scheduled test, the tester should know that all necessary materials are at hand. On the

day of the test, the tester should have ready for each teacher the correct number of pre-checked test booklets and answer sheets; pencils; scrap paper; instructions; time report; timing instrument; a Testing: "Do Not Disturb" sign and a proctor, or proctors, if needed.

Although schedules are developed cooperatively, before the testing date the examiner or test-coordinator should double-check with the principal to see that all rooms are in readiness, that no unforeseen event has arisen, that classes have been rescheduled if necessary, that no detail has been forgotten.

Student Readiness. The student, too, should be ready. He should not be led to believe his whole future is dependent upon a



Orienting Students to Test-Taking

single test, but he should know that the test is important; for unless he cares about the result he cannot be measured. Orientation of test-takers to promote readiness may be designed with many purposes in mind. Some of these may be:

1. What are standardized tests?
2. What do standardized tests measure?
3. What type questions can I expect?
4. How to take tests.

5. How to mark and change answers on an answer sheet.
6. Should I guess on a test?
7. How is my test scored?
8. What does the score mean?

Each student should be given a written test schedule and appropriate advance information concerning the type of test and its purpose. Nothing motivates a student to do his best more than having a purpose and knowing he will have his test results carefully interpreted.

Examiners Role. During the test the examiner should maintain an impartial and scientific attitude. He must maintain effective control with good rapport, by being firm yet pleasant. He should see to the physical comfort of each student and that each is in a position to hear and to see any demonstration.

The examiner and proctors should give particular attention to the name page or grid section of the test to see that all information needed is accurately supplied. These directions should be given simple, singly, clearly, with proper emphasis. It is imperative that the examiner be thoroughly familiar with the test and that he give the directions exactly as provided in the manual. Test instructions concerning questions about guessing, illness or other irregularities should be followed; but even with the most explicit directions, sometimes a situation arises which requires a decision based on the good judgment of the tester.

After the test starts, proctors should see that each student has started properly and that he is marking the answer sheet satisfactorily.

The tester, moving noiselessly when necessary to observe any individuals unusual behavior, to replace a pencil, or to answer those questions the directions permit, should maintain a quiet atmosphere in which the student can concentrate undisturbed. At no time should the tester or proctor desert his post for even a minute.

An accurate timing device and correct timing is of the utmost importance. Keeping a written record of the time the test starts is due to end greatly lessens timing inaccuracies.

When the final "stop" is given, all materials should be checked for stray marks and then quickly collected in order. Students should be commended for their cooperation and work and, be given an approximate date when the results can be interpreted.

Test booklets and pencils should be counted and stored. A note to replenish answer sheets should be placed in the order file.

SECTION VI

SCORING TESTS AND RECORDING THE RESULTS

If we can assume nothing else about a test score, we should at least be able to assume that it represents near-perfection in accuracy of scoring and of transformation into percentiles, grade equivalents, or I. Q.'s. If this assumption cannot be made and if large scoring errors occur with any degree of frequency, it is doubtful that the tests involved are useful, and they would better be discarded.[1]

Scoring Tests

Special training is necessary for proper scoring of standardized tests. The most commonly used group tests are not so difficult to score that a variety of personnel in each school system cannot be trained in scoring techniques. With today's attempts at mass education, a realistic approach to the problem of scoring tests would tend to favor use of the outside testing agency. These agencies are usually better equipped to do more accurate scoring than the typical school or teacher.

Many principals and authorities recommend use of regular classroom teacher to score her own students achievement tests for immediate results read for use and the opportunity to learn about student errors made on the test. Others because of unfavorable attitudes produced by additional work for the usually overworked teacher prefer to use the clerical staff or machine scoring.

Hand Scoring. If the scoring is done within the school system, adequate controls should be set up to insure accurate scoring. It must not be assumed that merely because the directions are clear, the key complete, the separate answer sheets well designed, and the process entirely objective, perfect protection against errors is therefore afforded.[2]

Numerous studies give abundant evidence to contradict this assumption. They reveal two distinct types of error in scoring, described or discussed by Ross and Stanley as *constant errors* and *variable errors.* A common example of the former type is misunderstanding

---

[1] Leo Goldman, Using Tests in Counseling, Appleton-Century-Crofts, Inc., N. Y., p. 308.

[2] Ross and Stanley, p. 231.

the scoring directions; for instance, by counting omissions the same as errors, when using scoring or correction formula. Such errors are especially serious, because there is no possibility of their offsetting each other according to any so-called "law of averages." Variable errors, on the other hand, sometime tend to make the score too high and at other times too low. While such errors may do serious harm to individual pupils, they tend to cancel each other in group measures such as averages. Examples of variable errors are errors resulting from carelessness, errors in counting the scores, errors in entering the scores on the front of the test booklet or on the record sheet, and errors in adding up the total score.[3] The following statements are stated by Ross and Stanley as a means to prevent errors in scoring.[4]

1. The Scorers must be taught how to score the papers and not merely told how to do it.
2. Check carefully the first few papers marked by beginners to detect errors at the outset.
3. Score each page or part of the test through all the papers in a set before going on to the second page or part of the test.
4. If the scorers work in groups, to reduce the risk of error and increase the speed of scoring, it is usually desirable, for each scorer to specialize in marking one part of the test.
5. Inexperienced scorers should mark each item in a uniform manner.
6. Experienced scorers will save time by marking only the incorrect and omitted items.
7. Each fifth or tenth paper, may be selected and carefully re-scored, and if only an occasional minor error is found in these sample papers, the whole set may be safely accepted.
8. It is important to have some person other than the original scorer check the totals for each part of the test and for the whole test, all substitutions in the scoring formulas, all transcribing of scores, and all transmitting of point scores into derived scores.
9. The standard of absolute accuracy should be accepted by all scorers.

Machine Scoring. It is a time-consuming job to systematically score and organize test results. This job can in many cases, be efficiently and economically farmed out to agencies who have special equipment for this purpose. Many testing agencies may score answer

[3] *Ibid.*

[4] *Ibid.*, p. 233-234.

sheets, summarize test results and provide special statistical treatment of test data.

Proper administration and practicing will relieve the necessity of detailed re-marking of answer sheets after the testing session.

Automation will not insure accuracy of scoring. Careful attention must be given to preparation of answer sheets to be sent to the scoring agency. The answer must be made with the proper sort of pencil and must fill the answer space completely. General instructions for sending answer sheets to be machine scored are listed below. Specific instructions from the scoring agency should be followed closely.

1. Be sure answer forms are properly marked. (Check marks or crosses cannot be properly scored by machine.)
2. Erase all stray marks on or near answer spaces.
3. Only one response (unless item calls for more than one) should appear for each item.
4. Check to insure that all pertinent student data is on the answer sheet, e.g., name, birth date, grade, etc.
5. Be careful not to crease or fold answer sheets.
6. Check for specific instructions from scoring agency being used.

Recording Test Results

The recording of test results is an important aspect of the testing program. Test results are meant to be used in the future as well as at present. The meaningfulness of any one test increases when it can be compared with the results of other tests given at the same time or at earlier dates. All tests should be recorded on some type of permanent record.

In view of the realities of the testing program in terms of time and expense, the recording of test information is usually a compromise. The best method would ideally include all the information concerning the individual. However, a compact and easily filed record is needed. The data should be recorded in an easily accessible place for the personnel who will use the information. Durability and availability are important when considering the recording of test results.

A convenient record card would keep test results together. For example, all achievement test results would be kept in one part of the card, cumulative record, etc. The record should have space provided for special test scores such as those resulting from individual intelligence tests.

All essential information should be given on the testing record. The following information is listed as desirable by Thorndike and Hagen.[5]

1. Grade in which test was given.
2. Date (month and year) of the testing.
3. Complete identification of the test, including name, form and designation of the specific subtests.
4. Norm group used for converting raw scores, where this norm is relevant.
5. All meaningful and reliable part scores, in addition to the total score. Part scores are needed because of their diagnostic and guidance value, much of which is lost in the total score.

---

[5] Thorndike, Robert L., and Hagen, Elizabeth, *Measurement and Evaluation in Psychology and Education,* pp. 461-462, New York: John Wiley & Sons, 1961.

SECTION VII

INTERPRETATION OF TEST RESULTS

The basic assumption must exist from the inception of the testing program that interpretation and usage are just as much parts of the program as are purchase, administration, and scoring. This implies that the school administration must be willing to buy time, human resources, and technical materials as readily as it purchases test booklets.[1]

Provisions for making the results meaningful to staff, parents and students will be necessary if the testing program is to achieve the purposes it was designed to accomplish. Any test given, uninterpreted or misinterpreted, unused or misused, should never have been given.

The two most important things that should happen as a consequence of testing are:

1. Positive changes in *behavior*
   . . of staff toward children in instruction and counseling and over-all planning.
   . . of children themselves

2. Positive changes in attitude and/or *understanding*
   . . of staff and parents about children
   . . of children about themselves[2]

If the consequences of testing listed above are to be realized, a transformation of quantative data to positive changes can take place by proper interpretation and use of the test results. The teacher or counselor attempting to interpret test results must have an understanding of the technical nature of test statistics and must be able to draw valid inferences and make a functional application of these inferences. Merely reporting test scores to students or parents does not assure the interpreter that a positive change in behavior and or attitudes, insights, or understanding has taken place. Pre-conceived notions, aspirations, ability of interpreter to communicate, readiness of the student or parent to receive the information and many other complex factors may influence any desired changes.

---

[1] *A Proposed 12 Year Testing Program,* p. 19.
[2] *Ibid.*

Types of Interpretation

Goldman[3] outlines four types of test interpretation in terms of the kinds of questions that are brought to it.

Descriptive. What kind of person is this man, woman, boy, or girl? How well does he handle numerical reasoning? How does his verbal intelligence compare with his non-verbal intelligence? What does he like to do?

Genetic. How did he get this way? Is his reading deficiency a result of emotional blockings, of inadequate development of basic skills or of lack of interest?

Predictive. How is he likely to fare at college? How much satisfaction will he have in a vocational high school course?

Evaluative. What course should he take? Which college should he attend?

Some counselors—probably the least directive or leading—emphasize the first of these categories, leaving inferences to the client alone, or to a mutual client-counselor discussion. The Genetic kind of interpretation may be less used than the others, but it has an important place in the work of the counselor who is concerned with adjustment and with needed remedial work, since the genesis of a disorder has implications for its correction or remediation. Predictive interpretation is probably the most frequent, since counselors deal primarily with future plans. Finally, evaluative interpretation is an area of controversy and confusion among counselors.[4] When it comes to advice giving, each counselor or teacher will have to decide if he should evaluate.

Merle M. Ohlsen[5] lists the following recommendations relating to the interpretation of test scores.

1. Orientation for acceptance and use of test results should precede testing.
2. Until someone is qualified to use and interpret a test, that test should not be given.
3. Tests and test scores should be released to only those persons who are qualified to use and interpret the tests.

---

[3] Goldman, Leo, *Using Tests in Counseling,* Appleton-Century-Crofts, Inc., N. Y., 1961, pp. 143-144.

[4] *Ibid.*, p. 145.

[5] *NSSE Yearbook*, p. 254-255.

4. Test scores should be interpreted to only appropriate individuals, e.g., students and their parents or legal guardians. In all instances, the pupils' scores should be interpreted within a setting in which unauthorized persons cannot listen in on the interpretation or see the results.
5. Inasmuch as test scores are often misinterpreted by laymen scores should be interpreted for pupils and parents, not merely distributed to them. Furthermore scores should be interpreted when pupils or their parents request information. Unless there is a genuine felt need for information, the odds are against an increase in self-understanding on the part of the parent or even against the acceptance by the parent of information about his child.
6. Before interpreting a test, a teacher or counselor should familiarize himself with the non-test data available on the student. During the test interpretation he should also encourage the pupil to supplement the test results with non-test data.
7. A test interpreter should encourage student participation in interpreting test scores. To help a student recall what a test, or a part of a battery, was like, the teacher or counselor can describe it in a non-technical language, and he will usually find that it is helpful to show the student sample items from the test, before encouraging the student to estimate how well he did on it. If the counselor is to do this successfully, it is obvious that he must know the student and be thoroughly familiar with the test in order to communicate accurate information to the student.
8. The test-interpreter must be sensitive to cues which suggest that the student does not comprehend the information which is being given to him.
9. The student should be encouraged to react to the test results—to raise questions or comment on how he feels about the way the test or tests describe him.
10. There is no justification for arguing with a student about his test scores. Moreover, little can be accomplished by either defending a test or criticizing it. What test-interpreters should do is explain how the results may be used by the student and to make certain predictions, and with what certainty.

The Meaning of Scores

Test scores without meaning are of no value to anyone. Explanation must be in terms the student can understand. Prior to any attempt at interpretation both the interpreter and the student should be aware that any test score reflects a sampling of a job, and there-

fore, cannot be a measure of all of any ability, that performance can change, and that every test score is a comparasion and an estimate. Many kinds of test scores are reported, e.g., raw score, percentiles, I. Q.'s, age and grade equivalents and various forms of standard scores based on the normal or probability curve. The most commonly used scores are discussed below. A graph is included in the appendix which shows the relationship between the various scores, use of which, will make the types discussed more meaningful.

Raw Score. A raw score, has very limited meaning. It is obtained by counting the number of right answers or applying some scoring formula to correct for missed answers. It generally has significance only for teacher-made tests.

Percentile. One of the most widely used score types is the percentile rank. It simply reports the proportion of students in the comparasion group who have been surpassed by the student who made this score. A score with a percentile rank of 40 means that the student who earned it did better than 40 out of each 100 in the comparasion group.

Although widely used because of apparent ease of interpretation, percentiles are often misused or misinterpreted because units of measurement are not equal at different points on the scale and are not subject to further statistical analysis for example, a common misuse is averaging percentiles as if they were actual scores rather than points on a scale. Raw score changes of one point near the middle of the distribution may result in a percentile change of five points due to central tendency. Whereas, near the bottom or top of the scale it may take a raw score change of five points to result in a percentile change of one. These characteristics of percentiles may result in an over-interpretation of differences near the middle and an under-interpretation at the top and bottom of the distribution.

Grade Equivalents. Grade equivalent scores are very popular especially in achievement testing. They are commonly misunderstood by students, parents, and teachers. It is a comparasion score determined by averaging scores of children in two successive grades and dividing the difference between the two grade averages by ten and designating this difference as a decimal, e.g., 6.1 grade equivalent is the average score of sixth grade plus one tenth of the difference between sixth and seventh grade average. For example, the arithmetic score of a seventh grade student on an achievement test designed for use with the seventh grade places him at the grade equivalent of 9.0.

This does not mean his arithmetic achievement is equivalent to an average ninth grade student. It simple means that he can do seventh grade arithmetic as well as the average ninth grade student can do seventh grade arithmetic. The seventh grader would most likely be lost in ninth grade algebra. He is a good seventh grade student who did as well as the average ninth grade student on the arithmetic test.

Intelligence Quotients. The intelligence quotient is a commonly used score type for reporting results of mental ability tests described in Section IV. The older method of dividing mental age by chronological age and multiplying by one hundred is commonly understood by test users. The newer method is the deviation I. Q. which has a mean of 100 with a standard deviation of 15. The I. Q. is subject to misinterpretation and the user must remain aware of the fact that the score is a comparative, estimate of learning ability or capacity, subject to change with any change in performance of the student.

Standard Scores. There are several types of standard scores frequently used. These scores are based on standard deviation units above and below the mean or average score in a normal distribution. The standard scores avoid difficulties of misinterpretation caused by unequal intervals on the percentile scale and can be used for further statistical analysis. Before the standard score can have meaning to the used, the standard deviation and mean score must be known and generally understood. For example, the College Board mean score is 500 with a standard deviation of 100. Stanines, a variant of the standard score gaining popularity in use, ranges over a nine point scale with a mean of five and a standard deviation of two. Each stanine is equal to one-half of a standard deviation unit. Not many teachers will use standard scores to report their own test result; however, it is essential that prospective users of standardized understand this score type. Use of the graph in the Appendix will aid in understanding the relationship of the most commonly used types of standard scores.

Adequate Norms Make Scores Meaningful

A fundamental knowledge of score types and their deviation is essential for proper interpretation and usage. For the score to have meaning just the awareness that the score is a comparasion is not sufficient. Composition of the comparison group or norm group must be clearly defined. Size, geographical distribution, age, sex and

other factors must be carefully specified before true interpretations of the tested population can be made.

National norms or general populational norms are supplied with most tests. Regional norms are sometimes available. Close scrutiny of the representative sample is necessary. Also of concern to the user is the nature of the sample. Small samples of many schools in each sample are more meaningful than norms based on large samples taken from a few schools in each section of the nation, region or particular normative group. Also of concern is when the norms were developed. Many factors may cause the norms to be outdated. To have meaning the individuals relative position in a group must be fixed as definitely as possible. For many types of tests this positioning needs to be made in his own group of associates. This means that local norms will need to be developed in order to achieve the best possible meaning of a test score.

As an example of increased meaning of test scores from different norm groups, Chauncey and Dobbin[6] have listed five kinds of comparative data in their likely order of usefulness and meaning in the interpretation of an individual students performance on an achievement test. If the teacher has these five kinds of comparison data available for American History test, then student A's test performance very well can have five different comparative meanings.

	Class Norms Percentile	School Norms Percentile	System Norms Percentile	State Norms Percentile	Publisher Norms Percentile
Score 142	48	44	56	60	65

Interpretation of these would mean that student A was slightly below the middle of his class on the test, somewhat farther below the average of all classes in his school, but above the average in his state and better than 65% of the students who took the test in the publishers sample. The various interpretations tell the teacher and the student how his performance stacks up in comparison with several kinds of competition.

Scores and Terms are Specific to Given Tests

Terminology used and scores derived from a given test are usually peculiar to that test and cannot be applied to others with

[6] *Chauncey & Dobbin,* p. 49.

accurate meanings . . . an Otis I. Q. and a California Mental Maturity I. Q. do not represent identical measurements and equivalent deductions cannot be drawn from them. In achievement tests, identical grade-level scores on two tests may have different meanings because of the difference between the tests. The same is true in interest and aptitude testing where vastly different scales may have identical titles. There are many tests having the term "mechanical aptitude" in their title and yet they are quite different tests—measure different things. The only way to interpret test scores accurately is to know the test and form used and the norms applied.[7]

Test Scores are Subject to Error

No test is perfect. Any test developed cannot be completely free of imperfections in the instrument. A test, being a sampling, can never measure all of any ability, and to some degree will be inaccurate in what it purports to measure. This margin of error based on reliability of measurement is referred to as Standard Error of Measurement. A particular test may have a standard error of three points. This means that the true test score may be three points above and below what was actually obtained. Some recently published tests are relating the standard error statistic to a band score to aid profile users in answering the question of when a difference in a subtest is actually a significant difference.

Tables* have been prepared for some of the commonly used tests showing the number of points difference between a pair of scores that is a minimum at a prescribed level of confidence. Tables can be prepared from information contained in the test manual of the number of points difference necessary for significance at a certain confidence level.

Presentation of Test Data

There is no universal technique or ready rule for presenting test data to students. A number of graphic devices are available to assist test users in making test data more usable and meaningful to parent and student. Test bands and tables have been referred to in discussing error of measurement. These may aid in presenting data.

---

[7] *A Proposed Twelve Year Testing Program,* p. 20.

* An example of tables referred to above are those prepared by Dr. Frederick B. Davis, Hunter College, N. Y.

A commonly accepted opinion is that parents and students have the right to know what schools learn about the students through the use of tests. This does not necessarily mean the actual score must be released, but does imply an obligation on the part of the school to keep them informed of progress periodically or upon request.

Two principles and one verbal technique for interpretation has been stated in a Test Bulletin by Psychological Corporation which can be used to provide a sound basis for communicating information to parents.[8]

1. Parents have the right to know whatever the school knows about the abilities, performance, and the problems of their children.
2. The school has the obligation to see what it communicates is understandable and usable knowledge.

The verbal technique is just a few words: "You score like people who . . ." or, to a parent, "Your son (or daughter) scores like students who . . ."

Some sample completions listed are:

". . . . people who are pretty good at office work, fast and accurate enough to hold a job and do it well."

". . . . people who don't find selling insurance a very satisfactory choice. Three out of four who score as you do and become insurance salesmen leave the job for something else in less than a year."

". . . . students who find getting into liberal arts college and getting a B.A. degree something they can attain only with extra hard work."

". . . . students who don't often, only about one out of four, manage to earn a C average their freshman year at State."

Many other samples will come readily to mind. The most important thing to note is that a satisfactory report combines two kinds of information:

1. The test results of the individual person
2. Something known about the test or battery and its relation to subsequent performance or others who have taken it.[9]

---

[8] *On Telling Parents About Test Results,* Test Bulletin, The Psychological Corporation, N. Y., p. 1-3.

[9] *Ibid.,* p. 3.

Presentation Depends on Type of Test

Test score information must be presented to the student or parent in a frank manner, and in such a way that it is not placed in improper perspective. There is no standard technique for doing this, but the counselor should see to it that the student does not feel threatened.

The degree of potential threat depends not only upon the rapport established by the interpreter but on the type of test employed. The following guides for interpreting scores for various major test types are included below. The following outline is adapted from a discussion by Merle Ohlsen.[10]

1. Before interpretive session
   a. Review notes made on the test during in-service training session.
   b. Try to determine which sub-scores are valid for what.
   c. Try to determine extent to which test reliability is a factor is use of results.
   d. Review manual for ideas which can be used to:
      (1) diagnose learning problems
      (2) describe test in meaningful way to parent or students.
   e. Write out what he expects to say about each score.
   f. Obtain reactions from someone who knows the test better than he before attempting interpretation.

2. During interpretive session
   a. Describe briefly the type test, sub-tests, and type of information gained from each, e.g., (about a month ago you took an achievement test. Perhaps you remember it had five parts. The first part was called a vocabulary test; compared with others in your grade, your score should give you some idea how well you understand words you read and hear others used.)
   b. Show pupil samples from the various parts of the test as they are described.
   c. Encourage pupil to make self-estimate of how well he did.
   d. Inform student of test scores; first in terms of local norms, then in terms of national norms.
   e. Be prepared to answer questions relative to whether scores are satisfactory.

[10] *NSSE Yearbook*, pp. 278-281.

Counselor Interpreting Test Results

3. After the interpretive session
   a. File rating scale (profile, etc.) in student folder.
   b. Complete and file notes on pertinent non-test data.
      (1) pupils and/or parents reactions
      (2) suggestions for using results
      (3) recommendations for further testing and/or counseling.
   c. Complete and file notes on test data (maybe on rating scale, profile, interpretive blanks, etc.)
      (1) name and forms of test
      (2) pupil's age and grade
      (3) date administered
      (4) date interpreted and to whom

Interpreting Intelligence Test Scores. Educators agree that we should not report specific I. Q.'s to parents. The I. Q.'s derived from different tests vary to an extent that can mislead and confuse even the most alert parents. Intelligence tests differ in the aspects of

mental ability they measure, in the reliability with which these aspects are measured, and in the nature of the norm population used to derive the mental age and/or I. Q.[11] Most educators will agree that the student and/or parent should not be told a specific score, but should know the general level as measured by mental ability test. It is probably wise to interpret I. Q. scores in terms of learning aptitude.

It is relatively easy to inform the above average student that his test scores indicate that he has the learning aptitude to finish college (other factors considered, of course.) The average student can see that with proper study he should have no great difficulty in finishing high school or other types of training. Interpretation of intelligence test score to the below average students presents a more difficult interpretive problem. Sometimes this is handled by telling a student that his scores show that better use of his time and more study will be necessary to reach his goal. Or, "You score like students who have difficulty completing a certain prescribed course."

Interpretation of Interest Inventory. The interest inventory probably is the instrument which presents the least amount of threat. Most interest tests compare the likes and dislikes of students with the likes and dislikes of persons employed in the given interest area. It is essential that the interpreter remain aware and communicate to the student that high interest in an area doesn't mean high aptitude or ability in that area.

The following statement is taken from Merle Ohlsen,[12] "Interpretation of Test Scores" and is cited to illustrate a way of interpreting the interest inventory.

> "These scores will help you discover how much more you seem to like some things than other things. Frequently this kind of study of interests can help you become acquainted with broad areas of interests and job families; it may also help you discover occupations which you could enjoy, it will not help you decide whether you have the ability to do the work. If for any reason my definition doesn't make sense, please feel free to ask questions . . ."

The Inventory is interpreted in above example by having the student estimate his score for ten scales of Kuder Preference Record with counselor going back over them and informing the student where

---

[11] *How to Tell Parents About Standardized Test Results,* Test Service Notebook, No. 26, Harcourt Brace and World, Inc.

[12] *NSSE Yearbook,* p. 291.

the test estimate agrees and disagrees with his estimate. This is followed up with suggestions that the student elect certain preferred occupations for study.

Interpretation of Multi-factor Aptitude Test. The aptitude battery may cause some concern on the part of the student but are still in the category of discussable test information. Principles or discussion of the other test types previously discussed also apply to the multi-factor batteries and or special area aptitude tests. The type interpretation most usually associated with these tests is predictive as their design is to predict future level of performance. The intent of the interpretation is not to tell the student whether he should try to be a musician or engineer, or a doctor; but allow a comparison of his skill in certain areas, with the skill of others. In this way, the student can add to his own knowledge about himself.

A counselor, using percentiles to interpret verbal reasoning aptitude, might say,

> "Your verbal reasoning score is probably higher than 60 of each 100 of the others who took the same test. This area refers to your aptitude to reason with words, written and oral. It relates to practically all school subjects and in practically all occupations, especially those which require communication with words."

Expectancy tables and/or prediction charts have been developed for the most common aptitude batteries. Extensive work has been done in this area with the GATB used by the USES.

Interpretation of Personality Test. Personality inventories are used by some counselors as a part of the systematic testing program. The problem checklist is a commonly used instrument yielding student stated problems which can be discussed with the student. Many personality instruments, however, contain emotionally toned material that requires the skills gained only through study of personality theory and wide clinical experience. The interpretation of such tests or inventories is beyond the province of this bulletin.

SECTION VIII

USING TEST RESULTS

The most important study of education is the study of the student. Without a knowledge about the individual very little effective teaching or counseling can be accomplished. Without complete information about students it would be extremely difficult to reorganize the curriculum or provide the parent with a report of progress.[1] Administering a test is a means to an end, the purpose being to obtain data about the individual which, in turn, may lead to the improvement of instruction and adjustment.[2]

Everyone who is concerned with education wants to know how much students learn in school. Teachers have for years attempted to make judgments, using marks to record their judgments. Making such judgments is baffling business because no completely satisfactory yardstick exists for making them. Wiley and Andrews[3] discuss a number of reasons or characteristics which generally explain why standardized tests have assumed such an important role in education. These are summarized below.

1. Foremost of the desirable characteristics of tests is their impartial, objective nature.
2. The time saving element offered by the use of tests is another desirable characteristic.
3. The results from standardized tests have approximately the same meaning for all trained personnel.
4. Tests provide clues permitting user to gain facts not readily observable.
5. The use of tests increases the probability of sampling all the traits of the individual.
6. School wide testing supplements information gathered by subjective observation in which we may have less confidence.

The versatile nature of standardized tests lead to an unlimited number of uses to which results can be put. The following outline suggests a few of these uses.[4] The list is not all inclusive and many others could be listed.

---

[1] *Willey and Andrews*, p. 125.
[2] *Ibid.*, 147-148.
[3] *Ibid.*, 148.
[4] *A Proposed 12 yr. Testing Program*, p. 23-25.

Testing for Understanding Children

1. Identifying *individual children* needing special attention
   a. Intellectual deviates, above or below norm
   b. Academic deviates, above or below norm
   c. Those whose achievement is inconsistent with their ability
   d. Those whose performance is erratic, inconsistent, or otherwise questionable or curious
   e. Those whose performance suggests the presence of physical difficulties
   f. Children needing referral for special study or other attention.

2. Identifying *groups* needing special attention
   a. Groups tending to be on the average, above, or below norm in achievement or ability
   b. Groups exhibiting special, specific academic, or mental strengths or weaknesses that have educational significance
   c. Groups tending to show academic efforts inconsistent with ability level
   d. Groups whose performance, on the whole, tends to be erratic and inconsistent.

3. Diagnosing all children's characteristics
   a. Understanding the child's specific mental strengths and weaknesses
   b. Understanding the child's specific academic strengths, weaknesses, and needs
   c. Understanding the unique nature of each mental and academic growth patterns
   d. Understanding the many differences in all characteristics which exist within the individual child.

Testing for Use in the Guidance Program

1. To help *teachers*
   a. To identify common problems and those of individual students
   b. To provide for individual differences in ability, interest, attitude, etc.
   c. To motivate students to achieve
   d. To recognize need for referral for special services
   e. To do a better job of understanding and helping students.

2. As an aid to effective school *administration*
   a. To identify areas of student needs
   b. To provide information important to curriculum revision
   c. To provide data for grouping, scheduling, etc.
   d. To provide facts for public relations activities
   e. To provide information basic to total school program evaluation.

3. As a basis for effective cooperation with *non-school agencies*
   a. To identify students in need of referral
   b. To determine eligibility of special service
   c. To provide background for service by non-school resources.

4. As an aid in *counseling*
   a. To help students develop and accept self-concepts
   b. To help students adjust in a social setting
   c. To help students develop long range educational and vocational plans
   d. To help students and their parents make immediate decisions concerning course choices, remedial or special programs, or exploratory activities.

Testing for Use in Curriculum and Instruction

1. *With individual children*
   a. To identify the grade-level at which instruction should be planned
   b. To determine how fast instruction should proceed
   c. To discover or confirm the *specific* kinds of skills which need reteaching, initial teaching, strengthening, etc.
   d. As an aid to selecting instructional materials of proper level, difficulty, etc.
   e. To identify those who need services of a special nature as readiness, remedial, enrichment, slow learning, etc.
   f. To evaluate the degree of severity of problems so time can be spent with pupils most needing it
   g. To assist in confirming the identification of children for whom retention or acceleration should be planned. This does *not* imply that children should be passed or failed by the evidence of test program results.

2. With groups (Points *a* through *d* under "Individual Children" above can be applied to groups as well.)
   a. To provide basis (along with other data) for grouping, either within a class or among classes.
   b. To identify groups of children with common, *specific* problems whose needs can be met with small group techniques
   c. To provide the teacher with diagnostic picture of the academic skills of the class or group as a whole
   d. To identify the *range* and *nature* of *individual* differences which occur in a class so the educational program can be planned to meet these differences in as many ways as possible
   e. To indicate which kinds of problems, if any, are more critical than others so time spent on them can be planned more efficiently.

3. With *building as a whole*
   a. To identify strengths and weaknesses in the total instructional program in terms of specific skills taught or not taught, extent to which the building program meets the needs or deviates, etc.
   b. To provide evidence for greater coordination of material taught from grade to grade; i.e., are there omissions, overlaps, etc.
   c. To identify the number of children needing special attention in the total building so plans can be made for their accommodation
   d. To provide data for effective vertical grouping, groups within groups, etc.
   e. To suggest areas of instruction or child study that might be appropriate for building-level curriculum study.

Appendix-A

A GLOSSARY OF MEASUREMENT TERMS

Ability — The power to perform a designated responsive act. The power may be potential or actual, native or acquired. The term implies that the act can be performed now, without further training, if the necessary circumstances are present.

Achievement Age — The performance level or achievement test score expressed in terms of the chronological age group for which this performance level or achievement test score is average.

Achievement Test — A test that measures the extent to which a person has achieved something—acquired certain information or mastered certain skills, usually as a result of specific instruction.

Age Norms — Values or scores representing typical or average performance for individuals classified according to chronological age, usually expressed as central tendencies, percentiles, standard scores, or stanines. In achievement test, which are standardized by grades, it is the "age equivalent" for grades, with interpolation between given reference points.

Age Grade Tables — A table showing the relationship between the chronological ages of pupils and the school grade in which they are classified.

Alternate-form reliability — The closeness of correspondence, or correlation, between results on alternate (i.e. equivalent or parallel) forms of a test.

Aptitude — The capacity to acquire proficiency with training.

Arithmetic Mean — The sum of a set of scores divided by the number of scores.

Average — A general term applied to measures of central tendency. The three most widey used averages are the arithmetic mean, the median, and the mode.

Articulated Test

A series of tests in which different levels of the test are used for different ages or grades and which have been constructed and standardized so that the same or comparable elements or objects are measured in the overlapping ranges among the various levels of the test. Well articulated tests have considerable overlapping from level to level in order to test the wide ranges of abilities and achievements in any given grade or class.

Battery

A group of several tests standardized on the same population so that results on the several tests are comparable. Sometimes loosely applied to any group of tests administered together, even though not standardized on the same subjects.

Ceiling

The upper limit of ability measured by a test.

Centile

A value on the scoring scale below which are any given percentage of cases. According to some statisticians, the term Centile is often superfluously called PERCENTILE.

Chronological Age Equivalent

The mean or median test score for a specified chronological age group. (See Derived Score)

Class Analysis Chart

A chart, usually prepared in connection with a battery of achievement tests, that shows the relative performance of members of a class on the several parts of the battery.

Class Interval

The divisions of a frequency distribution bounded by upper and lower score values.

Correction for Guessing

A reduction in score for wrong answers, sometimes applied in scoring true-false or multiple-choice questions. Scores to which such corrections have been applied—e.g. rights minus wrongs, or rights minus some fraction of wrongs—are often spoken of as "corrected for guessing."

Coefficient of Correlation (r)

A measure of the degree of relationship, or "going-togetherness," between two sets of measures for the same group of individuals. Unless otherwise specified, "correlation" usually means the product moment Correlation cooefficient, which ranges from .00, denoting complete absence of relationship, to 1.00 denoting perfect correspondence, and may be either positive or negative.

Criterion	A standard, norm, or judgment used as a basis for quantitative and qualitative comparison.
Decile	Any one of the nine percentile points (scores) in a distribution that divide the distribution into ten equal parts every tenth percentile. The first decile is the 10th percentile, the ninth decile the 90th percentile, etc.
Derived Score	A score that has been converted from a qualitative or quantitative mark on one scale into the units of another scale. Examples: Grade Placement Equivalent, Chronological Age Equivalent, Educational Age, Intelligence Quoitient, Mental Age.
Deviation I.Q.	A measure of intelligence based on the extent to which an individuals score deviates from a score that is normal for the individual's age.
Diagnostic Test	A test used to "diagnose," that is, to locate specific areas of weakness or strength, and to determine the nature of weaknesses or deficiencies.
Difficulty Value	The per cent of a specified group who answer a test item correctly.
Discriminating power	The ability of a test item to differentiate between persons possessing much of some trait and those possessing little.
Dual Standardization	The procedure of standardizing two tests simultaneously on one sample, thereby integrating the two instruments.
Educational Age	A pupil's achievement test score expressed in terms of the chronological age group for which his achievement score is average.
Equivalent Form	Any of two or more forms of a test that are closely parallel with respect to the nature of the content and the difficulty of the items included, and that will yield very similar average scores and measures of variability for a given group.

Error of Measurement — Indicates how closely the individual's score compares with his true score. If the standard error is 3.0, the chances are two to one that the score lies within 3.0 points either way of his true score, or nineteen to one that it is not more than 6.0 points from his true score.

Extrapolation — A process of estimating values of a function beyond the range of available data.

Frequency Distribution — A tabulation of scores from high to low (or low to high showing the number of persons who obtain each score or group of scores.

Grade Norm — The average test score obtained by pupils classified at a given grade placement.

Grade Placement Equivalent — A score or a scale developed to indicate the school grade and month in that school grade which is then assigned to the average chronological age, mental age, test score, or other characteristics of pupils classified at this school grade. A grade placement equivalent of 6.4 is interpreted as the fourth month of the sixth grade.

Intelligence — The ability to perceive and understand relationships, such as logical, spatial, verbal, numerical, and recall of associated meanings.

Intelligence Quotient — Originaly, the ratio of a person's mental age to his chronological age (MA/CA) or, more precisely, especially, for older persons, the ratio of mental age to the mental age normal for chronological age (in both cases multiplied by 100 to eliminate the decimal.) More generally, I. Q. is a measure of brightness that takes into account both score on an intelligence test and age.

Interpolation — A process of estimating intermediate values between two known points.

Inventory Test — As applied to achievement tests, a test that attempts to cover rather thoroughly some relatively small unit of specific instruction or training. The purpose of an inventory test, as the name suggests, is more in the nature of a "stock-taking" of an individual's knowledge or skill than an effort to

measure in the usual sense. The term sometimes denotes a type of test used to measure achievement status prior to instruction.

Many personality and interest questionnaires are designated "inventories" since they appraise an individual's status in several personal characteristics, or his level of interest in a variety of types of activities.

Item — A single question or execise in a test.

Item analysis — Any one of several methods used in test construction to determine how well a given test item discriminates among individuals differing in some characteristic. The effectiveness of a test item depends upon three factors: (1) the validity of the item in regards to curriculum content and educational objectives; (2) the discriminating power of the item in regards to validity and internal consistency; (3) the difficulty of the item.

Mean — The sum of a set of scores divided by the number of scores.

Median — The middle score in a distribution; the 50th percentile; the point that divides the group into two equal parts. Half of the group of scores fall below the median and half above it.

Mental Age — The age for which a given score on an intelligence test is average or normal. If a score of 55 on an intelligence test corresponds to a mental age of six years, ten months, then 55 is presumably the average score that would be made by an unselected group of children six years, ten months of age.

Mental Maturity — Intelligence, mental ability. (See Intelligence)

Modal age — That age or age range which is most typical or characteristic of pupils of specified grade placement.

Modal-age Norms — Norms based on the performance of pupils of modal age for their respective grades, which are thus free of the distorting influence of under-age or over-age pupils.

Mode

The score or value that occurs most frequently in a distribution.

Normal Distribution Curve

A derived curve based on the assumption that variations from the mean are by chance. It is bell-shaped in form and adopted as true because of its repeated recurrence in the frequency distributions of sets of measurements of human characteristics in psychology and education. It has many useful mathematical properties. In a normal distribution curve, scores are distributed symmetrically about the mean, as many cases at various equal distances above the mean as below the mean, and with cases concentrated near the average and decreasing in frequency the further one departs from it. (See Appendix E)

Normalized Standard Score

Usually called T-scores, made to conform to standard score values of a normal distribution curve by use of percentile equivalents of the normal curved area; and most frequently expressed with a mean equated to 50 and a standard deviation equated to 10.

Norms

Summarized statistics that describe the test performance of reference groups of pupils of various ages or grades in the standardization group for the test. Grade, age, standard score, and percentile are common types of norms. (See Derived Scores.)

Omnibus Test

A test in which items measuring a variety of mental operations are all combined into a single sequence rather than being grouped together by type of operation. The test has one time limit and yields a single score.

Percentile

One of the 99 point scores that divide a ranked distribution into groups, each of which contains 1/100 of the scores. A percentile rank is a person's rank in a standard group of 100 persons representative of the full range of the normative population. If a person obtains a percentile rank of 70, his standing is regarded as equaling or surpassing 70 per cent of the normative group on which the test was standardized; mean that 30 per cent of the normative group excel this person's test performance. (See Centile, Decile.)

Performance Test

Broadly speaking, every test is a performance test whether the performance is oral responses to questions, written responses to an essay test or an objective test, or the application of manual skills in a test situation. However, pencil-and-paper or oral tests are not usually regarded as performance tests. Performance tests generally require the use and manipulation of physical objects and the application of physical and manual skills in situations not restricted to oral and written responses. Shorthand or typing tests in which the response called for is similar to the behavior about which information is desired exemplify work-sample tests which are a type of performance test.

Personality

The sum total of everything that constitutes a person's mental, emotional, and temperamental makeup. Personality refers to the manner and effectiveness with which the whole individual meets his personal and social problems, and indirectly the manner in which he impresses his fellow.

Power test

A test intended to measure level of performance rather than speed of response; hence one in which there is either no time limit or a very generous one.

Profile

A graphic representation of the results on several tests, for either an individual or a group, when the results have been expressed in some uniform or comparable terms. This method of presentation permits easy identification of areas of strengths or weakness.

Projective Technique

A method of personality study in which the subject responds as he chooses to a series of stimuli such as ink-blots, pictures, unfinished sentences, etc.

Prognosis test

A test used to predict future success or failure in a specific subject or field.

Quartile

One of the three points that divide the cases in a distribution into four equal groups. The lower quartile, or 25th percentile, sets off the lowest fourth of the group, the middle quartile is the same as the 50th percentile, or median and the third quartile, or 75th percentile, marks the highest.

Random Sample — A sample of the members of a population drawn in such a way that every member of the population has an equal chance of being included.

Range — The difference between the lowest and highest scores obtained on a test by some group.

Raw Score — The first quantitative result obtained in scoring a test. Usually the number of right answers, or number right minus some fraction of number wrong, time required for performance, number of errors, or similar direct, unconverted, uninterpreted measure.

Readiness test — A test that measures the extent to which an individual has achieved a degree of maturity or acquired certain skills or information needed for undertaking successfully some new learning activity. Thus a reading readiness test indicates the extent to which a child has reached a developmental stage where he may profitably begin a formal instructional program in reading.

Reference Population — The total population from which a sample is selected for measurement. The term is synonymous with Standardization Population.

Reliability — The extent to which a test is consistent in measuring whatever it does measure; dependability, stability, relative freedom from errors of measurement.

Reliability Coefficient — The coefficient of correlation between two forms of a test, between scores on repeated administrations of the same test, or between halves of a test, properly corrected.

Representative Sample — A sample that corresponds to or matches the population of which it is a sample with respect to characteristics important for the purposes under investigation—e.g., in an achievement test norm sample, proportion of pupils from each state, from various regions, from segregated and non-segregated schools, etc.

Scholastic Aptitude — A combination of native and acquired abilities that is needed to do school work.

Sigma — Designation for STANDARD ERROR and most frequently applied to Standard Deviation.

Skewness — The tendency of a distribution to depart from symmetry or balance around the mean.

Skill — The efficient performance of mental or physical tasks.

Speed Test — A test in which performance is measured by the number of tasks performed in a given time.

Standard — A level of performance agreed upon by experts or established by local school personel as a goal of pupil attainment. Not to be confused with norm.

Standard Deviation — It is a statistic used to express the extent of the deviations from the mean for the distribution. It is obtained by taking the square root of the mean of the squares of the deviations from the mean of a distribution. If the group tested is a normal one their scores, if plotted graphically would yield a normal distribution curve. Approximately two thirds of the scores would lie within the limits of one standard deviation above and one standard deviation below the mean. One-third of the scores would be above the mean by one standard deviation, and one-third below the mean by one standard deviation. About 95% of the scores lie within the limits of two standard deviations above and below the mean. About 99.7% of the cases lie within the limits of three standard deviations above and below the mean.

Standard Error — See Error of Measurement.

Standardization Sample — Refers to that part of the reference population which is selected for use in norming a test. This sample should be representative of the reference population in essential characteristics, such as geographical representation, age, and grade.

Standardized Test — A test that is composed of empirically selected materials, has definite directions for administration, scoring and use, data on reliability and validity, and has adequately determined norms.

Standard Score (Sigma Score, T-Score, Z-Score)	A score expressed as a deviation from the mean in terms of the standard deviation of the distribution (raw score minus the mean, divided by the standard deviation.) (See Normalized Standard Score)
Stanines	A unit that divides the norm population into nine groups.
Stratified Sample	A sample in which cases are selected by the use of certain controls, such as geographical region, community size, grade, age, sex, etc.
Survey Test	A test that measures general achievement in a given subject area. It is used to test skills and abilities of widely varying types. A survey test may also yield diagnostic information. (See Diagnostic Test)
Talent	A relatively high order of aptitude. It refers to an individuals susceptibility to an unusually high degree of training.
Transmutation of Scores	The changing of scores from various tests to a common or equivalent scale so that the scores may be compared or combined.
Validity	The extent to which a test does the job for which it is used. The validity of a test is always specific to the purposes for which the test is used, and that different kinds of evidence are appropriate for appraising the validity of various types of tests.
Variability	The spread or dispersion of scores, usually indicated by quartile deviations, standard deviations, range of 90-10 percentile scores, etc.

Adapted from

"A Glossary of Measurement Terms", California Test Bureau, Monterey, California.

"A Glossary of 100 Measurement Terms", Test Service Notebook, No. 13 Harcourt, Brace and World, Inc.

APPENDIX B

Representative Standardized Tests for Use at Various Grade Levels*

Mental Ability Tests

(Includes Scholastic Ability, Academic Aptitude and/or Intelligence Quotient, I. Q. Tests)

Name of Test	Publishers	GRADE LEVELS						
		7	8	9	10	11	12	C
ACADEMIC APTITUDE TEST	1							
Non-Verbal Intelligence		7	8	9	10	11	12	C
Verbal Intelligence		7	8	9	10	11	12	C
AMERICAN COUNCIL ON EDUCATIONAL PSYCHOLOGICAL EXAMINATION FOR HIGH SCHOOL STUDENTS	8			9	10	11	12	C
AMERICAN SCHOOL INTELLIGENCE TEST	3	7	8	9				
ARMY CLASSIFICATION TEST	18							
First Civilian Edition				9	10	11	12	C
ARMY GROUP EXAMINATION, ALPHA	4	7	8	9	10	11	12	C
ARMY, BETA (revised)	13					11	12	C
CALIFORNIA CAPACITY QUESTIONNAIRE	5	7	8	9	10	11	12	
CALIFORNIA SHORT-FORM TEST OF MENTAL MATURITY	5							
Junior High Level		7	8	9				
Secondary Level					10	11	12	
Advanced Level					10	11	12	C
CALIFORNIA TEST OF MENTAL MATURITY (1957)	5							
Junior High Level		7	8	9				
Secondary Level				9	10	11	12	
Advanced Level					10	11	12	C
CATTELL CULTURE FAIR INTELLIGENCE TEST	3							
Scale II		7	8	9				
Scale III					10	11	12	C
CHICAGO NON-VERBAL EXAMINATION	13	7	8	9	10	11	12	C
COLLEGE QUALIFICATION TEST	13					11	12	C

*This list is not to be thought of as either an all inclusive or a recommended list of tests.

Name of Test	Publishers	GRADE LEVELS						
		7	8	9	10	11	12	C
COOPERATIVE SCHOOL AND COLLEGE ABILITY TESTS (SCAT)	8							
Level I							12	C
Level II				9	10	11	12	
Level III			8	9	10			
Level IV		7	8					
DETROIT GENERAL INTELLIGENCE EXAMINATION	16	7	8	9	10	11	12	C
DETROIT INTELLIGENCE EXAMINATION	16							
Alpha Level		7	8					
Advanced Level				9	10	11	12	C
HENMON-NELSON TESTS OF MENTAL ABILITY (rev.)	9							
Intermediate Level		7	8	9				
High School Level				9	10	11	12	
HENMON-NELSON TESTS OF MENTAL ABILITY (self-marking)	9							
High-School Level		7	8	9	10	11	12	
KENTUCKY CLASSIFICATION BATTERY	11							
Lower Level				9	10	11		
Upper Level						11		
KUHLMAN-ANDERSON INTELLIGENCE TEST	14							
Booklet G		7	8					
KUHLMAN-FINCH INTELLIGENCE TEST	2							
Junior High School		7	8	9				
Senior High School					10	11	12	
LORGE-THORNDIKE INTELLIGENCE TESTS	9							
Level IV								
Level V					10	11	12	
OHIO STATE UNIVERSITY PSYCHOLOGICAL TEST, FORM 21	18			9	10	11	12	
OTIS CLASSIFICATION TEST (rev.)	22	7	8	9				
OTIS GROUP INTELLIGENCE SCALE	22							
Advanced		7	8	9	10	11	12	C
OTIS QUICK-SCORING MENTAL ABILITY TESTS	22							
Beta		7	8	9				
Gamma				9	10	11	12	C

Name of Test	Pub-lishers	GRADE LEVELS						
		7	8	9	10	11	12	C
OTIS SELF-ADMINISTERING TESTS OF MENTAL ABILITY Intermediate Level	22	7	8	9				
Higher Level				9	10	11	12	C
PITNER-GENERAL ABILITY TESTS: Non-Language	22	7	8	9	10	11	12	C
PITNER GENERAL ABILITY TESTS: VERBAL Intermediate Level	22	7	8	9				
Advanced Level				9	10	11	12	C
PRESSEY CLASSIFICATION AND VERIFYING TEST Senior Classification Test	16	7	8	9	10	11	12	
Senior Verifying Test		7	8	9	10	11	12	
PURDUE NON-LANGUAGE TEST	18			9	10	11	12	C
SRA—NON-VERBAL FORM	18			9	10	11	12	C
SRA TESTS OF EDUCATIONAL ABILITY TEA—(6-9)	18	7	8	9				
TEA—(9-12)				9	10	11	12	
SRA TESTS OF GENERAL ABILITY TOGA (6-9)	18	7	8	9				
TOGA (9-12)				9	10	11	12	
SRA—VERBAL FORM	18			9	10	11	12	C
SCHRAMMEL GENERAL ABILITY TESTS	4			9	10	11	12	C
SURVEY OF MENTAL MATURITY: CALIFORNIA SURVEY SERIES Junior High Level	5	7	8	9				
Advanced Level					10	11	12	
TERMAN GROUP TEST OF MENTAL ABILITY	22	7	8	9	10	11	12	
TERMAN-McNEMAR TEST OF MENTAL ABILITY	22	7	8	9	10	11	12	
TEST OF WORD-NUMBER ABILITY	20				10	11	12	C
THURSTONE TEST OF MENTAL ALERTNESS (rev.)	18			9	10	11	12	C

Achievement Test Batteries

Name of Test	Pub-lishers	GRADE LEVELS						
		7	8	9	10	11	12	C
AMERICAN SCHOOL ACHIEVEMENT TEST	3							
Advanced		7	8	9				
CALIFORNIA ACHIEVEMENT TEST (1957 ed.)	5							
Junior High Level		7	8	9				
Advanced Level				9	10	11	12	C
CALIFORNIA BASIC SKILLS TEST	5							
Intermediate		7	8	9				
COOPERATIVE GENERAL AHIEVEMENT TEST	8						12	
ESSENTIAL HIGH SCHOOL CONTENT BATTERY	22			9	10	11	12	
FUNDAMENTAL EVALUATION TEST	20			9	10	11	12	
GRAY-VOTAW-ROGERS GENERAL ACHIEVEMENT TEST	20							
Advanced		7	8	9				
Abbreviated (ed.)		7	8	9				
IOWA EVERY-PUPIL TEST OF BASIC SKILLS	9							
Advanced		7	8	9				
IOWA TEST OF EDUCATIONAL DEVELOPMENT	18			9	10	11	12	
METROPOLITAN ACHIEVEMENT TEST	22							
1960 ed.)								
Advanced		7	8	9				
MINNESOTA HIGH SCHOOL ACHIEVEMENT TEST	3			9	10	11	12	
NATIONAL ACHIEVEMENT TEST	1							
Advanced		7	8	9				
PUPIL SCHOOL ACHIEVEMENT TEST	16	7	8	9				
SRA ACHIEVEMENT SERIES	18	7	8	9				
SRA HIGH SCHOOL PLACEMENT TEST	18							
Series A			8					
SEQUENTIAL TEST OF EDUCATIONAL PROGRESS (STEP)	8							
Level III		7	8	9				
Level II					10	11	12	
STANFORD ACHIEVEMENT TEST	22							
Advanced Battery		7	8	9				
TEST OF GENERAL EDUCATIONAL DEVELOPMENT	23			9	10	11	12	

Multiple Aptitude Test Batteries

Name of Test	Pub-lishers	GRADE LEVELS 7	8	9	10	11	12	C
ACADEMIC PROMISE TEST	13	7	8	9				
DETROIT GENERAL APTITUDE EXAMINATION	3	7	8	9	10	11	12	
THE DIFFERENTIAL APTITUDE TEST	13		8	9	10	11	12	
FLANAGAN APTITUDE CLASSIFICATION TEST	18							
19 test edition				9	10	11	12	
16 test edition					10	11	12	
GUILFORD-ZIMMERMAN APTITUDE SURVEY	19			9	10	11	12	
HOLZINGER-CROWDER UNI-FACTOR TEST	22	7	8	9	10	11	12	
THE MULTIPLE APTITUDE TEST	5	7	8	9	10	11	12	
SRA PRIMARY MENTAL APTITUDE TEST Intermediate Level	19	7	8	9	10	11	12	

SECTION IV.

Interest Inventories or Aptitudes

Name of Test	Pub-lishers	GRADE LEVELS 7	8	9	10	11	12	C
APTITUDE TEST FOR OCCUPATIONS	5			9	10	11	12	
BRAINARD OCCUPATIONAL PREFERENCE INVENTORIES	13		8	9	10	11	12	
GUILFORD-SHNEIDMAN-ZIMMERMAN SURVEY	19			9	10	11	12	C
KUDER PREFERENCE RECORD—OCCUPATIONAL	18			9	10	11	12	C
KUDER PREFERENCE RECORD—VOCATIONAL	18			9	10	11	12	C
OCCUPATIONAL INTEREST INVENTORIES	5							
Intermediate		7	8	9	10	11	12	C
Advanced				9	10	11	12	C
PICTURE INTEREST INVENTORIES	5	7	8	9	10	11	12	C
STRONG VOCATIONAL INTEREST PLANNING	7							
Men				9	10	11	12	
Women				9	10	11	12	
THURSTONE INTEREST SCHEDULE	13			9	10	11	12	
VOCATIONAL INTEREST ANALYSIS	5			9	10	11	12	
YOUR EDUCAIONAL PLANS (YEP)	18							
Junior High School		7	8	9				
High School				9	10	11	12	
PUPIL RECORD AND EDUCATIONAL PROGRESS (PREP)	18	7	8	9				

APPENDIX C

List of Some of the Major Test Publishers and Distributors

1. Acorn Publishing Company
   Rockville Centre, New York
2. American Guidance Service
   2106 Pierce Avenue
   Nashville 12, Tennessee
3. Bobbs-Merrill Co., Inc.
   170 East 38th Street
   Indianapolis 6, Indiana
4. Bureau of Educational Measurements
   Kansas State Teachers College
   Emporia, Kansas
5. California Test Bureau
   916 Williamson Street
   Madison 3, Wisconsin
6. Columbia University
   Teachers College
   Bureau of Publications
   New York 27, New York
7. Consulting Psychologists Press
   577 College Avenue
   Palo Alto, California
8. Cooperative Test Division
   Educational Testing Service
   P. O. Box 589
   Princeton, New Jersey
9. Houghton-Mifflin Company
   3130 Piedmont Road, N. E.
   Atlanta, Georgia
10. Institute for Personality and Ability Testing
    1602-4 Coronado Drive
    Campaign, Illinois
11. Kentucky Cooperative Counseling & Testing Services
    Room 303, Administration Building
    University of Kentucky
    Lexington, Kentucky
12. Psychometric Affiliates
    Box 1625
    Chicago 90, Illinois
13. Psychological Corporation
    304 East 45th Street
    New York 17, New York
14. Personnel Press, Inc.
    188 Nassau Street
    Princeton, N. J.
15. Psychological Research Services
    Western Reserve University
    Cleveland, Ohio
16. Public School Publishing Co.,
    345 Calhoun Street
    Cincinnati, Ohio
17. Scholastic Testing Service Inc.,
    3774 West Devon Avenue
    Chicago 45, Illinois
18. Science Research Associates, Inc.,
    259 East Erie Street
    Chicago 11, Illinois
19. Sheridan Supply Company
    P. O. Box 837
    Beverly Hills, California
20. The Steck Company
    Austin, Texas
21. Western Psychological Services
    10655 Santa Monica Blvd.,
    West Los Angeles 25, California
22. Test Department
    Hargrove-Brace and World
    2126 Prairie Avenue
    Chicago 16, Illinois
23. Veterans Testing Service
    American Council on Education
    1718 Massachusetts Ave., N.E.
    Washington 6, D. C.

APPENDIX D

Test Giver's Self-Inventory

by

Dr. Anton Thompson, Director of Research
Long Beach City Schools, Long Beach, California

Before Tests are Given

Yes	?	No	
——	——	——	1. In scheduling testing periods, do I avoid giving tests immediately after a vacation or on the last day before a vacation?
——	——	——	2. In scheduling tests, do I try to avoid giving tests in a period immediately following a strenuous physical education activity?
——	——	——	3. Have I ever taken the test myself?
——	——	——	4. Before attempting to give the test, do I study the content of the test manual?
——	——	——	5. If I am giving tests for a teacher, do I ask the teacher to serve as a proctor while I administer the tests?
——	——	——	6. Before the tests are given, do I make clear to the room teacher the importance of having the pupils in a state of "readiness" for the testing period? (Absence of hysteria or undue tension; absence also of lethargy or indifference; desk or table tops clear of books, etc.; no disorder; pupil understanding of fact that on standardized tests the pupil must work independently and at moderate speed, and that one expects to find some "hard" items that most pupils can't answer; general idea of separate answer; sheets if they are involved, etc.).
——	——	——	7. Do I understand and accept (and do I help my associates to understand and accept) this statement by Professor Lee J. Cronbach: "A test is necessarily a frustrating experience, since it is designed to include many items which the student will fail . . . . . the task of the tester is to make frustration mild."?

Yes	?	No	
—	—	—	8. Do I refrain from any special coaching on the subject matter of the test?
—	—	—	9. Before I give a test for the first time, do I rehearse to the extent of reading aloud all the directions and noting the timing for each section?
—	—	—	10. Before starting to test, do I make sure that I have an adequate supply of tests and an extra supply of pencils, erasers, and scratch paper if needed?
—	—	—	11. If I give a test requiring the use of a stop watch, do I make any check before testing time to see that the stop watch is working accurately?
—	—	—	12. Before the test begins, do I try to anticipate the questions the pupils will probably ask during the minutes just before they get the signal to begin? Do I have answers ready—answers from the manual, or, if necessary, from careful reasoning? (Example: "Should we skip the hard questions, or should we guess at the answers?")
—	—	—	13. Before starting to test, do I check the seating arrangements to see that the pupils (including the left-handed ones) can handle the test materials comfortably? Are pupils seated in such a way as to encourage self-reliance?
—	—	—	14. Before starting to test, do I check to see that the physical aspects of the testing room are satisfactory (heat, light, ventilation, etc.)?
—	—	—	15. Before the testing period, do I make arrangements so that pupils being tested won't be interrupted by administrative announcements, fire drills, visitors, etc.? Before I begin the test, do I also put a sign on the door (e. g., "TESTING—DO NOT DISTURB") to prevent unnecessary interruptions?
—	—	—	16. Before a testing period in which separate answer sheets are used, determine whether the questions in the booklet are in perfect alignment with the corresponding answer spaces on the answer sheet? Within the limits allowed by the test manual, do I alert pupils to the fact that the answer spaces don't line up exactly with the questions—if they don't?

Yes ? No

—— —— —— 17. Before giving a machine-scorable test, do I check the test booklets to see that they are clean and usable?

—— —— —— 18. If pupils are to take a machine-scorable test for the first time, do I make sure that a practice session covering the manipulation of separate answer sheets and the method of marking answers has been held **before** the day of testing arrives?

—— —— —— 19. Before testing begins, do I check to see whether pupils are properly motivated? If I'm testing older pupils, do they know why the tests are given and do they care about the results?

—— —— —— 20. When testing young children, do I give them a chance to go to the toilet and get a drink before starting the test?

—— —— —— 21. Do I make sure that pupils have removed all extraneous books, desks and tables?

—— —— —— 22. Before I administer a test, do I understand why the test is to be given?

During the Testing

—— —— —— 23. If all pupils are not present when the class period begins, do I delay the start of the test a minute or two in order to reduce the interruption caused by the entrance of the tardy?

—— —— —— 24. Do I try to be matter-of-fact in my manner so that the pupils will understand that the testing experience "is neither a crisis nor a lark"?

—— —— —— 25. When I given the group necessary directions, do I expect them to be followed?

—— —— —— 26. Without being rigid or stilted in manner do I follow the test authors' directions exactly?

—— —— —— 27. Do I refrain from trying to recite test directions from memory?

—— —— —— 28. Do I read the directions slowly, in a clear voice loud enough to be heard in all parts of the room, and do I give proper emphasis to key words and phrases?

Yes ? No

— — — 29. Within the limits allowed by the test manual, do I supplement oral instructions with blackboard or chart illustrations for filling out basic data, etc.?

— — — 30. If a test is to be machine-scored, do I check to see that **every** pupil uses the necessary type of special pencils for **all** parts of the test?

— — — 31. Do the room teacher and I refrain from exchanging semi-audible remarks during the testing? Does the room teacher also refrain from writing on the blackboard?

— — — 32. Do I allow pupils the precise amount of time which the directions specify?

— — — 33. Within the limits of the test directions, do I relieve pupils' tensions over the amount of time available for a long test by posting the **ending** time on the blackboard?

— — — 34. During the pupils' working period, do I move quietly around the room in order to note any individual's unusual behavior, to provide a replacement for a defective pencil, to answer those questions which the directions permit answering, to maintain order, and to encourage a good atmosphere for quiet work? Do I also avoid gazing over any pupil's shoulder so long that he becomes self-conscious?

— — — 35. Even though a test is easily administered and has long working periods for the pupils, do I refrain from the temptation to step out into the corridor (or the cafeteria, or library, etc.) for a minute or two?

— — — 36. Before collecting machine-scorable answer sheets and test booklets, do I ask the pupils to erase any stray marks or words on the materials? Do I ask pupils to blacken the answers well?

— — — 37. When the test is ended, do I collect materials promptly and completely?

After the Testing Period

Yes ? No

___ ___ ___ 38. Before sending a set of answer sheets to the scoring center for machine-scoring, do I take these minimum steps: inspection of papers to see that they are scorable; removal of any paper clips or rubber bands; careful packaging so edges of answer sheets are protected?

___ ___ ___ 39. When an individual pupil's behavior during test strongly suggests that his score will be an invalid measure, do I make a note to that effect on his completed paper and do I arrange for his retesting?

___ ___ ___ 40. Do I arrange for "pick-up" testing for those who miss tests because of absence?

___ ___ ___ 41. If tests are to be hand-scored by someone else, do I give the scorer an adequate explanation as to best procedure?

___ ___ ___ 42. Do I re-score a sample of the tests that others have scored (whether scored by hand or machine) before I proceed to record or use the test data?

___ ___ ___ 43. If there are part scores to be added to arrive at a pupil's total score, do I check the accuracy of my addition (or of the person who has done the adding)? If scoring involves the use of pupils' ages, do I check the accuracy of the age data?

___ ___ ___ 44. If the test results are to be recorded on pupil folders by someone else, do I review with my assistant the essential data that should be recorded? (Usually these are the essentials: title and form of test; date of testing; norm; raw score earned and its equivalent value in terms of the norm.)

___ ___ ___ 45. If the test results are to be used by someone else (e. g., a teacher), do I prepare an easy-to-understand summary, such as a ranking from high to low of the grade equivalents, percentile ranks, or aptitude indexes of all pupils, in a classroom group?

___ ___ ___ 46. Do I use judgment and discrimination when pupils ask: "How did I do on the test?" In such fields as arithmetic computation and language usage, do I help the pupils analyze the results on their scored papers—not in terms of norms but in terms of **correct** responses and **incorrect** responses?

At All Times

Yes ? No

— — — 47. Do I avoid making a statement like the following to any pupil or parent: "Your (Helen's) I. Q. is 97."?

— — — 48. Do I stress with my association the axiom that a standardized test must be given under **standard** conditions if it is to provide reliable information?

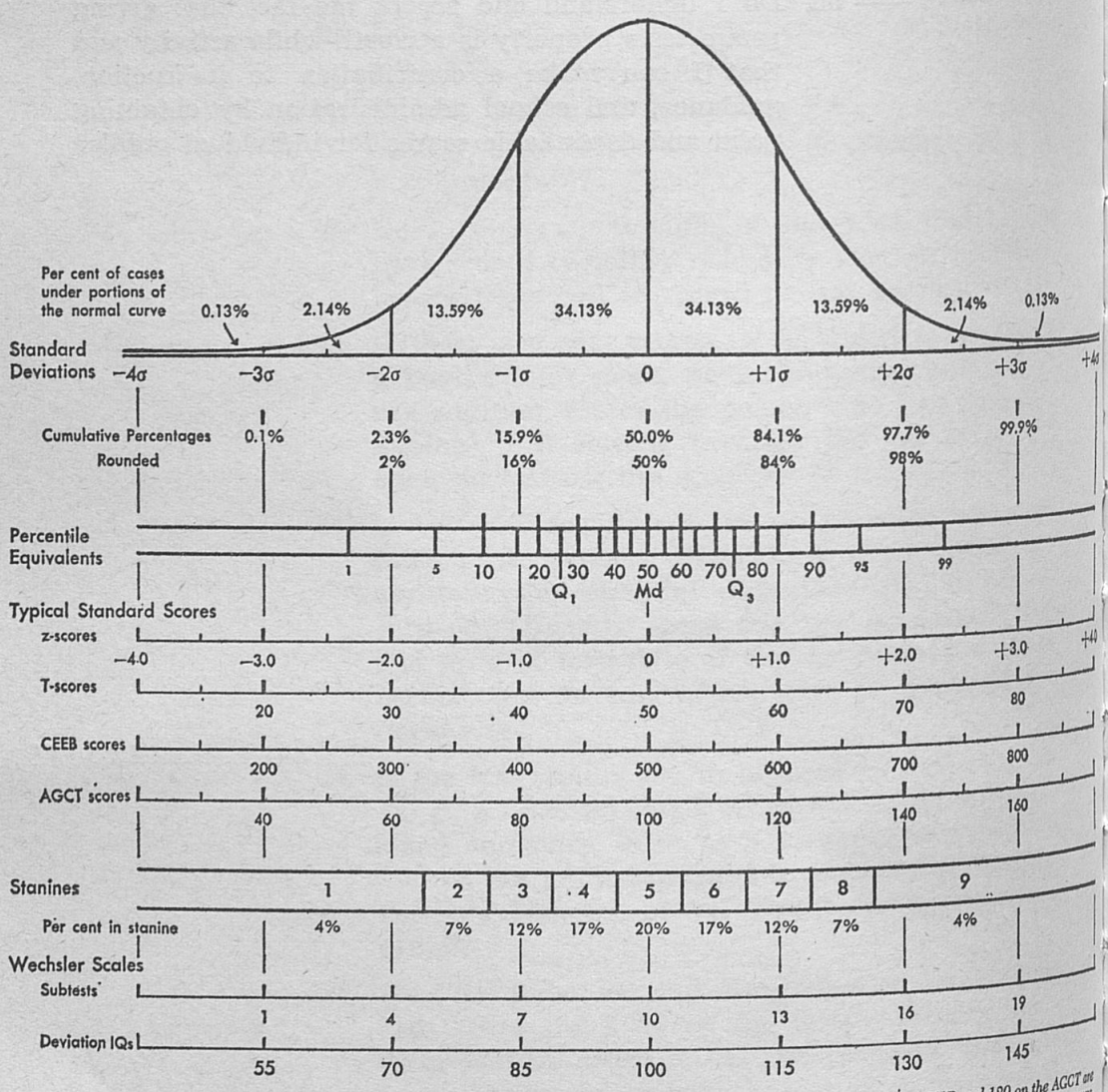
— — — 49. Do I meet periodically with my principal to review my work in the administering of standardized tests?

— — — 50. Do I understand and accept the fact that giving group tests properly is a worth-while activity and that I can make a contribution to instruction, guidance, and school administration by obtaining valid and dependable scores for individual pupils?

APPENDIX E

Methods of Expressing Test Scores

The fundamental equivalence of the most popular standard score systems is illustrated by the following chart. The chart is reprinted from Test Service Bulletin, No. 48, The Psychological Corporation. The discussion of the chart is adapted from the same bulletin. The chart represents the bell shaped normal curve and shows how various types of scores are related to the theoretical normal distribution.



NOTE: This chart cannot be used to equate scores on one test to scores on another test. For example, both 600 on the CEEB and 120 on the AGCT are one standard deviation above their respective means, but they do not represent "equal" standings because the scores were obtained from different groups.

There are no raw scores printed along the baseline. The graph is generalized; it describes an idealized distribution of scores of any group on any test. We are free to use any numerical scale we like. For any particular set of scores, we can be arbitrary and call the average score zero. In technical terms we "equate" the mean raw score to zero. Similarly we can choose any convenient number, say 1.00, to represent the scale distance of one standard deviation. (The mathematical symbol for the standard deviation is the lower case Greek letter sigma σ. These terms are used interchangeably in this article.) Thus, if a distribution of scores on a particular test has a mean of 36 and a standard deviation of 4, the zero point on the baseline of our curve would be equivalent to an original score of 36; one unit to the right, $+ 1\ \sigma$, would be equivalent to 40, (36 + 4); and one unit to the left, $-1\ \sigma$, would be equivalent to 32, (36–4).

The total area under the curve represents the total number of scores in the distribution. Vertical lines have been drawn through the score scale (the baseline) at zero and at 1, 2, 3, & 4 sigma units to the right and left. These lines mark off subareas of the total area under the curve. The numbers printed in the subareas are per cents-percentages of the total number of people. Thus 34.13 per cent of all cases in a normal distribution have scores falling between 0 and –1. For practical purposes we rarely need to deal with standard deviation units below –3 or above +3; the percentage of cases with scores beyond ± 3 is negligible.

The fact that 68.26 per cent fall between ± 1 gives rise to the common statement that in a normal distribution roughly two thirds of all cases lie between plus and minus one sigma. This is a rule of thumb every test user should keep in mind. It is very near to the theoretical value and is a useful approximation.

Below the row of deviations expressed in sigma units is a row of per cents; these show cumulatively the percentage of people which is included to the left of each of the sigma points. The simplest facts to memorize about the normal distribution and the relation of the percentile system to deviations from the average in sigma units are seen in the chart. They are

Deviations from the mean	-2σ	-1σ	0	$+1\sigma$	$+2\sigma$
Percentile equivalent	2	16	50	84	98

We can readily see that ten per cent of the area (people) at the middle of the distribution embraces a smaller distance on the baseline of the curve than ten per cent of the area (people) at the ends of the range of scores, for the simple reason that the curve is much higher at the middle. A person who is at the 95th percentile is farther away from a person at the 85th percentile in units of test score than a person at the 55th percentile is from one at the 45th percentile.

The remainder of the chart, that is the several scoring scales drawn parallel to the baseline, illustrates variations of the deviation score principle. As a class they are called standard scores.

z-scores — These are the same numbers as shown on the baseline of the graph. z-scores are computed by equating the mean to 0.00 and the standard deviation to 1.00.

T-scores — These are directly related to z-scores. The mean is equated to 50, with a standard deviation of 10.

CEEB — The College Entrance Examination Board sets the mean at 500 and the standard deviation at 100.

Stanine — The stanine plan divides the norm population into nine groups with a row of percentages below indicating the per cent of the total population in each of the stanines. Thus 7 per cent of the population will be in stanine 2, and 20 per cent in the middle group, stanine 5.

Reprinted and Adapted from:

"Methods of Expressing Test Scores", Test Service Bulletin, No. 48, The Psychological Corporation, January, 1955.

APPENDIX F

Checklist of Successful Organization and Administration of the Testing Program

A. Coordination and Leadership

	Yes	No
1. A system wide testing program coordinator is appointed		
2. A system wide testing committee has been appointed.		
3. Each elementary school has functional testing committee.		
4. Each secondary school has functional testing committee.		
5. Specific responsibilities in the testing program has been assigned.		
6. In-service training opportunities have been scheduled for those who administer, score, and interpret tests.		
7. In-service training opportunities have been scheduled for staff to assure better and wider use of test results.		

B. Selection of Tests

1. Tests have been well selected with respect to		
a. Reliability		
b. Validity		
c. Adequacy of Norms		
d. Correspondence with the school program		
e. Cost and ease of administration and scoring		
2. There is periodic evaluation by appropriate staff members of each test to determine whether it should continue to be used.		

C. Administration of Tests

1. Persons assigned responsibility for administering tests have been thoroughly prepared.		

2. There is an established maximum pupil-tester ratio. _____

3. Efforts are made to help children understand and accept the importance of the test they are going to take. _____

4. Proctors have been assigned in advance to help with testing before the testing session. _____

D. Testing Facilities

1. Private rooms with minimum noise level is used for group testing. _____

2. Testing rooms are well lighted and ventilated. _____

3. There is adequate, carefully supervised facilities for storage and control of testing materials. _____

4. All testing supplies are subject to continuing inventory and checked as to usability. _____

E. Facilities and Means for Use of Test Results

1. The results of tests are filed in places easily available to the persons who should use them. _____

2. In-service training opportunities have been scheduled for staff to assure better and wider use of test results. _____

3. Private rooms are available, in which staff members can counsel pupils. _____

4. Cumulative folders are carefully reviewed by appropriate staff member, to insure that test data are properly organized and as easily usable as possible. _____

5. Assistance is provided in the interpretation of test results to teachers. _____

Adapted from:

1. George E. Hill, "The Staff Evaluate the Schools' Testing Program." Center for Educational Service, College of Education, Ohio University. Pupil Services Series, 1960, No. 1, pp. 21-23.

2. "Checklist of a Successful Testing Program," Bulletin, Guidance Services in Kentucky's Program of Education, Department of Education, Volume XXXII, April, 1964, No. 4, p. 90.

SELECTED REFERENCES

Anastasi, Anne. *Psychological Testing.* 2nd ed. New York: The Macmillian Co., 1961. 657 p.

Bauernfeind, Robert H. *Building a School Testing Program.* Boston: Houghton Mifflin Co., 1963. 343 p.

Chauncey, Henry, and Dobbin, John E. *Testing: Its Place in Education Today.* 1st. ed. New York: Harper & Row, Publishers, Inc., 1963. 224 p.

Cronbach, Lee J. *Essentials of Psychological Testing.* 2nd ed., New York: Harper & Bros., 1960. 650 p.

Davis, Frederick B. *Educational Measurements and Their Interpretation.* Belmont, California. Wadsworth Publishing Company, 1964.

Durost, Walter N. *The Characteristics, Use and Computation of Stanines.* Test Service Notebook, No. 23. New York: Harcourt, Brace & World, Inc., 1961. 6 p.

How To Tell Parents About Standardized Test Results. Test Service Notebook, No. 26. Tarrytown, N. Y.: Harcourt, Brace & World, Inc., 1961. 4 p.

Findley, Warren G. (ed.) *The Impact and Improvement of School Testing Programs.* 62nd Yearbook, Part II, 5835 Kimbark Avenue, Chicago: National Society for the Study of Education, 1963. 304 p.

Froehlich, Clifford P. and Hoyt, Kenneth B. *Guidance Testing and Other Student Appraisal Procedures for Teachers and Counselors.* 3rd ed. Chicago: Science Research Associates, Inc., 1959. 438 p.

Gannon, F. B. and Telschow, Earl. *Tests and Interpretations* — a teachers handbook. Rochester, New York: City School District, 1960. 31 p.

A Glossary of Measurement Terms, Ninety-six Concepts Which Constitute a Basic Vocabulary in Evaluation and Testing. Del Monte Research Park, Monterey, California: California Test Bureau, 1959. 16 p.

Garrett, Henry E. *Elementary Statistics,* 2nd ed., New York: David McKay Company, 1962.

Goldman, Leo. *Using Tests in Counseling.* New York: Appleton-Century-Crofts, Inc., 1961. 434 p.

Hill, George E. *Evaluating the School's Testing Program.* Center For Educational Service, College of Education, Ohio University, 1959. 28 p.

Kent Area Guidance Council. *A Proposed Twelve Year Testing Program.* Columbus, Ohio: Ohio Scholarship Tests, State Department of Education, March 1959. 57 p.

Lennon, Roger T. et al. *A Glossary of 100 Measurement Terms.* Test Service Notebook, No. 13. Tarrytown, N. Y.: Harcourt, Brace & World, Inc. 6 p.

Lindquist, E. F. (ed.) *Educational Measurement.* Washington, D. C.: American Council on Education, 1951. 820 p.

McLaughlin, Kenneth F. *Interpretation of Test Results.* Washington. U. S. Government Printing Office, 1964. 63 p. (U. S. Office of Education OE-25038).

Understanding Testing, Purposes and Interpretations for Pupil Development. Washington: U. S. Government Printing Office, 1962. 24 p. (U. S. Office of Education, OE-25003.) reprinted 1963.

Noll, Victor H. *Introduction to Educational Measurement.* Boston: Houghton Mifflin Co., 1957. 437 p.

North Central Association Today. Chicago: North Central Association of Colleges and Secondary Schools, August 1961.

Nunnally, Jum C. *Educational Measurement and Evaluation.* New York: McGraw-Hill Book Company, Inc., 1964. 440 p.

Ohlsen, Merle M. *Guidance Services in the Modern School.* New York: Harcourt, Brace & World, Inc., 1964.

Ricks, James H., Jr., *On Telling Parents About Test Results.* Test Service Bulletin No. 54. December, 1959, New York: The Psychological Corporation. 4 p.

Ross, C. C. and Stanley, Julian C. *Measurement in Today's Schools.* 3rd ed. New York: Prentice-Hall, Inc., 1954. 485 p.

Rothney, John W., Danielson, Paul J., and Heimann, Robert A. *Measurement for Guidance.* New York: Harper & Bros., 1959. 378 p.

The Role of Intelligence Testing. Education, XXCI, October, 1960.

The Testing Service. State Department of Education, Hartford, Connecticut, June 1959.

Thorndike, Robert L. and Hagen, Elizabeth. *Measurement and Evaluation in Psychology and Education.* 2nd ed. New York: John Wiley & Sons, Inc., 1961. 602 p.

Traxler, Arthur E. *Techniques of Guidance. Rev. ed.* New York: Harper and Brothers, 1957. 374 p.

Tyler, Leona E. *Tests and Measurements.* Foundations of Modern Psychology Series. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963. 116 p.

Willey, Roy D. and Andrew, Dean C. *Modern Methods and Techniques in Guidance.* New York: Harper and Brothers, Publishers, 1955. 653 p.