1961

Testing and evaluation practices in New England public schools

BOSTON UNIVERSITY

SCHOOL OF EDUCATION

Dissertation

TESTING AND EVALUATION PRACTICES IN NEW ENGLAND PUBLIC SCHOOLS

Submitted by

Winston B. Keck

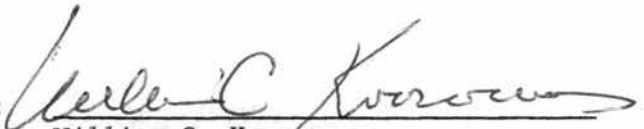
(B.S., Bates College, 1938)

(Ed.M., Boston University, 1950)

In Partial Fulfillment of Requirements for

the Degree of Doctor of Education

1961

First Reader: 

William C. Kvaraceus
Professor of Education

Second Reader: 

James F. Baker
Associate Professor of Education

Third Reader: 

John G. Read
Professor of Education

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

APPENDIX Page

LIST OF TABLES

Table | Page

Table | Page

Table Page

Table Page

Table Page

Table | Page

Table Page

Table Page

CHAPTER I

NATURE OF THE STUDY

1. The Problem

Purpose of the study.-- A comprehensive testing and evaluation program is a necessary and integral part of a sound educational program. It has become an indispensable aid to schools in tailoring a mass educational system to the needs of the individual child and in revealing both the attained and unattained objectives of the schools. Increased emphasis upon individualized instruction, early identification of superior ability, and an expanding special-education program add to the desirability and necessity of securing and utilizing adequate data concerning each pupil. The maximum growth and development of each child is the controlling aim of the schools. Only when the school knows the potentialities and accomplishments of each individual is it in a position to plan for and to achieve this aim. The testing and evaluation program represents the most economical, scientific, and professional means for the school to identify varying differences in individual pupils, measure them, and attempt to gear the educational program to them.

Basically, it is the purpose of this study to contribute specifically to what is known regarding the testing and evaluation practices in the public schools of New England. It thus concerns

itself with the status of the pupil appraisal programs in these schools and determines practices by answering the following questions:

1. How are the programs organized and administered?
2. What tests and other evaluation devices are used?
3. How are the results of the program organized and reported?
4. What use is made of the results of the testing and evaluation program?

The study analyzes and compares practices of school systems of different enrollment categories. Inferences regarding aspects of the program are made and patterns of evolving practices that appear to emerge from the New England scene are noted.

Another purpose of the study is to bring into focus what is being done in testing and evaluation, thus allowing the administrator or supervisor to compare practices in his schools with those in systems of comparable size. The study may also serve as a means of assessing present status and so provide a basis for future evaluation and a starting point for improved testing and evaluation practices in general.

This study does not pretend to appraise the adequacy of practices. Some of these are embodied in the questions posed by Wrightstone:[1/]

"1. Is the design of the evaluation program comprehensive, so that it includes not only abilities, skills, and understandings, but also the less tangible objectives of learning and instruction?

2. Are changes in the behavior of the individual the basis for

---

1/J. Wayne Wrightstone, Joseph Justman, and Irving Robbins, Evaluation in Modern Education, American Book Company, New York, 1956, p. 27.

evaluating his growth and development, since the total behavior of the individual--mental, physical, emotional, and social--should be the concern of the teacher and supervisor in every situation?

3. Are the results of the evaluation organized into a meaningful interpretation so that a portrait of the individual's growth and development and the interrelationships of such growth become evident?

4. Is the evaluation program continuous and interrelated with curriculum?"

Neither is it concerned with measurement practices of individual classroom teachers nor the technical aspects of test construction. It is principally a status study to determine just what is being done and does not purport to evaluate or suggest remedies for present deficiencies.

Definition of terms.-- The terms testing and evaluation are defined and interpreted in a number of different ways. Since they are sometimes used interchangeably, they may be the cause of some confusion. Wrightstone, et al.[1/] refer to evaluation in the following manner:

> "Modern evaluation differs from older forms of appraisal in several ways. First, it attempts to measure a comprehensive range of objectives of the modern school curriculum rather than subject-matter achievement only. Second, it uses a variety of techniques of appraisal, such as achievement, attitudes, personality, and character tests. Included also are rating scales, questionnaires, judgment scales of products, interviews, controlled-observation techniques, sociometric techniques, and anecdotal records. Third, modern evaluation includes integrating and interpreting these various indices of behavior into an inclusive portrait of an individual or an educational situation."

1/Wrightstone, Justman, and Robbins, op. cit., p. 3.

Noll[1] conceives of evaluation as encompassing qualitative as well as quantitative instruments in an attempt to measure and interpret the "whole" child. Monroe[2] also notes that the emphasis in evaluation is upon broad personality changes and major objectives of an educational program rather than upon the more limited aspects of subject matter achievement or specific skills and abilities. Ross and Stanley[3] take this point of view in referring to a comprehensive evaluation program.

The term evaluation, as used in this study, has the broad connotation referred to previously and involves the process of collecting and interpreting data pertaining to the "whole" child as he progresses through school. It is more comprehensive than testing and uses a variety of techniques of appraisal, testing and nontesting devices, in order to obtain an inclusive portrait of the individual pupil. Tests are the tools which are most frequently used and form the groundwork of the evaluation program. They are thus emphasized in this study and are considered as an integral part of the total evaluation program.

Scope of the study.-- This study is concerned with the status of testing and evaluative practices in the New England public schools from kindergarten through high school. By means of an inquiry form,[4] it

---

1/Victor H. Noll, Introduction to Educational Measurement, Houghton Mifflin Company, Boston, 1957, p. 12.

2/Walter S. Monroe, "Educational Measurement in 1920 and in 1945," Journal of Educational Research (January, 1945), 38:334-340.

3/C. C. Ross and Julian C. Stanley, Measurement in Today's Schools, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1954, p. 18.

4/See Appendix C.

analyzes practices within the local school system in the following areas: (1) organizing and administering the program, (2) tests and other measuring devices used in the program, (3) organizing and reporting the results of the program, and (4) administrative and supervisory uses made of test results. These areas are the major concern of the study, since a review of the literature and opinions of such authorities as Noll,[1/] Ross and Stanley,[2/] Traxler,[3/] and Lennon[4/] revealed that these four areas represent the major aspect of any system-wide testing and evaluation program.

Three hundred ninety-five school systems from all the New England states are included in the study. This represents an aggregate enrollment of approximately 1,120,000 pupils. It includes large city schools as well as small systems which are members of a union district or a regional organization.

In addition to determining the testing and evaluation status of the total New England sample, the study analyzes and compares practices of school systems in the following enrollment categories:

1. School systems enrolling under 500 pupils
2. School systems enrolling 500-999 pupils

1/Noll, op. cit., pp. 318-353.

2/Ross and Stanley, op. cit., pp. 209-246.

3/Arthur E. Traxler, et al., Introduction to Testing and the Use of Test Results in Public Schools, Harper and Brothers, New York, 1953, pp. 13-19.

4/Roger T. Lennon, Planning & Testing Program, Test Service Bulletin No. 55, World Book Company, Tarrytown-on-Hudson, New York.

3. School systems enrolling 1,000-2,999 pupils

4. School systems enrolling 3,000-4,999 pupils

5. School systems enrolling 5,000-9,999 pupils

6. School systems enrolling more than 10,000 pupils.

Delimitation of the study.-- The study is primarily concerned with investigating major aspects of the local system-wide testing and evaluation program and is not concerned with evaluative practices carried on by individual teachers in the classroom. No attempt is made to investigate the technical aspects or complexities of test construction or nontesting devices.

Although it is recognized that a comprehensive pupil appraisal program must be concerned with evaluating all aspects of pupil behavior, a review of the literature reveals that a profile of the pupil's behavior in aptitudes, interests, achievement, and adjustment provides most of the evidence needed for an adequate guidance and instructional program.[1/]

This study is thus delimited to the following areas of the appraisal of the pupil:

1. intelligence

2. special aptitudes

3. academic achievement

4. interests

5. personality.

1/Joseph E. King, "Using Tests in the Modern Secondary School," The Bulletin of the National Association of Secondary School Principals (December, 1948), 32:51.

The traditional test approach continues to be the most widely used method for pupil appraisal. Other means of appraisal are considered in this study, but this method receives major emphasis in this investigation. The research is limited also in the following ways:

1. No attempt was made to construct an instrument which would include every possible practice in the testing and evaluation program. Only major items under the four main areas of the program were included.
2. The inquiry form used to collect data was not prepared as an instrument to evaluate the program under investigation. It was considered solely as an instrument to ascertain practices in the areas under investigation.
3. The inquiry form exemplifies the limitation of all inquiry forms in that it depends for its validity on the truthfulness of the person responding.

2. Justification and Need for the Study

Lack of research in the field.-- A review of the literature reveals that no scientific attempts have been made in the past to obtain a clear picture of the testing and evaluation practices in the New England area. A few studies have been made in other areas of the country investigating limited aspects of the program. There is no objective evidence to indicate what the schools in New England are actually doing and to what extent they have incorporated theories and

ideas advanced by measurement authorities. Lindquist,[1] Ross and Stanley,[2] and others note that there is a great need for a critical appraisal of the program as a basis for guiding both theory and practice.

Need for valid data.-- Data obtained from this study will provide valuable information which will aid the administrator in obtaining a realistic picture of practices in his system in relation to other public schools throughout the area. This may thus serve as a means of comparison and as an aid in program justification and upgrading. It will also aid State Departments of Education, colleges serving the area, consultative services, and test publishers in further program planning and development at the local level.

Basis for comparison and future evaluation.-- More emphasis is being placed upon testing and evaluation. Traxler[3] notes that during 1958 over 122 million standardized tests were administered in the United States public schools. The Fifth Mental Measurements Yearbook[4] lists 957 commercially available tests during the 1952-58 period. It also indexes 485 books on measurement and related fields as well as 535 excerpts from book reviews in 81 journals during the same period.

1/E. F. Lindquist (Editor), Educational Measurement, American Council on Education, Washington, D. C., 1951, p. v.

2/Ross and Stanley, op. cit., p. vii.

3/Arthur E. Traxler, "Standardized Tests," National Education Association Journal (November, 1959), 48:18-20.

4/Oscar K. Buros (Editor), The Fifth Mental Measurements Yearbook, Gryphon Press, Highland Park, New Jersey, 1959.

The National Defense Education Act of 1958 and expanding large-scale national testing programs are having far-reaching effects upon programs at the local level. This study may be used as a point of reference in assessing progress since the operation of this Act and point the way for needed improvements in program expansions.

Further research and study.-- The study may serve as a framework for more detailed research in specific areas of the program and for improved educational measurement practices in general. A follow-up study would clearly indicate the impact of the National Defense Education Act on testing and evaluation practices in the schools.

CHAPTER II

REVIEW OF RELATED LITERATURE

1. Introduction

The literature of testing and evaluation is scattered and diffused, encompassing a wide range of books, reports, monographs, bulletins, and periodicals. Buros[1] lists 435 books on measurement and closely related fields and 535 excerpts from book reviews in 81 journals for the period 1952-58. A glance through the two most recent issues of the Review of Educational Research[2], [3] on the topic, "Educational and Psychological Testing," reveals 805 references cited in the 1959 issue and 657 references in the 1956 edition. Wrightstone[4] and his committee noted 187 contributions of research in tests and measurements during the past 25 years. Research in testing and measurement is vast and specialized in such specific topics as aptitude, intelligence, achievement, personality, interests and attitudes, child

---

1/Oscar K. Buros (Editor), The Fifth Mental Measurements Yearbook, Gryphon Press, Highland Park, New Jersey, 1959.

2/American Educational Research Association, "Educational and Psychological Testing," Review of Educational Research (February, 1956), 26:1.

3/American Educational Research Association, "Educational and Psychological Testing," Review of Educational Research (February, 1959), 29:1.

4/J. Wayne Wrightstone, "Educational Measurements," Review of Educational Research (June, 1956), 26:268-292.

study and techniques, and statistical methods related to test construction.

Books and articles dealing with testing and evaluation programs were reviewed. These were valuable sources in providing a background for the study and in developing the inquiry form. The literature, relying heavily upon opinions of competent educators, was primarily concerned with principles, procedures, and scope of the testing and evaluation program. Great emphasis was noted upon the multitude of problems relating to use and interpretation of tests. Most of the writings were concerned with setting forth certain suggestions regarding various aspects of the program. Although these writings and studies may not satisfy the criteria of rigorous research, they do reflect the developments and state of the field.

No studies related in purpose and method to the present study were found. A limited number of surveys were reported, none pertaining to the New England states. These studies were restricted to specific geographical areas and were not concerned with the degree of comprehensiveness of the present study.

This chapter reviews previous studies of practices in programs of evaluation and reports current studies and selected writings which, though they may not represent pure research, nevertheless reflect the best thinking and state in testing and evaluation practices.

2. Current Evaluation Practices

The literature contained several surveys of evaluation programs in action. These studies are considered under the following categories:

(1) system-wide practices involving kindergarten through grade 12, (2) secondary school practices, (3) elementary school practices, (4) measurement practices of teachers, and (5) other testing and evaluation practices. Since the first three categories are most pertinent to this study, major emphasis is directed to these investigations.

System-wide practices.-- Michaelis and Howard[1/] reported on current evaluation practices in 38 unified school districts in California. From data obtained by individual interviews and analysis of guidebooks on evaluation and measurement, they reported that 30 of the districts had a full-time director of the testing program. These persons were given a wide variety of titles. Thirty-two per cent of the districts provided teachers' manuals to explain the program and guide the staff in its operation. A wide variety of topics were noted in these guidebooks. Cumulative pupil record forms were also provided for each pupil in 35 of the districts. Although a variety of evaluative devices were used, the major emphasis was upon tests. Tests of mental ability, reading, and general achievement were used extensively and administered at regular intervals by all districts. Tests of interest, special aptitudes, and personality were used only for special needs or purposes at the secondary level in the majority of districts. Mental tests were administered at grades 1, 2, and 3, or at grades 2, 3, 5, or 6. Achievement tests were administered at grades 4, 5, and 6 or 8.

Of particular interest was the use of various evaluative devices.

1/John W. Michaelis and Charles Howard, "Current Practices in City School Systems in California," Journal of Educational Research (December, 1949), 43:250-260.

Listed in order of decreasing frequency, they were: tests, interviews, case studies, case conferences, group discussions, anecdotal records, observations, files of materials, questionnaires, rating scales, check lists, inventories, logs, diaries, and sociograms. The area of personal-social development evidenced the greatest need for improvement. Very little was being done in this area. Problems of major importance to administrators were: (1) selection of evaluative instrument, (2) interpretation and use of test data, and (3) organization and administration of the program.

Michaelis 1/ also conducted a similar investigation in 68 city school districts throughout the country. Fifty-two of the districts reported that they had directors of evaluation. As in the case of California, many varying titles were given to these positions. Twenty per cent of the districts developed handbooks on testing and evaluation covering a wide variety of topics. A variety of evaluative devices, similar to those reported in the California study, were employed in the program. Individual pupil cumulative records were maintained in 54 districts. Major problems of administrative concern were the same as those reported in the California study.

Traxler 2/ investigated the testing practices in 40 cities of more

---

1/John W. Michaelis, "Current Practices in Evaluation in City School Systems," Educational and Psychological Measurement (Spring, 1950), 9:15-22.

2/Arthur E. Traxler, "The Status of Measurement and Appraisal Programs of Large City School Systems," 1953 Achievement Testing Program in Independent Schools and Supplementary Studies, Educational Records Bureau, Bulletin No. 61, Educational Records Bureau, New York, 1953, pp. 75-86.

than 250,000 population in the United States by means of a questionnaire. He reported that 38 systems had unified testing programs under the supervision of a director. Again, a variety of titles was noted for these positions. All cities reported extensive and widespread use of intelligence and standardized achievement tests. Although the over-all test of intelligence continued to be widely used, a trend toward academic aptitude tests which provide differentiated measures of aptitudes was noted. Interest tests were used to some extent in most cities. Very limited use was made of personality measures. Approximately 53 per cent of the cities administered locally constructed achievement tests as part of the program. A variety of appraisal devices was utilized in all cities. Results of tests were made readily available to counselors in 98 per cent of the cases. Three fourths of the cities indicated wide use of tests for guidance purposes, and three fifths reported wide use for instructional purposes. Test results were used extensively for administrative and supervisory purposes, but very little as a public relations medium. Seventy-five per cent of the cities reported an in-service training program in measurement for teachers. In the majority of cases, the administration of the evaluation program centered in the Bureau of Research.

Testing practices were investigated in 161 public school districts of Connecticut.[1/] One hundred fifty-four districts tested both mental

1/Connecticut State Department of Education, Testing Programs in the Public Elementary and Secondary Schools of Connecticut, Bureau of Research and Statistics, Connecticut State Department of Education, Hartford, 1955.

ability and achievement in the elementary grades. Of the remaining seven districts, four tested achievement only and three limited testing to intelligence. The typical testing program at the elementary grades included a test of mental ability and an achievement test battery--reading, arithmetic, spelling, and language--below grade 4, with social studies and science added for the upper grades. All secondary schools administered mental ability and standardized subject achievement tests. Interest and special aptitude tests were included for guidance purposes in 85 per cent of the schools. Testing patterns, specific tests used, and typical testing programs for towns in various population groups were also investigated in this study.

North[1/] surveyed the testing programs of 48 member public schools of the Educational Records Bureau. Twelve states were represented in this survey. All 48 schools had a systematic testing program in operation. A wide variety of tests was used. Ninety-four per cent of the schools based their testing programs upon stated educational objectives and 90 per cent recorded test results on a uniform pupil cumulative record form. Forty-seven members made testing results available to counselors, 83 per cent interpreted achievement test results to the Board of Education, and 67 per cent made results available to parents and students. In the majority of cases, members indicated that they used special appraisal devices and held in-service training programs

---

1/Robert D. North, "Testing Programs of Public School Members of the Educational Records Bureau--Report of a Questionnaire Survey," 1956 Achievement Testing Program in Independent Schools and Supplementary Studies, Bulletin No. 68, Educational Records Bureau, New York, 1956, pp. 76-90.

to aid the staff members in the use and interpretation of tests. Member schools reported an improvement in the instructional program as a result of testing and a desire for additional consultative services. In 21 systems, teachers helped to score tests. Only four systems developed locally-made achievement tests for school-wide use. A wide variety of titles was mentioned for the persons charged with the responsibility of directing the evaluation programs.

Rosinski[1/] reported a status survey, done by means of a questionnaire, of 49 member schools of the Western New York Study Council. This investigation revealed that 81 different standardized tests were being used in the schools, some schools using only four tests and others using as many as eighteen different tests. Tests generally were administered and scored by teachers. Tests were administered by 70 per cent of the schools during September, October, and/or May. Reading, achievement, and academic aptitude were the major types of tests administered in the schools. Infrequent use was made of other types of tests. Considerable variation exists among schools in the use of certain specific tests. Two concerns published the majority of all tests used in the schools reporting.

The Kent Area Guidance Council[2/] conducted a survey of testing

---

1/Edwin F. Rosinski, "A Status Survey of Existing Testing Programs in Member Schools of the Western New York School Study Council," in The Fourteenth Yearbook of the National Council on Measurements Used in Education, Edith M. Huddleston (Editor), Educational Testing Service, Princeton, New Jersey, 1957, pp. 37-42.

2/Kent Area Study Council, A Proposed Twelve-Year Testing Program, Ohio Scholarship Tests, State Department of Education, Columbus, Ohio, 1959, pp. 31-39.

practices in 43 school systems in Ohio. This study investigated (1) the extent of testing, and (2) evaluation of tests in actual use. Findings revealed that academic aptitude and standardized achievement tests were used regularly in all schools. Academic aptitude tests were given most frequently in grades 1, 3, 6, 9, and 12, and standardized achievement tests in grades 3, 4, 6, and 8. General achievement tests were used more extensively at the elementary level than at the high school level. Special aptitude tests frequently were used in grades 8 and 12, and interest inventories were used primarily in grades 10 and 12. No significant findings were presented relative to the evaluation of tests in actual use, since the data appeared to be inconclusive with respect to the relative merits of particular tests.

Secondary school practices.-- Two significant studies were conducted in the 1930's to investigate testing practices in the secondary schools. Leonard and Tucker[1/] surveyed 870 high schools as part of the 1938 biennial study of the United States Office of Education. Lee[2/] investigated the testing practices of 500 secondary schools and 1,600 secondary school teachers covering 48 states. These studies revealed that intelligence and standardized achievement tests were used by a large majority of the schools. Very limited use was made of aptitude

1/Eugenie A. Leonard and Anthony C. Tucker, The Individual Inventory in Guidance Programs in Secondary Schools, United States Office of Education Vocational Division, Bulletin No. 215, Occupational Information and Guidance Series No. 7, United States Government Printing Office, Washington, D. C., 1941.

2/J. Murray Lee, Testing Programs for Secondary Schools, Teachers College Contributions to Education, Teachers College, Columbia University, New York, 1934.

tests, interest inventories, and personality measures. Fewer than one fourth of the schools had a professionally trained staff member in the field of tests and measurements. Large cities, in many cases, reported that research and guidance departments were responsible for directing the testing program. In almost all other cases the main testing decisions were made by administrative officials.

Surveys were conducted in the high schools of New York City by Alpern,[1] Fox,[2] and Lass and Wrightstone[3] to determine the kind and scope of evaluative practices in the secondary schools of the city. Widespread diversity in practices was noted, ranging from comprehensive programs geared to evaluate the "whole" pupil, to traditional programs hopelessly mired in the past. Varied tests and records were used by individual schools with little city-wide uniformity. The traditional test approach dominated the scene, with newer evaluative techniques neglected for the most part. All investigations revealed that additional consultative services and increased specialized personnel were necessary to provide comprehensive evaluation programs vital to the schools.

Belanger[4] investigated, by means of a questionnaire, the testing

---

1/Hyman Alpern, "Evaluation in the Academic High Schools of New York City," High Points (March, 1946), 28:13-26.

2/Benjamin Fox, "Evaluation in the Vocational High Schools," High Points (March, 1946), 28:26-31.

3/A. N. Lass and J. Wayne Wrightstone, "Evaluation in the Secondary Schools of New York City," High Points (March, 1946), 28:11-13.

4/Laurence I. Belanger, "Testing in California Secondary Schools," California Journal of Secondary Education (February, 1947), 22:108-111.

practices and problems of 187 secondary schools in California. The findings of this investigation revealed that scholastic aptitude, academic achievement, reading, aptitude, interest, and personality were, in the order noted, the areas of appraisal receiving attention. Of particular concern was the fact that more than one half of the schools received no test data from sending elementary schools. Tests were administered about evenly by specialists and teachers. In the majority of cases, tests were scored by teachers, and the results were recorded on individual pupil folders or cards. In decreasing order of frequency, norms and scores receiving most use were: grade equivalent, age equivalent, percentile, and raw scores. Major uses of test data were for sectioning and placement, individual counseling, appraisal of the individual, diagnosis, and individualization of instruction. Major emphasis was directed toward individual appraisal, while some note was made of test usage for checking teaching efficiency and curriculum adequacy. Major administrative problems mentioned were: effective use of test results by teachers, scoring, time factor in recording the results, lack of specially trained personnel, finding tests to parallel the curriculum, maintaining uniform test conditions, and adequate budgetary allowance.

A study of testing practices in 510 California public secondary schools was conducted by the California Association of Secondary School Administrators and the State Department of Education[1/] to determine

1/Carl A. Larson and William H. McCreary, "Testing Programs and Practices in California Public Secondary Schools," California Journal of Secondary Education (November, 1956), 31:389-402.

current programs and the use made of test results. Findings revealed widely diversified practices and comprehensiveness of programs. Extensive testing in the areas of achievement and academic aptitude was reported, with interest, special aptitudes, and personality following in that order. Peak testing activity occurred at the ninth grade level, with the twelfth grade having the second heaviest amount. Interest testing at grades 9 and 12 showed the greatest concentration. Specific tests were also reported.

Of particular concern was the use made of test results. The chief uses were reported as follows:[1/]

> "(1) diagnosing learning difficulties; (2) placing students in appropriate classes or courses, and (3) evaluating instructional programs.
>
> The main uses of academic aptitude tests consisted of: (1) identifying students with either low or high learning ability, (2) placing students in appropriate high school courses, and (3) helping students make educational and vocational plans.
>
> The major uses of aptitude tests other than academic aptitude were listed as: (1) helping students make sounder curricular and vocational choices, and (2) identifying students with special abilities and talents.
>
> Interest inventories were used in most instances for: (1) helping students gain insight into their interests, (2) identifying student interests in broad fields such as scientific and social, and (3) motivating students to see relationships of classroom subjects to various vocational fields.
>
> Personality inventories received their greatest use in: (1) helping to structure interviews with students, and (2) identifying students with possible emotional problems for further study."

Elementary school practices.-- A limited number of studies concerned with testing practices at the elementary level were found. Brown[2/]

1/Larson and McCreary, op. cit., p. 397.

2/Woodrow A. Brown, Testing in Pennsylvania's Public Kindergartens, Test Service Bulletin No. 67, World Book Company, Tarrytown-on-Hudson, New York.

investigated measurement and evaluative practices in kindergartens in 142 public school districts in Pennsylvania. He found that the majority of districts determine the level of attainment of pupils by direct observation and teacher opinion. Formal, as well as informal, testing was used by many schools. Intelligence tests used alone and in combination with reading readiness tests were the chief types of formal measures reported. In almost all cases, testing was done during the month of May.

Shane 1/ investigated evaluative instruments developed during the last several years to determine current interpretations of the term "evaluation" as used in the elementary schools. A study of the evaluative instruments revealed that the term "evaluation" had at least five interpretations, as follows: (1) as a synonym for testing program, (2) as appraisal of instructional competence, (3) as concerned with improving curriculum practice, (4) as the development of a statement of objectives, and (5) as concerned with studying and guiding behavior.

The National Association of Guidance Supervisors and Counselor Trainers 2/ reported wide use was being made of different tests in the elementary schools.

Practices of teachers.-- Several investigations were conducted to

---

1/Harold G. Shane, "Recent Developments in Elementary School Evaluation," Journal of Educational Research (March, 1951), 44:491-500.

2/National Association of Guidance Supervisors and Counselor Trainers, A National Study of Existing and Recommended Practices for Assisting Youth Adjustment in Selected Elementary Schools of the United States, Ann Arbor Publishers, Ann Arbor, Michigan, 1953.

determine measurement practices of teachers. Lee and Segel[1] made a detailed analysis of testing practices of secondary school teachers. Noll and Durost[2] followed up this study in 1955 and reported the kinds of measuring instruments made by high school teachers for their own purposes and the kinds of standardized tests teachers like to use or would use if such were available. Other surveys noted in the literature which were concerned with various aspects of testing practices of teachers were investigations made by Crook[3], [4] and Torbet.[5]

Other testing and evaluation practices.-- The literature also revealed several studies concerned with practices in other aspects of the measurement program. Allen[6] surveyed measurement courses and opinions relative thereto in 288 teacher-training institutions. Stuit,

---

1/J. Murray Lee and David Segel, Testing Practices of High School Teachers, United States Office of Education Bulletin No. 9, United States Government Printing Office, Washington, D. C., 1936.

2/Victor H. Noll and Walter N. Durost, Measurement Practices and Preferences of High School Teachers, Test Service Notebook Number 8, World Book Company, Tarrytown-on-Hudson, New York.

3/Frances E. Crook, "The Classroom Teacher and Standardized Tests," Teachers College Record (December, 1956), 58:159-168.

4/Frances E. Crook, "Elementary School Testing Programs, Problems, and Practices," Teachers College Record (November, 1959), 61:76-85.

5/David P. Torbet, "The Attitude of a Select Group of Colorado Secondary School Teachers Toward Informal Teacher-made Tests as Measured by a Projective Interview," Journal of Educational Research (May, 1957), 50:691-700.

6/Margaret E. Allen, "Status of Measurement Courses for Undergraduates in Teacher-Training Institutions," Thirteenth Yearbook, National Council on Measurements Used in Education, The Council (Robert D. North [Secretary], 21 Audubon Avenue, New York), 1956, pp. 69-73.

Helmstadter, and Frederiksen[1/] investigated the problems of college evaluation. Noll[2/] studied the requirements for measurement courses for certification in various states and the course work offered in measurement in 80 selected teacher-training institutions. Harvey[3/] investigated testing usage and practices in 296 institutions participating in the Graduate Record Examination Institutional Testing Program for 1955-56.

Summary.-- The literature reveals that very few investigations have been made during the past decade to determine current testing and evaluation practices. There is a dearth of data to indicate what the schools are actually doing in this field. From the data reported in this study, it is evident that there is widespread diversity in practices, ranging from the traditional approach mired in the past to the comprehensive program using a wide range of techniques and methods to evaluate the growth and development of pupils.

An evaluation program patterned on the data obtained from surveys reviewed in this study has the features noted below. The majority of surveys reported in this study were concerned with practices in California and the larger school districts. Features presented do not

---

1/Dewey B. Stuit, Gerald C. Helmstadter, and Norman Frederiksen, Survey of College Evaluation Methods and Needs, A Report to the Carnegie Corporation, December, 1956, Educational Testing Service, Princeton, New Jersey, 1957.

2/Victor H. Noll, "Requirements in Educational Measurement for Prospective Teachers," School and Society (September 17, 1955), 82:88-90.

3/Philip R. Harvey, The Use of the Graduate Record Examinations by Colleges and Universities, Educational Testing Service, Princeton, New Jersey, 1957. (Mimeographed.)

thus necessarily represent practices in the typical school district throughout the country.

1. The program is under the direction of a full-time director who has a variety of titles. In the large city system the administration is centered in the Bureau of Research.
2. A wide variety of measuring devices is used in the program. The major emphasis is upon tests.
3. Academic aptitude and standardized achievement tests are used extensively in all schools. Although the over-all intelligence test is widely used, the trend is toward the test which provides differential measurement of aptitudes.
4. Measures of interest, special aptitude, and personality are limited to special needs and purposes.
5. Mental ability is most often tested at grades 1, 3, 6, 8, and 12. Reading measures are administered at grades 1, 3, 6, and 9. General achievement tests are given in grades 3, 5, and 6 or 8. Special aptitude and interest measures are used in grades 9 and 12.
6. Most tests are administered in September, October, and/or May.
7. Little use is made of locally-constructed achievement tests.
8. Limited use is made of the following evaluative devices: interviews, case studies, case conferences, group discussions, anecdotal records, observations, rating scales, and check lists.
9. Test results are recorded on an individual pupil cumulative record.

10. An in-service training program is conducted to aid teachers in using and interpreting tests and measuring devices.
11. Tests are used for improvement of instruction, grouping and sectioning of pupils, diagnostic teaching, individualization of instruction, counseling, and administrative and supervisory purposes.
12. Major administrative problems are selecting evaluative instruments, interpreting and using test data, organizing and administering the program, making effective use of test results, lack of skilled personnel, providing teacher time for scoring, and securing adequate budgetary provisions.
13. Greatest need for improvement is in the area of personal-social development. There is also need for increased use of informal devices and research on improved methods of measurement.

CHAPTER III

ESSENTIAL FEATURES OF A TESTING AND EVALUATION PROGRAM

1. Introduction

Aspects of the testing and evaluation program have been treated by many writers. Source material pertaining to a general overview of the program is presented in recent issues of the Review of Educational Research.[1/, 2/, 3/] Several textbooks on testing and evaluation[4/] treat the over-all program, including aspects of organization and administration. Excellent resources are also to be found in periodicals and journals.

Many phases of the over-all program are the concern of bulletins published by test bureaus and major test publishers. Typical of these are bulletins by Durost,[5/] Clark,[6/] and the Educational Testing Ser-

---

1/Frederick B. Davis, "Testing and the Use of Test Results," Review of Educational Research (February, 1953), 23:5-10.

2/Max D. Engelhart, "Testing and Use of Test Results," Review of Educational Research (February, 1956), 26;5-13.

3/Samuel T. Mayo, "Testing and the Use of Test Results," Review of Educational Research (February, 1959), 29:5-14.

4/See Appendix F for bibliography of textbooks treating aspect of testing and evaluation programs.

5/Walter N. Durost, What Constitutes a Minimal Testing Program for Elementary and Junior High School, Test Service Notebook No. 1, Revised, World Book Company, Tarrytown-on-Hudson, New York, 1956.

6/Willis W. Clark, Articulated and Integrated Measuring Instruments for Practical Evaluation Programs, Educational Bulletin No. 20, California Test Bureau, Los Angeles, California, 1954.

vice.[1] Publications by the State Departments of Education, the United States Office of Education, and professional organizations contribute valuable information on testing and evaluation programs. The School Testing Program[2] and The Fourteenth Yearbook of the National Council on Measurements Used in Education[3] are examples of the latter type of bulletins and reports.

The literature relies heavily upon opinions of competent educators and is primarily concerned with principles, procedures, and scope of testing and evaluation programs. Considerable emphasis is placed upon enumerating and suggesting practices in selected areas of measurement and evaluation. Much of the literature is directed to descriptions of measuring instruments, procedures for performing measurement and evaluation tasks, and the use and interpretation of measurement tools by classroom teachers and school administrators. The literature, generally speaking, does not meet the criteria of research. It does, however, reflect developments and the state of the field. It is thus the concern of this study.

This chapter reviews the literature pertaining to the testing and evaluation program and presents the essential features of this program

---

1/Educational Testing Service, Essential Characteristics of a Testing Program, Educational and Advisory Service Series No. 2, Educational Testing Service, Princeton, New Jersey, 1956.

2/New York State Education Department, The School Testing Program, University of the State of New York Bulletin No. 1415, State Education Department, Albany, New York, 1953.

3/Edith M. Huddleston (Editor), The Fourteenth Yearbook of the National Council on Measurements Used in Education, Educational Testing Service, Princeton, New Jersey, 1957.

as reflected in current writings. Technical considerations involved in various phases of testing, i.e., inter-item correlation, test scaling, test validity, etc. are not reported. Statistical and measurement theories are not treated. Selected writings are reviewed, synthesized, and treated under the following aspects of the pupil-centered testing and evaluation program:

1. Basic considerations
2. Organizing and administering the program
3. Appraising various aspects of student behavior
4. Organizing and recording test results
5. Utilizing test results.

The literature reveals that a comprehensive appraisal program must consider all aspects of the pupils' growth and development. Physical, biological, and biographical data must be included in this total picture. A review of the literature concerning all aspects of the pupil's behavior and learning pattern is beyond the scope of this study. Major emphasis in this review is thus directed to the following areas of the pupil-centered testing and evaluation program:[1/] (1) intelligence or academic aptitude, (2) special aptitude, (3) achievement, (4) interest, and (5) social-emotional development.

Many opinions and ideas of writers concerning various aspects of the testing and evaluation program are noted throughout the literature. Statements of opinions and ideas in this review represent views ex-

---

1/See Delimitation of the study, supra, p. 6.

pressed by these writers which reflect developments and the state of the field of testing and evaluation.

2. Basic Considerations

A concept of testing and evaluation.-- Education today is regarded as a process of growth and development of the "whole" child, and not merely an act of acquiring specific skills and factual information.[1/] The modern school is not only concerned with subject matter achievement, but also with the development of attitudes, interests, ways of thinking, personal and social development, work habits, and like processes.

This broader concept of the function of the school has made its imprint on the literature pertaining to testing and evaluation programs. As the school's objectives have been clarified and defined in instructional practices, appropriate methods of assessment--formal and informal--have been devised to gauge the adequacy of the school's program. The emphasis has thus changed from that of testing single aspects of subject matter achievement to that of gathering all kinds of evidence which indicate the degree of attainment of important outcomes of the educational process.[2/] The literature indicates that tests still greatly exceed in number and variety all other types of evaluative instruments and play a most important and vital role in the

---

1/Pedro T. Orata, "Evaluating Evaluation," Journal of Educational Research (May, 1940), 33:641-661.

2/J. Wayne Wrightstone, "Evaluation," in Encyclopedia of Educational Research, The Macmillan Company, New York, 1950, p. 404.

evaluative process. They represent the best means of measuring pupil attainment in subject-matter areas and tangible aspects of educative growth; but they no longer alone meet the needs for appraising the total aims of a comprehensive educational program.

In recent years the thinking has changed from the narrow aspect of testing to the broad philosophy of evaluation. Monroe[1] indicates that this has had a far-reaching influence on all phases of the measurement movement and represents a significant development in education. In function evaluation involves the following:[2]

> ". . . (1) the identification and formulation of a comprehensive range of major objectives for a curriculum, (2) their definition in terms of pupil behavior to be realized, and (3) the selection or construction of valid, reliable, and practical instruments for appraising major objectives of the educative process or characteristics of personal growth and development."

Wrightstone, Justman, and Robbins[3] further define evaluation in the following terms:

> "Broadly defined, educational evaluation is the estimation of the growth and progress of pupils toward objectives or values in the curriculum. The purposes of evaluation are to provide for the collection of evidence which will show the degree to which pupils are progressing toward curricular goals, and to permit teachers and supervisors to evaluate the effectiveness of curricular experiences, activities, and instructional methods. The functions of evaluation are to make provisions for guiding the growth of individual pupils, to diagnose their weaknesses and strengths, to point out areas

---

1/Walter S. Monroe, "Educational Measurement in 1920 and in 1945," Journal of Educational Research (January, 1945), 38:334-340.

2/J. Wayne Wrightstone, Joseph Justman, and Irving Robbins, Evaluation in Modern Education, American Book Company, New York, 1956, p. 4.

3/Ibid., p. 16.

where remedial measures may be desirable, and to provide a basis for the modification of the curriculum or for the introduction of experiences to meet the needs of individuals or groups of pupils."

Segel and Gerberich[1] report certain premises which are widely accepted and are consistent with the point of view expressed in the literature:

> "Certain premises seem to be widely accepted today. One is that the modern teacher must be concerned with the basically important and functional learning outcomes, usually far less tangible and less easily measurable than knowledges and skills, which were largely disregarded some years ago both by teacher and tester. Another is that the whole child must be evaluated in the dynamic social situation in which he inevitably finds himself. The demand, therefore, seems to be for an approach to evaluation which is founded upon a sound understanding of child growth and development, the nature of learning, and individual group, and trait differences. The measurement specialist may well be faced with the necessity of meeting this type of demand. It seems likely to entail a broader equipment of testing techniques and a greater use of the principles of expanding social and educational psychology than were demanded during the first fifty years of development in educational and psychological testing."

The development of orderly procedures and techniques for evaluating the "total" pupil involves many basic considerations. These are stated succinctly by Burton[2] in discussing the source of the factors which complicate the problem:

> "1. The outcomes to be evaluated are complex integrations of many items.

1/David Segel and Raymond J. Gerberich, "Overview of Educational and Psychological Testing 1946 to 1943," Review of Educational Research (February, 1950), 20:13.

2/William H. Burton, "Implications for Organization of Instruction and Instructional Organization," Learning and Instruction, Part I, Forty-ninth Yearbook of the National Society for the Study of Education, University of Chicago Press, Chicago, 1950, pp. 247-248.

2. The outcomes to be evaluated are difficult to state in unequivocal language.

3. The scope of evaluation has been greatly expanded. Early emphasis upon testing fact and skill outcomes have been supplemented by attention to evaluating such outcomes as understandings, attitudes, appreciations, values, and other personal-social-moral traits which control conduct.

4. Directional progress goals must be related eventually to relatively fixed standards. Standards of achievement and growth must of necessity be adapted to levels of maturity and to individual differences; hence there must be flexibility in the appraisal of pupil growth. The learner is to progress along lines indicated by the aims and objectives in the general situation. The standards move constantly upward with the maturing learner. Eventually current social standards enter as mature levels of the goals which pupils have been progressively achieving. Efforts to evaluate in terms of arbitrary standards, the growth of immature children with differences in abilities, interests, and rates of growth result in serious disturbances of mental hygiene. Children allowed to progress through a series of directional progress goals and standards will come to see quite naturally the necessity for achieving socially desirable levels. Evaluation must be carried on with these facts in mind.

5. The outcomes to be evaluated are often stated in general and abstract terms, but pupil behavior is specific and concrete. This calls for the gathering of specific data from observations of behavior and then making a judgment, i.e., evaluating the behavior as indicative or not indicative of growth in given values.

6. Modern evaluation within the everyday on-going classroom situation is continuous and adjusted to the nature of the learner. The close of a unit or series of assignments will include an appraisal of the degree to which the objectives and purposes of the unit or assignments have been achieved.

7. Modern evaluation is participatory; everyone from test technician to the pupils themselves takes part. It is of basic importance to have pupils learn to evaluate objectively their own achievement. Evaluations should be included in which pupils do not participate, but for

which, on the basis of their experience, they can clearly see the why and the how.

8. The techniques of modern evaluation are such that the teachers should be given definite training in the development and use of the processes. Children should be given continuous opportunity to evaluate materials, classroom procedures, their own achievements and growth.

9. Objective measurement and subjective evaluation are both essential. The current scene is marked by considerable development in the area of subjective evaluation. New instruments and procedures are constantly being developed."

Writers agree that an adequate program of testing and evaluation must, in design, be comprehensive and include tangible as well as intangible objectives of instruction and learning. Changes in the behavior of the individual must be the basis for evaluating his growth and development. The program must be continuous and well articulated from kindergarten through high school. The program must be functional and practical, and the results must be organized into a meaningful interpretation of the individual's growth and development.

Wrightstone[1/] sets forth six criteria that may well serve as a guide in developing a concept of testing and evaluation:

"The 1st criteria is that the major purposes of testing and evaluation should be to appraise the outcomes of educational experience and that the design of such evaluation should be comprehensive. It should not be limited to a few objectives or outcomes, but should include the major objectives of the modern school curriculum.

The 2nd criteria is that the outcomes of learning experiences are the changes which take place in the behavior of students. This behavior should be interpreted to mean

1/J. Wayne Wrightstone, "Design for Evaluation for High Schools," High Points (February, 1944), 26:32.

not only observable skills but also interest, attitude, ways of thinking, and the like.

The 3rd criteria is that for some of the major objectives of instruction no adequate methods or instruments for collecting reliable evidence are available. Until valid and reliable technics are evolved, such outcomes should be appraised by as careful subjective means as possible.

The 4th criteria is that reliable and valid instruments of measure are by their very nature restricted to an appraisal of limited aspects of behavior. It is impossible to measure the whole result of the educative experience by any one test or battery of tests. The hope remains, however, that by measuring the major, important, and vital objectives of educative experience some valid appraisals can be obtained of the relative merits of diverse educational practice.

The 5th criteria is that the test results should be summarized into a meaningful pattern of scores either statistical, graphic, or verbal, and that these results should be organized and integrated into helpful patterns of interpretations. In this interpretation an effort should be made to see the relation between and among the indexes and scores obtained and to interpret the results in terms of probable contributing causes.

The 6th criteria is that a dynamic evaluation program is one that is continuous and interrelated with the curriculum. The curriculum objectives determine the outcomes which should be evaluated. The test, questionnaire, and other instruments by means of which evidence is gathered provide evidence of development towards these objectives. The evidence gathered by techniques of appraisal, in turn, affects, first, by indicating those desirable, and second, by throwing light on the meaning and clarification of objectives and outcomes."

Place of testing and evaluation in the total education program.--The literature states that evaluation is a major activity of the school and represents an important phase of good teaching. Education in terms of individual needs must be based upon a knowledge of the "whole" child. To appraise pupil progress and meet individual needs, adequate data must be secured concerning each child. A comprehensive testing and

evaluation program is needed to accomplish this. Cromwell[1/] notes the necessity of such a program if the school is to secure an adequate assessment of its product.

> "A study of any group reveals differences among the individuals composing it. These differences may be in terms of intelligence, command of fundamental tools of learning, physical and emotional maturity, social abilities, specific aptitudes, interests and other characteristics. Furthermore, not only are there infinite variations among individuals, but each differs from himself from time to time. So vitally do these variations affect mental and emotional growth, that their existence must never be ignored by anyone who works with youth.
>
> To understand others and to work effectively with them, one must understand the nature and extent of their differences. It is not enough for the school to know that a boy is a poor reader; how great his deficiency is and what specific difficulties he encounters must be understood if real help is to be given. This information should be a part of the school record, available to all the school officials and teachers who have contact with the pupil concerned. Mere access to records will not suffice, however. Each teacher should grow in his ability to interpret and use them; and the best measure of his success in doing so will be his ability and willingness to interpret unpromising data in an objective manner.
>
> An analysis of individual differences should lead to the discovery of individual needs, and the extent to which a school discovers and provides for individual needs determines the adequacy of its program."

Testing and evaluation are integral parts of the total educational program and function in facilitating learning in the following areas:[2/] over-all educational planning, educational placement, guidance and

---

1/Floyd R. Cromwell, A Basic Program of Guidance, Maryland School Bulletin, Vol. XXV, No. 1, State Department of Education, Baltimore, Maryland, 1943.

2/Walter W. Cook, "The Functions of Measurement in the Facilitation of Learning," Chapter I in Educational Measurement, E. F. Lindquist (Editor), American Council on Education, Washington, D. C., 1951.

counseling, improvement of instruction, and improvement of the learning situation. Segel[1] notes that testing and evaluation should assume a position of administrative importance for the following reasons:

> "1. To enable teachers, administrators, and counselors to keep themselves more intimately and reliably acquainted with the continuing educational development of each individual pupil, in order that the instruction and guidance may be better adapted to his peculiar and changing interests, needs, and abilities.
>
> 2. To provide the school administrator with a more dependable and objective basis for the over-all evaluation of the total education offering of the school, in order that any need for curricular revision may more surely be brought to his attention, and that his supervisory efforts may be more wisely distributed."

Writings cite that the problems of evaluation and the problems of instruction are intricately interwoven. What is to be evaluated and the methods of evaluation depend on what the schools teach or should teach. Deriving its direction from the major objectives of the schools, the testing and evaluation program clarifies these objectives into goals and purposes which are effective guides to instruction. The controlling purpose of the appraisal program is to facilitate learning experiences aimed to meet the needs of the pupil. To accomplish this aim, the program must meet the following conditions:[2]

> "1. The objectives set up to guide instruction should specify unambiguously and realistically the behaviors

---

1/David Segel, "Survey and Trend Studies," Review of Educational Research (December, 1942), 22:494.

2/American Association of School Administrators, American School Curriculum, Thirty-first Yearbook, 1953, Washington, D. C., p. 313.

for the development of which the schools accept responsibility. These are the ultimate objectives.

2. The selection, organization, and sequential arrangement of learning experiences (for example: immediate objectives, units of instruction, and course content) should be clearly relevant to the attainment of these ultimate objectives. Appraisals, in turn, should be relevant to the immediate objectives so determined.

3. Evaluation, as an integral part of the instructional program, should be as comprehensive as that program. It must embrace all important behaviors and must show periodically the degree to which these behaviors have been achieved. It will utilize all technics which perform important services.

4. The data obtained from evaluation should be fully utilized for the improvement of instruction.

5. The staff responsible for instruction should have the primary responsibility for the program of evaluation; that is, for determining what should be tested, what forms of evaluation are appropriate, and how the results are to be interpreted and used. It should, however, include persons with special competence in the technics of evaluation.

6. Classroom teachers should participate in the program of evaluation to the fullest practicable extent, not only because of the contributions they can make to it but also because of the beneficial effect on their teaching.

7. Pupils should share in the appraisal of their achievements."

Lee,[1/] in discussing the elementary school program, notes the place of tests in the instructional program:

"No elementary school can have an effective instructional program without knowing what it is accomplishing. Tests are a means of finding out what is happening to pupils in order that further progress may be made. The emphasis on

1/J. Murray Lee, "Essential Elements of an Adequate Testing Program," in "Appraising the Elementary School Program," Sixteenth Yearbook, The National Elementary Principal (July, 1937), 16:465-470.

> the individual requires not only a general knowledge of what is happening to the group, but also specific knowledge for each individual."

The foregoing clearly indicates that testing and evaluation are vital and integral elements of a comprehensive educational program and are essential parts of the process of teaching, curriculum making, and guidance.

Kinds of information needed about pupils.-- Writings reveal that a basic function of the school is to provide for the common and individual needs of the pupil in terms of his abilities and interests in relation to the environment. If this function is to be carried out on any rational basis, it is necessary for the schools to obtain data on all factors which influence the pupil's behavior and learning. Schwartz and Tiedman[1/] classify these factors in the following categories: (1) physical, (2) health, (3) psychological, (4) educational, (5) and emotional. These authors state that a true picture of the pupil requires evidence concerning each of these factors and the influence each has upon the pupil's progress. They further list specific factors which affect the pupil's progress and development under each category in the following summary:

> "I. Physical factors
> A. Speech
> B. Hearing
> C. Vision
> D. General physical (bodily) condition
> 1. Motor coordination
> 2. Vigor

1/Alfred Schwartz and Stuart C. Tiedman, Evaluating Student Progress in the Secondary School, Longmans, Green and Company, New York, 1957, pp. 19-20.

3. Vitality
4. Growth

II. Health Factors
- A. Eating habits
- B. Sleeping habits
- C. Personal hygiene
- D. Complexion

III. Psychological factors
- A. Personality adjustment
  1. Social competence
  2. Aggressive or submissive tendencies
  3. Attitudes
  4. Group (peer) relationships
- B. Aptitudes
  1. Scholastic
  2. Specific
- C. Interests
- D. Reading skill
- E. Mental health

IV. Educational factors
- A. Past achievement--areas of superiority and weakness
- B. Suitability of curriculum
- C. Cocurricular participation
  1. Special successes or failures
  2. Areas of activity chosen
- D. Suitability of instructional materials and methods
- E. Study and work habits
- F. Teacher-student relationship
- G. Student-school relationship

V. Environmental factors
- A. Family background
  1. Cultural and social characteristics and influences
  2. Attitudes toward child, school, society
  3. Economic and social status
  4. Marital status of parents
  5. Home duties and responsibilities
- B. Neighborhood factors
  1. Geographical location and characteristics
  2. Companions
  3. Recreation facilities
  4. Employment (full- and/or part-time) opportunities."

The importance of obtaining evidence of understandings is discussed by Douglas and Spitzer.[1/] It is noted that although understandings are

---

1/Harl R. Douglass and Herbert F. Spitzer, "The Importance of Teaching for Understanding," Chapter II in The Measurement of Understanding, Part I, The Forty-fifth Yearbook of the National Society for the Study of Education, University of Chicago Press, Chicago, 1946, p. 25.

basic to rich and effective living, they are often neglected in favor of other learning outcomes such as verbalism, barren factual information, and mechanical skills. In analyzing curriculum objectives, the Eight-Year Study[1/] set forth the following specific aspects of pupil growth that should be measured: thinking, social sensitivity, significant interests, civic and social beliefs, appreciation in literature and art, and personal and social development.

Remmer and Gage[2/] indicate that the mutual needs of both society and the individual should determine the kinds of information and data with which pupil evaluation must be concerned. They further suggest that the following aspects of the pupil need to be evaluated: (1) achievement of instructional objectives, (2) physical aspects, (3) mental abilities, (4) emotional and social adjustment, (5) attitudes, (6) environment and background. Traxler[3/] states that there are ten areas of the history and development of the individual within which information is needed for guidance and instructional purposes. They are as follows: (1) home background, (2) school history and home adjustment, (3) mental ability or academic aptitude, (4) achievement and growth in different fields of study, (5) health of the individual, (6) out-of-school experience, (7) educational and vocational interests of

---

1/Eugene R. Smith, Ralph W. Tyler, and others, Appraising and Recording Student Progress, Harper and Brothers, New York, 1942, p. 18.

2/H. H. Remmer and N. L. Gage, Educational Measurement and Evaluation, Harper and Brothers, New York, 1943, pp. 19-117.

3/Arthur E. Traxler, Techniques of Guidance, Harper and Brothers, New York, 1957, pp. 20-25.

individual pupils, (8) special aptitudes, (9) personality, and (10) plans for the future. Dunlap[1/] states that if the school is to carry on a well-founded educational program, it must secure the following information about each pupil: (1) measures of achievement, (2) estimates of intellectual ability, (3) a profile of the individual's interests, (4) estimates of artistic capacity, (5) a measure of socio-economic status, (6) ratings of adjustments to home, (7) ratings of health, (8) ratings of adjustments to social environment, (9) measures of personality traits, (10) some measure of delinquency, and (11) measures of attitude toward society and the state.

Thorpe[2/] states that measures of achievements, potentialities, and state of personal and social adjustment are a prerequisite to effective instruction. To plan a constructive program and forestall failures with their attendant nervous symptoms or delinquent behavior, he notes that the teacher must secure objective measurements of the status of each pupil regarding: (1) physical health, (2) intelligence or mental maturity, (3) competence in basic academic skills, (4) general and occupational interests, (5) personality adjustment, (6) mental health, (7) special abilities and aptitudes, and (8) social relationships with other pupils.

Considerable duplication and overlapping are noted in the liter-

---

1/Jack W. Dunlap, To Test or Not to Test, Test Service Bulletin, No. 37, World Book Company, Tarrytown-on-Hudson, New York, p. 3.

2/Louis P. Thorpe, Guiding Child and Adolescent Development in the Modern School, Educational Bulletin No. 16, California Test Bureau, Los Angeles, California, 1951, p. 7.

ature. It is evident, however, that the school should be concerned with gathering data on all aspects of pupil behavior and achievement. It is not enough to be concerned simply with academic achievement. To have a developing picture of the individual pupil and be cognizant of the factors which affect the pupil's school adjustment and progress, all areas must be tapped for information. If decisions are to be made regarding the educational program and the pupil's pattern of development, a continuous appraisal and recording of the progress of each pupil in each important area of his growth over a long period of time are necessary.

Means and techniques used in the program.-- The field of evaluation includes testing, both educational and mental. It also includes the much wider area of pupil behavior study by observation and rating in every conceivable situation. A broad testing and evaluation program requires a great variety of measures or samples of an individual's behavior used interrelatedly in the process of understanding and helping the individual. Two basic principles underlie evaluative measures and techniques:[1] (1) the instrument or technique should be chosen to fit the objective to be measured or evaluated; and (2) no technique is worth using unless the results it yields can be depended upon.

Wrightstone, Justman, and Robbins[2] point to the necessity of using a variety of techniques in the following comments:

1/Victor H. Noll, Introduction to Educational Measurement, Houghton Mifflin Company, Boston, 1957, p. 13.

2/Wrightstone, Justman, and Robbins, op. cit., p. 16.

"Techniques of evaluation range from such informal measures as teacher ratings on oral recitations and teacher-made tests to more refined and standardized measures or aptitudes, abilities, skills, interests, and attitudes. Formerly, the major emphasis in measurement was on appraisal of pupil mastery of information and the skills. The changing concepts of the curriculum have required the evaluation of pupil growth in other areas as well, such as physical and mental health, social relationships, critical thinking, appreciations and creative expression, interests, and attitudes. In other words, new curricular emphases have required the development of new techniques of measurement and evaluation."

Schwartz and Tiedman 1/ give the following list of available techniques for measuring various aspects of the pupil's behavior:

"I. Tests
  A. Achievement
    1. Informal teacher-made
    2. Standardized
  B. Mental Ability
  C. Personality
  D. Aptitude
  E. Interest
II. Rating scales
III. Checklists, surveys, inventories, and questionnaires
IV. Observation
V. Interviews
VI. Records and reports
  A. Cumulative folders
  B. Anecdotal records
  C. Diaries and logs
VII. Sociometry
VIII. Role-playing
  A. Sociodrama
  B. Psychodrama
IX. Situational tests
X. Student projects
  A. Papers
  B. Notebooks
  C. Reports
  D. Autobiographies
  E. Personal data sheets
XI. Case studies
XII. Case conferences"

1/Schwartz and Tiedman, op. cit., pp. 11-12.

This list is essentially the same as that developed by Burton[1/] in 1950. The Eight-Year Study[2/] describes a variety of measuring devices for determining the degree to which students are attaining major curriculum objectives. They are noted below:

> "1. For aspects of thinking: tests of interpretation of data, application of principles, logical reasoning, and nature of proof.
>
> 2. For social sensitivity: tests of application to social problems of social values, social facts, and generalization.
>
> 3. For civic and social beliefs: scales of social, political, and economic beliefs.
>
> 4. For aspects of appreciation in literature and art: a variety of techniques.
>
> 5. For interests: an inventory of personal, social, and school interests.
>
> 6. For personal and social development: various self-reporting scales and anecdotal records."

Orata[3/] also notes the necessity of using a variety of techniques in a comprehensive appraisal program:

> "Aside from tests and scales, there are other instruments of evaluation that are not ordinarily regarded as such, as, for example, anecdotal records, data from questionnaires, records of attendance, records of physical examinations, diaries kept by the teacher, pupils, parents, case study records and other kinds of records that show or give evidence of significant changes in the behavior and development of the boys and girls in school. They may be formal or informal records, complete or incomplete records, just as long as they give lists of significant and valid evidence of changes in

---

1/Burton, op. cit., pp. 248-249.

2/Smith, Tyler, and others, op. cit., pp. 3-34.

3/Orata, op. cit., pp. 645-646.

> the pupils they are properly regarded as evaluation instruments. What seems important is not that each instrument is complete in itself, but that it contributes, along with other instruments, some evidence, slight though it may be, toward a comprehensive and dependable evaluation of the total school program."

Traxler[1] contends that the greater part of information may be obtained from tests and other objective techniques. The use of these techniques are recommended wherever possible, since the data obtained in this manner will tend to be reliable and impersonal. He points out, however, that certain kinds of information, including home background, social history, certain aspects of health history, extracurricular activities, work and summer experience, interests, school subjects preferred and those disliked, voluntary reading, and educational and vocational plans, cannot be obtained by objective measures. This type of information must be gathered directly from the pupil or parent by such means as the interview and questionnaire.

Greene, Jorgensen, and Gerberich[2] state that although the test constitutes the major type of evaluation instrument, other means such as the anecdotal record, the interview, and the rating scale have significant places in the evaluation of pupil behavior and achievement. They also point out that although certain aspects of pupil behavior lend themselves to objective measurement, there are many elements in the total understanding of the child which are represented by the

---

1/Traxler, op. cit., p. 25.

2/Harry A. Greene, Albert N. Jorgensen, and Raymond J. Gerberich, Measurement and Evaluation in the Elementary School (Second Edition), Longmans, Green and Company, New York, 1953, p. 7.

intangibles.[1] These elements lie beyond the objective measurement field and must be appraised on the basis of keen observation. Smith and Tyler[2] suggest that any device which provides valid evidence regarding the progress of students toward the educational objective is appropriate and should be included in the program. Schwartz and Tiedman[3] note the place of various devices in the evaluative process by the following comment:

> "There is no one method or technique of evaluation that is best for measuring the wide variety of objectives found in the usual school program, and the choice of technique depends almost entirely on the kind of objective to be measured. Measurement is not limited to the administration of objective tests, nor is it limited to the use of essay examinations, nor to the use of observational techniques. Each of these has its place in a total program of evaluation, and each can be uniquely useful when properly used."

A wide variety of devices is available in measuring pupil progress and adjustment. Writers agree that it is the task of the school to know the appropriate techniques, their uses and limitations, and how they may be employed to indicate the degree to which the pupil has attained the important outcomes of the educational process.

Summary.-- The ultimate success of a testing and evaluation program is dependent upon basic considerations which act as guideposts for the entire program. The operation of the program will hinge on the concept developed by the school. The type of program that emerges will be directly related to the philosophy of testing and evaluation, an

---

1/Greene, Jorgensen, and Gerberich, op. cit., p. 633.

2/Smith, Tyler, and others, op. cit., p. 13.

3/Schwartz and Tiedman, op. cit., p. 12.

understanding of the place of the evaluative process in the total educational program, an awareness of the aspects of pupil behavior with which evaluation must be concerned, and cognizance of various techniques that may be used in the pupil appraisal program.

A review of the literature reveals the following general principles:

1. It is only through being well informed regarding all aspects of growth and development of the individual pupils that the school staff can plan an effective educational program to meet the needs of the pupils.
2. The evaluation program is an integral phase of the educational process and should be considered one of the major tasks of the school.
3. Evaluation should not be regarded as an end process, but the means to an end--aiding the individual pupil in his growth and development.
4. A comprehensive program of evaluation should include the gathering of evidence on most or, if possible, all of the major objectives of the school program. It should not be limited to collecting data for a few instructional objectives of the conventional subject curriculum.
5. Adequate objective methods and instruments are not available for collecting evidence of the attainment of many objectives of the school. Appraisal in these areas must be accomplished by subjective methods.

6. A variety of techniques and measuring devices must be used for collecting data. These will range from simple pencil-and-paper tests to observation and sociometric techniques.
7. Major evaluation techniques available are: (1) objective tests including mental ability, aptitude, academic achievement, attitudes,[1] interests,[1] personality;[1] (2) anecdotal records; (3) observational techniques; (4) oral and essay examinations; (5) questionnaires, inventories, and interviews; (6) checklists and rating scales; (7) projective techniques; (8) case studies; (9) sociometric techniques; and (10) cumulative records.
8. The choice of the evaluative device is dependent upon the specific objective to be measured. No technique is worth using unless the results it yields can be depended upon.

3. Organizing the Program

Basic operational considerations.-- The administrative organization for carrying on a testing and evaluation program will differ considerably from community to community, depending on the size, type, and problem of the particular school system. Wrightstone, Justman, and Robbins[2] state that the type of curriculum, pupil personnel, and the like will, in large measure, determine its scope and nature. The competence with which the staff can use the results will also affect the

---

1/These are sometimes classified as scales or inventories.

2/Wrightstone, Justman, and Robbins, op. cit., p. 61.

program. Smith and Tyler[1] contend that the evaluation program should serve the local needs and purposes of each school and that the particular emphasis, as well as the extent of the program, should be determined largely by the needs of each school for data. King[2] points out that the extent and comprehensiveness of the program and organization will be directly proportional to the facilities available and to the training of those who are commissioned to operate it. The literature corroborates these points of view and further emphasizes the fact that there is no one set organizational pattern that must be adopted by all schools. Each school system should develop a plan in keeping with its particular needs and facilities.

The organizational plan may also vary considerably in a particular school system, depending on various local situations which may develop. Kirby[3] emphasizes this axiom by stating that it is not a good policy to set up a formal testing program to be followed without deviation year in and year out, since problems change, new objectives develop, and different areas need varying degrees of emphasis. He further suggests that the over-all program should not be crystalized, but rather flexible and adaptable. Any plan must be subject to adjustment to the individual school situation and be organized so that it may expand, contract, or shift emphasis as the needs require. Fullmer[4] contends

---

1/Smith and Tyler, op. cit., p. 443.

2/Joseph E. King, "Using Tests in the Modern Secondary School," The Bulletin of the National Association of Secondary School Principals (December, 1948), 32:46.

3/B. C. Kirby, "Minimum Testing Programs," Journal of Education (January, 1948), 131:24.

4/Daniel W. Fullmer, "The Testing Program: What Constitutes Minimum

that the important consideration is a test pattern to meet the individual's needs and not a stereotyped set of test results.

Despite the uniqueness of the organization for each school situation, the literature reveals that there are certain accepted general principles and procedures common to all testing and evaluation programs. This study presents these principles and procedures as set forth in the literature.

General principles.-- The testing and evaluation program must be set up as definitely as any other part of the educational structure. It should not be a hit-or-miss affair, but represent a well-ordered program worthy of the name.[1] Spence[2] suggests that the program should be "supplementary not duplicative, usable not confusing, economical not burdensome, comprehensive not sporadic, suggestive not dogmatic, progressive not static." He further suggests that this may serve as a good rule of thumb in planning the program.

Wrightstone, Justman, and Robbins[3] set forth four characteristics of an adequate evaluation program which may be used to appraise the adequacy of a program in a modern school and serve as a guide for program organization and development:

"1. Is the design of the evaluation program comprehensive,

Essentials," The Bulletin of the National Association of Secondary School Principals (May, 1958), 42:87.

1/C. C. Ross and Julian Stanley, Measurement in Today's Schools, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1954, pp. 175-178.

2/Ralph B. Spence, "A Comprehensive Testing Program for Elementary Schools," Teachers College Record (January, 1953), 34:281.

3/Wrightstone, Justman, and Robbins, op. cit., p. 27.

so that it includes not only abilities, skills, and understandings, but also the less tangible objectives of learning and instruction?

2. Are changes in the behavior of the individual the basis for evaluating his growth and development, since the total behavior of the individual--mental, physical, emotional, and social--should be the concern of the teacher and supervisor in every situation?

3. Are the results of the evaluation organized into a meaningful interpretation so that a portrait of the individual's growth and development and the interrelationships of such growth become evident?

4. Is the evaluation program continuous and interrelated with the curriculum development?"

Several writers suggest general principles pertaining to testing programs. These are equally applicable to the over-all evaluative program. The Kent State study[1] notes that the program should be three-dimensional, having (1) length--the program must be continuous, testing certain basic factors on two or more occasions over a period of time; (2) breadth--the program must include measurement of various traits or characteristics so that interpretations of one test may be made in the light of other test data as well as nontest information; and (3) depth--including provisions for individual or clinical testing. This study[2] indicates that (1) tests provide only part of the information needed to aid pupils; (2) the adequacy and effectiveness of a testing program can be measured only by the use made of the results; (3) the testing program should be closely related to the needs of the

---

1/Kent Area Study Council, A Proposed Twelve-Year Testing Program, Ohio Scholarship Tests, State Department of Education, Columbus, Ohio, 1959, p. v.

2/Ibid., p. 2.

local school; and (4) testing should be planned as an integral part of the total school program. Lennon[1/] lists the following four general principles of a sound high school testing program: (1) the program should be continuous; (2) the program should be comprehensive; (3) testing should be jointly planned; and (4) the testing programs should be integrated with the total educational program.

Erickson[2/] sets forth the following criteria that may be used as guiding principles in program planning and organization:

> "1. A program that is planned, administered, and utilized co-operatively by teachers, supervisors, curriculum workers, counselors, and administrators.
>
> 2. A program which is devoted to the improvement of the effectiveness of teaching and guidance in the school.
>
> 3. A program which will help pupils appraise and understand their relative strengths and weaknesses in achievement, aptitude, adjustment, personality, and interests.
>
> 4. A program that will build favorable public relations with school patrons and the community at large."

The literature reveals that the testing and evaluation program should be purposive, long-ranged, cooperatively planned, practical, and professionally administered. These represent organizational aspects of the program and are considered further in this study.

Purposive.-- For any program to be justified, it must be purposive.

1/Roger T. Lennon, Testing in the Secondary School, Test Service Notebook No. 20, World Book Company, Tarrytown-on-Hudson, New York, 1957, p. 4.

2/Elmer J. Erickson, "What Kind of Testing Program for Today's Schools?" The Bulletin of the National Association of Secondary School Principals (April, 1952), 36:160.

Before any attempt is made to collect data about pupils, the purposes for which they are to be used must be determined and thoroughly understood by all members of the staff.[1] Wertz[2] indicates that the purposes must be determined by the instruction and guidance staff members who are going to use the program and must reflect the needs felt by the staff.[3] Ross notes that the purposes should be stated specifically in terms of both the scope and purpose.

Noll[4] lists the most common purposes for which educational measurements are used:

"A. Classification of pupils
B. Homogeneous grouping
C. Diagnosis and remedial work
D. Counseling and guidance
E. Marking
F. Motivation
G. Identification and study of exceptional children
H. Interpreting schools to the community
I. Improvement of school staff
J. Educational research."

Torgerson[5] suggests that the purpose of the program is to collect information which may be used to solve important educational problems of vital concern to the school staff. He further notes that these

---

1/J. Murray Lee and Doris May Lee, The Child and His Curriculum, Appleton-Century-Crofts, Inc., New York, 1950, p. 674.

2/Henry Wertz, "Minimum Essentials for a Testing Program," American School Board Journal (September, 1957), 135:42.

3/Ross, op. cit., pp. 175-178.

4/Noll, op. cit., p. 321.

5/Theodore L. Torgerson and Georgia S. Adams, Measurement and Evaluation for the Secondary School Teacher, Dryden Press, New York, 1956.

problems should be concerned with factors that condition the growth and development of pupils in harmony with the objectives of education. Durost[1] states that tests may be used for many purposes in that they may serve the administrator, supervisor, and teacher. Traxler[2] states that most tests have multiple uses and that the specific purposes will, in large measure, determine the nature of the program and the kind of measures used. Wertz[3] further points out the necessity of having the community understand and accept the purposes of the program.

Writings reveal that the testing and evaluation program may be multipurposive in that it may serve the varying needs of all staff members by supplying a great variety of information about each pupil. The program will be successful to the extent that it meets the purposes for which it is designed.[4] The program must therefore be planned in accordance with the stated purposes. Adequate follow-up work must be carried on to determine the degree to which the purposes have been achieved.

Long-range and continuous program.-- The program which looks toward completeness must be instituted gradually over a period of years. Ross and Stanley[5] state that the total program should be

---

1/Durost, op. cit., p. 1.

2/Arthur E. Traxler, "Ten Essential Steps in a Testing Program," Education (February, 1959), 79:358.

3/Wertz, op. cit., p. 42.

4/Noll, op. cit., p. 320.

5/Ross and Stanley, op. cit., p. 210.

planned for an extended period rather than being adopted piecemeal. In this way a varied program, omitting gaps and avoiding needless duplication, may result. They further advocate systematically recurring schedules as opposed to sporadic testing. Lennon[1/] holds this view and further notes that it is only through a continuing program that measurement of growth and progress, and evaluation of changes from year to year, may be accomplished.

The literature discloses that the chief value of testing and evaluation lies in its regularity. Instructional planning is most effective when it is based upon regular appraisal of the growth and development of the pupil. Sporadic appraisal may help solve immediate problems, but it does not provide sufficient information for effective guidance and instructional planning. The literature points to the value of planning a systematic, sequential program covering a period of years, with a degree of flexibility which will allow the gradual introduction of new measuring devices when they prove to be better than the older ones.

Beer[2/] recommends that the program be introduced and proceed at a reasonable rate of development so that all staff members may recognize its needs, sense its values, and learn how to analyze and use its data. He further suggests that additions to the program be instituted grad-

---

1/Lennon, op. cit., p. 4.

2/Carl R. Beer, A County Program of Standardized Testing, Test Service Bulletin No. 60, World Book Company, Tarrytown-on-Hudson, New York, p. 5.

ually, so that teachers will have time to determine values of the new information and see the relationship between the new information and that already known. Ross and Stanley[1/] feel that the scope of the program should be rather limited at first, but that plans be established to broaden it as the teachers become aware of the needs for additional information and trained in interpreting and using such data. They believe that it is the best plan to run the risk of undertaking too limited a program rather than one too broad.

Cooperative enterprise.-- The testing and evaluation program represents a cooperative enterprise on the part of parents, pupils, teachers, counselors, and administrators. The entire staff must have a voice in determining the purpose and in formulating the plans for the program.[2/] Traxler[3/] recommends that the program be undertaken cooperatively by the faculty and that planning and operation be made as democratic as possible.[4/] Ross and Stanley believe that the program will fail without the cooperation of all staff members. All the literature reveals that the program of evaluation cannot be separated from the total educational program and thus must enlist the cooperative endeavors of all personnel in administration, guidance, and instruction.

1/Ross and Stanley, op. cit., pp. 180-182.

2/Smith and Tyler, op. cit., p. 441.

3/Arthur E. Traxler, Robert Jacobs, Margaret Selover, and Agatha Townsend, Introduction to Testing and the Use of Test Results in Public Schools, Harper and Brothers, New York, 1953, p. 13.

4/Ross and Stanley, op. cit., p. 179.

King[1] states that it is important that every phase of the testing program be planned cooperatively with all staff members participating actively. He further emphasizes the importance of teacher participation by the following comment:

> "Testing and learning are first cousins. The classroom teacher is responsible for the learning progress of her pupils, within reasonable limits, so she is surely entitled to the privilege of full participation in the program of evaluation designed to measure the learning progress of pupils. If test results are to facilitate effective teacher appraisal of curriculum and teaching methods, she must be a full partner in the school's co-operative testing enterprise."

Boyer[2] emphasizes this point and states that the cooperation of the teacher must be obtained, since the failure or success of the program is largely determined by what the teacher does as the result of his thinking.

The Commission on the American School Curriculum[3] strongly recommends that teachers share the responsibility for evaluation as well as for curriculum construction, and cite the following specific reasons for teacher participation in the evaluation program:

> "1. Classroom teachers can make valuable contributions, often indispensable, in determining what behaviors should be evaluated, the degree of accomplishment which it is reasonable to expect at the various grade levels, the appropriateness of proposed evaluation

---

1/King, op. cit., p. 57.

2/Philip A. Boyer, "The Administration of Learning Groups in Elementary Schools," The Grouping of Pupils, Part I, Thirty-fifth Yearbook of the National Society for the Study of Education, Public School Publishing Company, Bloomington, Illinois, 1936, p. 213.

3/American Association of School Administrators, op. cit., p. 339.

instruments as measures of the curriculum objectives, and particularly the relevance of proposed measures to the objectives that guide instruction in their own classes.

2. They can aid in the experimental tryout of standardized tests and of other technics of appraisal. In many cities such tests are not chosen for system-wide use until they have been tried out in a number of classrooms in order to secure the judgment of teachers as to their appropriateness in terms of the school objectives, difficulties of administration and scoring, and the interpretation and use made of results. Such preliminary tryouts are equally valuable for any form of appraisal intended for city-wide use.

3. Committees of teachers, with the aid of the central staff, may undertake to devise methods of evaluating special units and to assist in planning for the evaluation of areas not now adequately measured by standardized tests.

4. Participation in the evaluation program is one of the best methods of helping classroom teachers clarify their own objectives as well as their understanding of the ultimate objectives set up for the entire curriculum.

5. Participation also helps classroom teachers grow in their understanding of the purpose of evaluation and to improve their own evaluation procedures. It goes a long way toward removing the distrust and uneasiness which teachers have when the evaluation program is planned solely in the central office.

6. The most compelling argument for the participation of classroom teachers is that the effective utilization of the results of any type of evaluation largely depends upon them. Moreover, the teacher's attitude toward evaluation is certain to be reflected in the attitudes of the students."

Froehlich and Benson[1/] feel that the entire program must be based on the results of a cooperative study by all members of the school

1/Clifford P. Froehlich and Arthur S. Benson, Guidance Testing, Science Research Associates, Chicago, Illinois, 1948, p. 6.

staff having need of the data in dealing with pupils. Lennon[1] states that the program should be jointly planned. The writings imply that the testing and evaluation program should be an outgrowth of important educational problems of vital concern to teachers. Only in this way can all staff members develop common understandings of the purposes of the program and a realization of its benefits and vital role in the educational process.

Practical.-- To be justified, the program must be practical.[2] Froehlich and Benson[3] state that what is practical for one school may be out of the question for another. They further note that a practical program is one that is adapted to the local needs and conditions. Ross and Stanley[4] recommend that the program be planned to obtain data helpful in solving some practical school problem or issue.[5] Spence notes that the program should be economical from the standpoint of time and money, and subject to adjustment to the individual school community, depending on the facilities and personnel. Lass and Wrightstone[6] recommend that the program not exceed the practical school uses

---

1/Lennon, op. cit., p. 4.

2/Ross and Stanley, op. cit., p. 179.

3/Froehlich and Benson, op. cit., p. 6.

4/Ross and Stanley, op. cit., pp. 182-183.

5/Spence, op. cit., pp. 283-284.

6/A. N. Lass and J. Wayne Wrightstone, "Evaluation in the Secondary Schools of New York City," High Points (March, 1946), 28:26.

that can be made of data that are obtained. He further notes that the program should be initiated gradually so that the staff will not be overburdened with too many tests and techniques.[1] The findings reveal that the school should obtain only as much information as it is staffed to utilize effectively and assimilate efficiently.

Froehlich and Benson[2] suggest that routine clerical work in scoring, tabulating, and recording data be kept at a minimum. They note that the program loses much of its value if teachers become overloaded in carrying out the details and thus become "fed up" with the entire operation. The loss in time in the regular school schedule should not be out of proportion to the expected instructional and guidance gains.

Professional.-- The program should possess a high degree of professionalism.[3] The measures of the program must be properly administered and the results interpreted and used to improve the instruction, distribution, and adjustment of individual boys and girls.[4] To insure this, King[5] recommends that a well-trained person be given the responsibility of organizing the program. This person should be responsible for training teachers in interpreting and using evaluative results in a professional manner, and should make certain that the

---

1/Wrightstone, Justman, and Robbins, op. cit., p. 61.

2/Froehlich and Benson, op. cit., p. 6.

3/Ibid.

4/Arthur E. Traxler, Techniques of Guidance, Harper and Brothers, New York, 1957, p. 266.

5/King, op. cit., p. 58.

program operates on a professional basis. Lass and Wrightstone[1] hold the same opinion as King and note that the program aims not only at measuring and inducing pupil growth, but also provides an opportunity for continuous teacher education and professional growth. Noll[2] expresses concern for the professional aspect of the program and lists the following qualities needed by the person administering standardized group tests: (1) ability to understand and follow directions, (2) ability to maintain the attention and cooperation of a group, (3) ability to read directions aloud clearly and distinctly, and (4) ability to be objective.

The study by Leonard[3] reveals that fewer than one fourth of the school districts surveyed had a professionally trained person in testing and evaluation. Michaelis[4] found that the majority of city school districts had a director of evaluation. These studies reveal that, although improvements have been made, the schools have some distance to go in reaching the goals recommended in current writings.

1/Lass and Wrightstone, op. cit., p. 26.

2/Noll, op. cit., pp. 331-334.

3/Eugenie A. Leonard and Anthony C. Tucker, The Individual Inventory in Guidance Programs in Secondary Schools, United States Office of Education Vocational Division, Bulletin No. 215, Occupational Information and Guidance Series No. 7, United States Government Printing Office, Washington, D. C., 1941.

4/John W. Michaelis, "Current Practices in Evaluation in City School Systems," Educational and Psychological Measurement (Spring, 1950), 9:15-22.

4. Administering the Program

Stages in administering the program.-- The extent to which testing and evaluation fulfills its purpose is dependent upon the administration of the program and the utilization of the results. Certain administrative and planning aspects have been the concern of several writers.

Traxler[1] listed ten essential steps in a testing program, setting forth procedures involved in planning and carrying on a program of objective testing. These steps are aimed at assisting schools in making certain that tests are administered correctly, interpreted wisely, and used effectively. They set forth, in outline form, important procedural aspects of the program as follows:

> "1. Decide what is to be measured.
> 2. Decide for what purposes the measuring devices are to be used.
> 3. Decide when the tests are to be given.
> 4. Decide the grade levels at which the tests are to be given.
> 5. Decide how the test should be chosen, administered, and used.
> 6. Choose the test to be used in each testing program.
> 7. Make certain that the tests are carefully administered, with meticulous attention to the directions.
> 8. Provide for rapid and accurate scoring of the tests and statistical treatment of the results.
> 9. Organize the test results for use by teachers, counselors, principals, and other school functionaries who have access to the scores.
> 10. Inform and train the faculty in the nature and purposes of tests and testing and the use of test results."

Several checklists in the form of planning guides were also noted. These are aimed at assisting the school staff in working out details

---

1/Arthur E. Traxler, "Ten Essential Steps in a Testing Program," Education (February, 1959), 79:357-362.

and acquainting the staff with the specific factors involved in establishing and administering the program. Typical of these guides is Planning the Testing Program,[1] which covers the following aspects:

"1. Determining the purposes
2. Selecting grades to test
3. Choosing the tests
4. Selecting the testing days
5. Rearranging the daily schedule during testing
6. Administering tests
7. Scoring tests
8. Planning teacher conferences
9. Making test scores meaningful
10. Using test results."

Steps in administering an evaluation program were the concern of Jordan,[2] Ross and Stanley,[3] Noll,[4] Traxler et al.,[5] and several others.[6] Wrightstone, Justman, and Robbins list five steps in the evaluation program:

"1. Formulation of the major objectives of the curriculum
2. Definition and clarification of these major objectives
3. Selection of available tests and measures for each objective
4. Construction of needed test scales and techniques
5. Application of the various formal and informal tests and techniques to the appraisal of individual growth and development."

1/World Book Company, Planning the Testing Program--Testing Program Organizational Chart, Tarrytown-on-Hudson, New York.

2/A. M. Jordan, Measurement in Education, McGraw-Hill Book Company, New York, 1953, pp. 67-73.

3/Ross and Stanley, op. cit., Chap. XIII.

4/Noll, op. cit.

5/Traxler, and others, Introduction to Testing and the Use of Test Results in Public Schools, Harper and Brothers, New York, 1953, pp. 13-19.

6/Wrightstone, Justman, and Robbins, op. cit., pp. 17-21.

The writings reveal at least eight steps in a comprehensive and complete testing program. Regardless of the scope, these steps represent common procedural aspects for administering any program of testing and evaluation. The following represent various administrative stages in the program:

1. Establishing the purposes of testing and evaluation
2. Selecting appropriate instruments to be used in the program
3. Scheduling the program
4. Administering the evaluative devices
5. Scoring the tests
6. Recording and reporting test results
7. Analyzing and interpreting results
8. Using and applying test results.

The place of objectives in the evaluation program.-- Smith and Tyler[1] state that the first step in the development of an appraisal program is for the staff to cooperatively decide what to evaluate, what kind of evidence to secure, and how to go about securing and using it. This involves securing a statement of the objectives which are representative of the work done by the school and the phase of growth represented by the pupil. Traxler[2] notes that the objectives represent the clues as to what should be measured in any school and must receive primary attention in any evaluative procedure. Taba[3] makes

1/Smith and Tyler, op. cit., pp. 439-459.

2/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 357.

3/Hilda Taba, "The Function of Evaluation," Childhood Education (February, 1939), 15:249.

the following statement in regard to the place of objectives in the pupil appraisal program:

> "A program of appraisal should start with a careful formulation and listing of objectives by individual teachers, or by groups of teachers within one school or by groups of schools. This formulation and listing is important for the purpose of developing instruments of appraisal which give the kinds of evidence that are significant to teachers and which bring results which they can appreciate and use. This formulation is also important for the purpose of assuring that the total program of appraisal covers all the significant types of child development and is in that sense comprehensive enough. This process also points out to all participating teachers the values important to the development of the whole personality and thus promotes a cooperative attack for their development."

Smith and Tyler[1] recommend that the whole staff participate in forming the basic platform, each teacher or department submitting a list of objectives of their particular concern.[2] Ahman and Glock stress the importance of teacher participation in this important phase by the following comment:

> "The teacher is the most prominent single person involved in the process of identifying educational objectives for the classroom. It is he who must amalgamate the general and specific objectives of the kind mentioned with a galaxy of pertinent local factors in order to arrive at the proper educational objectives for his class and to afford each objective its proper emphasis."

They further note that teacher-derived educational objectives serve the following two functions: (1) the basis on which the curriculum is developed and the teaching process is organized, and (2) the basis on

---

1/Smith and Tyler, op. cit., p. 18.

2/J. Stanley Ahman and Marvin D. Glock, Evaluating Pupil Growth, Allyn and Bacon, Boston, 1958, p. 41.

which the evaluation procedures are planned.

Objectives are value judgments and represent choices by educators and other individuals.[1] Accordingly, they are determined from a variety of sources.[2] The Commission on the American School Curriculum lists the following three sources as examples: (1) accumulated knowledge of the scholars, (2) analysis of society, and (3) the nature of learner. It further notes that the objectives provide the foundation for all curriculum planning and the basis for the evaluative process. Wrightstone, Justman, and Robbins[3] note that although the pattern of objectives may vary from school to school, a comprehensive range of major objectives would be concerned with physical and mental health, social relationships, skills and knowledges, appreciations and creative expression, critical thinking, interests, and attitudes.

Smith and Tyler[4] offer the following classification of objectives in order to focus attention on areas requiring evaluative measures:

> "1. The development of effective methods of thinking
> 2. The cultivation of useful work habits and study skills
> 3. The inculcation of social attitudes
> 4. The acquisition of a wide range of significant interests
> 5. The development of increased appreciation of music, art, literature, and other aesthetic experiences
> 6. The development of social sensitivity
> 7. The development of better personal-social adjustment
> 8. The development of important information
> 9. The development of physical health
> 10. The development of a consistent philosophy of life."

1/American Association of School Administrators, op. cit., p. 55.

2/Ibid., p. 56.

3/Wrightstone, Justman, and Robbins, op. cit., p. 17.

4/Smith and Tyler, op. cit., p. 18.

The Taxonomy of Educational Objectives[1] represents an excellent source for the classification of educational objectives that deal with the recall of knowledge and the development of intellectual abilities and skills. It further provides an excellent basis for building evaluative instruments in these areas. An excellent listing of specific objectives of the general education program in the secondary school is provided by French.[2] Kearney[3] provides a companion list for the elementary school years, as recommended by the Mid-Century Committee on Outcomes in Elementary Education.

Burton[4] states that all schools should define what is being appraised. He further notes that if the program is purposive, the first step in the evaluative process is a clear statement of what is being evaluated.

In his survey, North[5] found that 94 per cent of the schools

---

1/Benjamin Bloom (Editor), Taxonomy of Educational Objectives, Longmans, Green and Company, New York, 1956.

2/Will French and associates, Behavioral Goals of General Education in High School, Russell Sage Foundation, New York, 1957.

3/Nolan C. Kearney, Elementary School Objectives, Parts II and III, Russell Sage Foundation, New York, 1953.

4/William H. Burton, "Implications for Organization of Instruction and Instructional Organization," op. cit., p. 249.

5/Robert D. North, "Testing Programs of Public School Members of the Educational Records Bureau--Report of a Questionnaire Survey," 1956 Achievement Testing Program in Independent Schools and Supplementary Studies, Bulletin No. 68, Educational Records Bureau, New York, 1956, pp. 76-90.

claimed that their testing programs were based on defined educational objectives. This claim needs investigation.

The literature pertaining to objectives is voluminous. A comprehensive review is entirely beyond the scope of this study. Objectives represent an essential element of any testing and evaluation program and should not be neglected in any treatment of the program.

Selecting measuring devices and instruments.-- Once the staff has determined the general scope of the program, what is to be measured, the method for securing the needed evidence of the degree of attainment of the objectives must be considered.[1/] This will involve selecting tests or measures. Although a variety of measuring and evaluative devices is necessary in a comprehensive program, most of the literature is directed toward test selection, since test selection is a complicated procedure and there are many tests available from which to make a choice. Traxler[2/] notes that there are over 108,000,000 tests administered annually in the schools of the United States. He further notes that there are about 20 test publishers with numerous test offerings. More than 5,000 different objective tests have been published. At least 1,000 of these are on the active list of publishers. The Fifth Mental Measurements Yearbook[3/] reviews 957 commercially available tests during the 1952-58 period. With more emphasis being

1/Smith and Tyler, op. cit., p. 444.

2/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 358.

3/Oscar H. Buros (Editor), The Fifth Mental Measurements Yearbook, Gryphon Press, Highland Park, New Jersey, 1959.

placed upon tests, the writings give more import to this phase of the over-all program. This study considers important aspects of test selection. Principles presented by the literature apply to a considerable extent to measuring instruments and devices of all types.

The literature points to several basic issues that are involved in test selection, namely: (1) who shall select the test, (2) what are the sources of information about available measuring instruments, (3) what types of measuring instruments shall be used, (4) what procedure shall be followed in making the selection. The first, second, and fourth issues are the immediate concern of this review. The third aspect is treated in a later section dealing with measures used in the program and specific testing patterns.

Great care must be exercised to secure tests most appropriate for the intended purpose. Wertz[1/] states that the responsibility for selection rests with the school faculty and administration. Ross and Stanley[2/] note that the best qualified person, or persons, available should make the selection. In large cities this will be the Research Director or Testing Coordinator. In the small district it will undoubtedly be the principal. He further recommends that a staff committee be set up to aid in this task. The Kent Area Study[3/] recommends that, although the test coordinator will ordinarily have the responsibil-

1/Wertz, op. cit.

2/Ross and Stanley, op. cit., p. 214.

3/Kent Area Study Council, op. cit., p. 11.

ity for test recommendation, other appropriate staff members should be involved in the selection process. A test selection committee composed of measurement specialists, curriculum specialists, and classroom teachers is advocated. Traxler[1/] states that the decision regarding test selection should not be made by one person. He recommends that the selection should be made by a testing committee representing all departments of the school. The use of outside consultative service is recommended by Jordan.[2/] The literature supports the views presented, and holds that test selection is a coordinated group enterprise with each of several groups making a definite contribution. Writers in the field appear to agree that it is a sound principle to rely on the combined judgments of a group of competent people rather than upon the judgment of one individual. The study of Belanger[3/] reveals that this principle is common practice in California schools where counselors, teachers, and principals participate in this process.

Writings cite that test selection involves a systematic procedure aimed at procuring the instrument best suited to measure the stated objectives. Wrightstone, Justman, and Robbins[4/] advocate that the school maintain an extensive and up-to-date file of standardized tests. Noll[5/]

1/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 359.

2/Jordan, op. cit., p. 68.

3/Laurence L. Belanger, "Testing in California Secondary Schools," California Journal of Secondary Education (February, 1947), 22:108-111.

4/Wrightstone, Justman, and Robbins, op. cit., p. 62.

5/Noll, op. cit.

recommends the use of catalogs and other advertising material distributed by test publishers. Test manuals and the Mental Measurements Yearbook[1] are recommended in the writings as primary sources. Testing service publications by the World Book Company, California Test Bureau, and the Psychological Corporation are referred to rather extensively. Textbooks such as the following represent valuable sources of test selection: Greene, Jorgensen, and Gerberich,[2] Ross and Stanley,[3] Traxler,[4] Noll,[5] and Remmer and Gage.[6] Publications by the American Psychological Association[7] represent additional sources.

In selecting a standardized test, various qualities which characterize the instrument must be considered. Wrightstone, Justman, and Robbins[8] recommend preparing a checklist of qualities desired and then using a rating device to quantify the judgment. Several rating scales

---

1/Buros, op. cit.

2/Greene, Jorgensen, and Gerberich, op. cit.

3/Ross and Stanley, op. cit.

4/Arthur E. Traxler, Techniques of Guidance, Harper and Brothers, New York, 1957.

5/Noll, op. cit.

6/Remmer and Gage, op. cit.

7/American Psychological Association, "Technical Recommendations for Psychological Tests and Diagnostic Techniques," Supplement to the Psychological Bulletin (March, 1954), Vol. 51, No. 2, Part 2, American Psychological Association, Washington, D. C.

8/Wrightstone, Justman, and Robbins, op. cit., p. 63.

are mentioned prominently, among which are Otis,[1] Rinsland,[2] and Cole and von Borgersrode.[3]

Several criteria by which selection of tests may be guided are noted in the literature. Lennon[4] sums up these criteria under the following five headings: (1) validity, (2) reliability, (3) appropriate difficulty level, (4) norms, and (5) administrative considerations. Noll[5] lists nine characteristics of a good measuring instrument as follows: (1) reliability, (2) validity, (3) objectivity, (4) ease of administration, (5) ease of scoring, (6) ease of interpretation, (7) adequate norms, (8) equivalent forms, and (9) economy.

Brownell[6] emphasizes the close relationship between teaching and testing in setting forth the following criteria:

> "1. Does the test elicit from pupils the desired types of mental processes?
> 2. Does the test enable the teacher to observe and analyze the thought processes which lie back of the pupil's answers?

1/Arthur S. Otis, Otis Scale for Rating Standardized Tests, World Book Company, Tarrytown-on-Hudson, New York.

2/H. D. Rinsland, "A Form for Briefing and Evaluating Standardized Tests," Journal of Educational Research (January, 1949), 42:371-375.

3/Robert D. Cole and Fred von Borgersrode, "A Scale for Rating Standardized Tests," School of Education Record of the University of North Dakota (October, 1928), 14:11-15.

4/Roger T. Lennon, Planning a Testing Program, Test Service Bulletin No. 55, World Book Company, Tarrytown-on-Hudson, New York, pp. 4-7.

5/Noll, op. cit., pp. 66-87.

6/William A. Brownell, "Some Neglected Criteria for Evaluating Classroom Tests," National Elementary Principal (July, 1937), 16:485-492.

3. Does the test encourage the development of desirable study habits?
4. Does the test lead to improved instructional practice?
5. Does the test foster wholesome relationships between teacher and pupils?"

The Kent Area Guidance Council[1] lists the following considerations in choosing a test:

"1. Test should give usable information.
2. Test should fit the range of capacities of the pupils being tested.
3. The length should be appropriate to the use.
4. The test should possess a degree of ease of administering and scoring.
5. Cost and service factors should be reasonable.
6. Equivalent forms should be available.
7. The test should have high reliability.
8. The test should have high validity.
9. The test should have adequate published information."

Test selection depends primarily on the type of program planned. According to Zehrer[2] selection and use are governed by the extent to which the findings contribute to the better understanding of the individual. He further notes that each instrument is chosen for a definite purpose. Sister Afra[3] holds these views and adds that the amount of time available for interpreting the results is also a determining factor. Test selection involves many factors. Decisions must be based on criteria established to meet the particular needs of the school. The criteria reviewed in the literature may be summed up under

---

1/Kent Area Study Council, op. cit., p. 11.

2/Frederick A. Zehrer, A Basic Testing Program for the High School, Test Service Bulletin No. 47, World Book Company, Tarrytown-on-Hudson, New York, p. 4.

3/Sister M. Afra, S.C.L., Standardized Tests: Their Uses and Abuses, Test Service Bulletin No. 58, World Book Company, Tarrytown-on-Hudson, New York, 1945, pp. 5-8.

the following headings:

1. Validity
2. Reliability
3. Administrability
4. Interpretability
5. Practicability
6. Economy

Scheduling the program.-- Scheduling the testing program is an administrative problem of major import. When to give tests, how much time to devote to testing, and so forth, require decisions prior to administering the tests.[1/] Frequency and grade levels of testing represent aspects of this problem.[2/] These are considered in a later section of this study.

Smith and Tyler[3/] recommend that the total amount of time devoted to testing be such that the faculty not feel overburdened with tests. They state that an adequate amount of time must be devoted to the program, but note that the purpose of the program may be defeated if too much time is spent in testing. Undue concentration of formal testing toward the end of the school year is frowned upon. It is further recommended that the schedule call for a fair distribution among subject fields so that no undue amount of time is taken from any one class.

1/Ross and Stanley, op. cit., pp. 225-226.

2/Noll, op. cit., p. 321.

3/Smith and Tyler, op. cit., pp. 447-449.

Writings reveal that there are two seasons of the year in which testing is usually done--fall and spring. The literature points out advantages and disadvantages in each of these periods.

Ahman and Glock[1/] recommend fall administration, giving the following reasons:

> "1. With today's itinerant population, fall administration assures the teacher a record for each pupil. Transfer pupils may not bring adequate records with them from other schools. If tests are administered in the spring, the teacher is more likely to have inadequate data for some pupils. This prevents him from focusing his attention on pupil needs rather than on what is supposed to be taught in a certain class. If pupils have specific weaknesses, he can plan remedial work. If they rank high in achievement, he won't make the serious mistake of having them repeat what they already know just because it appears in the syllabus. The information enables him to meet the needs of individuals through grouping or through any other practical approach.
>
> 2. Fall testing provides a more realistic measure of pupil's achievement. During the summer vacation, certain skills may improve; others may deteriorate. If the child knows arithmetic, he may not be as proficient in September as he was in June. On the other hand, he may have done considerable reading during the summer, thereby improving his skill in this area.
>
> 3. Testing can be helpful to pupils in self-evaluation. If the results are used wisely, they can provide aid in formulating goals. Knowledge of the results of learning is a very important factor in motivation. The objective data from standardized tests can give the pupil direction and purpose for the new school year.
>
> 4. Too often it appears that tests are administered to determine the effectiveness of the teacher rather than the status of the pupils. This may result in coaching pupils for tests or invalid administration such as lengthening time limits. Fall testing lessens this possibility.

1/Ahman and Glock, op. cit., pp. 490-491.

5. Testing in the spring is generally done under pressure of time. As the school year draws to a close teachers are busy filling out reports and completing final details. It tends to more of a chore than an aid to good teaching. The teacher does not become enthusiastically involved in testing."

Jordan[1] leans toward fall testing since results obtained at this time may be used for planning programs of improvement, for grouping of pupils within the class for purposes of instruction, and for deciding upon differential procedures for slow and fast learners. Beer[2] states that the beginning of the term is best for testing since (1) teachers can make the most effective use of test information gathered at the opening of the term, and (2) early administration makes tests a starting point rather than goals in themselves and thus gives impetus away from rather than toward themselves. Morrison[3] points out that fall testing serves many functions which the end-of-the-year testing cannot serve, since the results of the latter will be of use primarily to subsequent teachers. Traxler[4] prefers fall testing for scholastic aptitude tests, tests of basic skills, and tests of broad field so that results may be used during the year by the instructors, counselors, and remedial teachers. Writers also indicate that fall testing does not encourage teachers to teach for the test.

1/Jordan, op. cit., p. 69.

2/Beer, op. cit., pp. 5-6.

3/Harriet Barthelmess Morrison, The Fall Testing Program, Test Service Bulletin No. 62, World Book Company, Tarrytown-on-Hudson, New York, 1948, pp. 3-7.

4/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 358.

Traxler[1] notes that spring testing is advisable when tests of achievement in specific courses such as algebra, biology, or world history are employed. He further points out that results of spring testing should be reported to the teachers before the school year ends. Ross and Stanley[2] favor giving tests at the end of the school year when the pupils' status is more normal. They argue that a comparison between the records made by pupils at the end of each of two successive years is usually more trustworthy than that between tests at the beginning of one year. It is stated that an analysis of errors revealed can serve equally well as the basis of remedial teaching in the succeeding grade as if the new teacher had given the test at the beginning of the year.

The Division of Research and Test Service of the World Book Company[3] recommends that the scheduling of tests depends on the type of test to be used. The following outline is offered as an aid in arriving at scheduling decisions:

"Diagnostic Tests	When any pupil gives evidence of difficulty in learning a subject or any part of it.
Instructional Tests	Upon the completion of each unit of subject matter at fairly frequent intervals throughout the year.
Prognosis Tests	Before the students undertake the study of a subject.

1/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 358.

2/Ross and Stanley, op. cit., p. 226.

3/World Book Company, Types of School Tests, Test Service Bulletin No. 19, World Book Company, Tarrytown-on-Hudson, New York, pp. 4-5.

Aptitude Tests	Before pupils enter a course in which special training is to be given, as in mechanical types of work.
Survey Tests	At any time during the year.
Intelligence Tests	Preferably at the beginning of the school term or year, but may be given any time during the school year. When there is a marked discrepancy between the intelligence test record and achievement, another intelligence test should be given.
Achievement Tests	Preferably at the end of the school year or term, but may be given any time during the school year.
Subject Tests	Preferably at the end of each school term or year, but may be given any time during the school year.
Analytical Subject Tests	Preferably at the beginning of the school term, but may be given any time during the year."

The literature indicates that the decision of timing is largely dependent upon the purpose for which the tests are intended. Generally speaking, tests administered in the spring have greater value for administrative uses, while those administered in the fall have greater instructional utility. The applicability of test norms is another factor affecting timing. Most of the writings lean toward autumn testing.

Studies of current practices reveal that 70 per cent of the schools in the Western New York area administer tests during September, October, and/or May.[1] Lee[2] found that 11 per cent of the schools administer

---

1/Edwin F. Rosinski, "A Status Survey of Existing Testing Programs in Member Schools of the Western New York School Study Council," in The Fourteenth Yearbook of the National Council on Measurements Used in Education, Edith M. Huddleston (Editor), Educational Testing Service, Princeton, New Jersey, 1957, p. 40.

2/J. Murray Lee, Testing Programs for Secondary Schools, Teachers Col-

tests at the end of the year and 42 per cent administer tests at the beginning of the year.

Administering evaluative instruments.-- The literature indicates that tests and other evaluative devices must be carefully administered with meticulous attention to the directions. Traxler[1] states that staff members who are to participate in this task should be carefully chosen and thoroughly briefed in advance.

Writers generally agree that individual tests and tests for special purposes should be given by the school psychologist or specially trained personnel. In the general group testing program it is recommended that the classroom teacher administer the tests. Ross and Stanley[2] make the following comment regarding this procedure:

> "When tests are used for the purposes of research, or when they are used to compare one grade, class, or school with others, they should usually be given by one person, or a small group of specially trained examiners. But in the ordinary testing program, employing group intelligence tests, the regular classroom teachers should usually administer the tests."

Many writers point to the advantage of having the teacher administer tests. Ahman and Glock[3] make the following comment pertaining to this practice:

> "It aids in identifying the teacher with the testing program since he must familiarize himself with the tests. Because the teacher knows his pupils, he can more effectively

---

lege Contributions to Education, Teachers College, Columbia University, New York, 1934.

1/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 359.

2/Ross and Stanley, op. cit., p. 227.

3/Ahman and Glock, op. cit., p. 488.

motivate them. He can interpret their reactions and establish the rapport necessary for adequate performance."

Noll[1] advocates that the administration be handled by only a few people if a considerable amount of testing is to be done. By following this procedure, he points out that these individuals may be given intensive training in administration. This plan tends to increase accuracy, uniformity, and efficiency.

The literature points to the necessity of test administration being done as well as possible. It stresses accuracy of timing and of adhering to instructions. With proper training, it indicates that teachers can obtain practically the same results with group tests as can be obtained by special examiners.

Rosinski,[2] in surveying practices in Western New York State, found that tests are generally administered by the classroom teachers. Belanger[3] found that in California secondary schools, tests were administered by specialists and teachers evenly.

Scoring tests.-- The literature points to the necessity of rapid and accurate scoring of the tests and statistical treatment of the results. Several writers consider the pros and cons of manual and machine scoring and present debatable points relative to scoring procedures. The entire subject of scoring is treated extensively by Traxler.[4] All

1/Noll, op. cit., p. 331.

2/Rosinski, op. cit.

3/Belanger, op. cit.

4/Arthur E. Traxler, "Administering and Scoring the Objective Test," Chapter X in Educational Measurement, E. F. Lindquist (Editor), American Council on Education, Washington, D. C., 1951.

writers emphasize the importance of exercising great care to insure the accuracy of the scores.

Several basic principles applying to the problem of test scoring are enumerated by the Kent Area Guidance Council[1] as follows:

"1. In the primary grades, consumable hand-scored booklets should normally be used.

2. At other levels, machine-scored self-marking answer sheets reduce scoring time and can be handled by clerical staff.

3. In some cases, the tester should be aware of the response to individual test items. In such situations, a consumable booklet should be hand-scored.

4. Test scorers often include teachers, counselors, psychologists, clerical staff, and sometimes students under supervision.

5. A system of spot checking should be established whether the tests are machine or hand-scored. Frequently every tenth or fifteenth paper is rechecked. Spot checking should include scoring, conversion or computation of score, and posting.

6. Students in group guidance programs may gain additional insight by scoring and profiling their own tests."

Except in large cities where there are scoring facilities available, hand-scoring is still the common procedure.[2] Ross and Stanley[3] indicate that a variety of personnel is used in this task and that hand-scoring is probably best done by the classroom teacher. Wrightstone, Justman, and Robbins[4] advocate this practice and note the ad-

---

1/Kent Area Study Council, op. cit., pp. 4-5.

2/Noll, op. cit., p. 337.

3/Ross and Stanley, op. cit., p. 320.

4/Wrightstone, Justman, and Robbins, op. cit., p. 68.

vantages of teacher-scoring by the following comment:

> "In most instances there are decided advantages in having the teacher score the test paper of pupils in his own class. Not only will the results be available at an early date, but the teacher will gain valuable information concerning errors and gain a better understanding of the meaning of the test scores."

Traxler[1] states that scoring should not be turned over to teachers without proper training or supervision. Ideally, teachers should not be required to score standardized tests at all. He feels that efficient objective scoring is a rapid clerical task and that local scoring can be carried on more effectively and at less actual cost by a specially trained clerical staff.[2] The following comment is made by Traxler:

> "The most common but least satisfactory use of personnel for this purpose is to have the teachers or counselors do the scoring. Staff members who have full teaching or counseling loads and who are given the work of scoring as an added duty can hardly be expected to be as efficient or accurate as a group of trained clerks. Moreover, the routine drudgery of scoring takes valuable time and energy which should be used in planning and carrying on instruction and guidance. It is sometimes assumed by administrators and school boards that the comparing of a pupil's objective responses with an answer key will, by means of some mysterious process, give a teacher or a counselor a diagnostic insight into the pupil's strengths and weaknesses. This could be a valid assumption only if a scorer of objective tests were obliged to read each question in order to evaluate each answer, but efficient scoring is not done in that way."

Wellman[3] advocates the use of separate answer sheets and machine

---

1/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 360.

2/Traxler, "Administering and Scoring the Objective Test," op. cit., pp. 396-398.

3/Frank E. Wellman, "Administration of the Testing Program," in Understanding Testing, United States Department of Health, Education, and Welfare, Office of Education, United States Government Printing Office, Washington, D. C., 1960, p. 20.

scoring for the fourth grade and up. Many other writers advocate this procedure and point to economy both in time and money, and a higher degree of confidence regarding the accuracy of results. Ahman and Glock[1] note the following advantages of machine-scoring:

"1. Accuracy is greater; machine scoring results in fewer errors.

2. It saves time for the busy teacher who can spend his time more profitably than doing clerical work.

3. Item-analysis data can be quickly prepared when needed."

The literature contains several varying views regarding the scoring procedure. In the interest of efficiency, it appears advisable to machine-score tests for the fourth grade and above. If the tests are hand-scored, accuracy in scoring cannot be taken for granted. The scorers must be taught how to score the papers and carefully be supervised in this process. Papers should be spot-checked for accuracy. The writings also favor the use of clerks rather than teachers for this task when feasible.

Survey studies conducted by North[2] and Rosinski[3] reveal that the majority of tests are scored by teachers. Belanger[4] reported that only 24 per cent of the schools used machine-scoring. Leonard[5] reported only one per cent of the systems making extensive use of

---

1/Ahman and Glock, op. cit., p. 491.

2/North, op. cit.

3/Rosinski, op. cit.

4/Belanger, op. cit.

5/Leonard, op. cit.

machine-scoring. Current practices thus indicate that the schools have considerable distance to go to reach the procedures advocated in the literature.

Financing the program.-- To a large extent, the comprehensiveness of a testing and evaluation program is dependent upon financial provisions made for its operation. This aspect of the program needs investigation and has received practically no attention from writers. The literature contains a dearth of references concerning this important element.

Lennon[1/] states that "as nearly as the publishing industry figures reveal, the average expenditure per elementary and secondary school pupil for testing materials is in the neighborhood of twenty cents--certainly not more than twenty-five cents." Boag[2/] reports that only seven cents per pupil per year was being spent for standardized tests.

From available figures, the per pupil expenditure for testing is exceedingly low when viewed in terms of the total per pupil educational expenditure.

5. Responsibility for Program Direction

Program organization.-- The administrative organization for carrying on a testing and evaluation program will differ considerably depend-

1/Roger T. Lennon, "Discussion of the School Administrators' Problems," in Invitational Conference on Testing Problems, 1957 Proceedings, Educational Testing Service, Princeton, New Jersey, 1958, p. 97.

2/Audrey K. Boag, "Standardized Tests: How, When, Why," Instructor (October, 1935), 65:24.

ing on the type and size of the school system. Certain basic principles of organization prevail in all situations. The following principles of organization are applicable to all schools:[1]

> "1. The program should be centrally coordinated with appropriate functions decentralized among the school staff.
>
> 2. Specific assignments and lines of responsibility should be defined as clearly as possible. Although individual staff members may have various responsibilities, it should be kept in mind that the teacher in the classroom is the focal point of the testing program.
>
> 3. The person assigned responsibility for the testing program should have appropriate training, adequate time, and designated authority to coordinate the development and operation of the program. His duties should be defined."

Program director.-- The problem of planning and organizing the details of any testing and evaluation program is a technical job requiring special skills and knowledge. Jordan,[2] King,[3] and Lass and Wrightstone[4] advocate that a responsible leader, specially trained in the measurement field, be charged with the responsibility of the total program. Other writers strongly advocate this procedure and further note that this individual be given free time in order to carry out this function. In the large system this person will undoubtedly devote full time to these responsibilities. In their studies of current practices,

---

1/Kent Area Study Council, op. cit., p. 3.

2/Jordan, op. cit., p. 68.

3/King, op. cit., p. 58.

4/Lass and Wrightstone, op. cit., p. 26.

Michaelis[1] and Michaelis and Howard[2] found that almost all the large school districts reported a full-time director of the program. Many varying titles were noted for this individual. In the small system, where it may not be feasible to employ full-time specialists, Durost[3] advocates that a teacher be given special training in the measurement field and then be allowed some released time to function as the specialist for the program.

The Kent Area Study Council[4] recommends that, regardless of title, the person charged with the direction of the program perform the following functions:

> "a. He should constantly evaluate new tests and should be prepared to recommend any changes in the program. It is desirable for him to work with a representative advisory committee to select and to evaluate tests periodically.
>
> b. He should assume responsibility for the mechanical and the administrative aspects of the group testing program: i.e., planning for, ordering, and distributing test materials; issuing test bulletins and announcements; scheduling test dates; checking receipt of test materials; and performing other details.
>
> c. He should be responsible for the statistical summaries of test results that are needed for teacher and administrative use.
>
> d. He should be responsible for test interpretations to the school staff and to the public.

1/Michaelis, op. cit., pp. 15-22.

2/John W. Michaelis and Charles Howard, "Current Practices in City School Systems in California," Journal of Educational Research (December, 1949), 43:250-260.

3/Durost, op. cit., p. 5.

4/Kent Area Study Council, op. cit.

e. He should, in cooperation with principals and supervisory personnel, make provision for the encouragement of planned teacher use of test results."

Coordinated endeavor.-- The literature advocates that a school-wide committee, representing the whole staff, be set up to plan and develop the program. Remmer and Gage[1] and Womer[2] stress the importance of this type of organization working closely with, and under the leadership and guidance of, the program director.

Writings clearly indicate that the testing and evaluation program will be received by the teachers to the extent to which each one knows his specific responsibility and function in the program.[3] This involves a clear delineation of duties and an understanding of the part that each member of the staff plays in the total program. The literature also discloses that the most successful program is one in which teachers participate in carrying out many of the functions. Detailed directions and procedures must therefore be made available to all participants.[4] This, together with other administrative aspects, is the responsibility of the director working with the staff committee. In the interest of efficiency and economy all organizational and administrative aspects of the program must be carefully organized. Good administrative procedure demands utmost staff cooperation and close coordination of all staff functionaires.

1/Remmer and Gage, op. cit., p. 78.

2/Frank B. Womer, "Initiating a Testing Program," Elementary School Journal (January, 1957), 57:195.

3/Durost, op. cit., p. 5.

4/Ross and Stanley, op. cit., pp. 227-228.

6. Personnel Orientation

Necessary measurement understandings for staff.-- Writers agree that the value and usefulness of a testing and evaluation program are strengthened if the problems being investigated are thoroughly understood by all participants. Since the efficacy of a test depends largely upon the skill with which the results are interpreted and used, all personnel using and interpreting these data must be adequately orientated.[1/] Cook[2/] reported certain knowledges that teachers should possess regarding tests. The following items selected from his list are pertinent to this problem:

"1. Be able to analyze standardized tests in the subject and the grade level of teaching.

2. Know proper procedure and ethics of test administration, reporting results, and using teacher-made and standardized tests.

3. Know how to compute percentile ranks, grade placement scores, and various scores and interpretations.

4. Know how to compute and interpret coefficients of reliability and validity.

5. Interpret test scores with reference to standard error of measures.

6. Construct and interpret individual and class educational profiles.

7. Interview parents in analyzing pupil results.

8. Know the relative value of various measures of predicting achievement in school subjects."

1/King, op. cit., p. 59.

2/Walter W. Cook, "What Teachers Should Know About Measurement," in Fifteenth Yearbook of the National Council on Measurements Used in Education, Educational Testing Service, Princeton, New Jersey, 1958, pp. 16-19.

Wilkinson[1] states that, "beyond the understanding of school policy concerning measurement and evaluation, the teacher should be guided toward an understanding of the purpose of educational tests and see them in relation to other facts involved in the teaching process." Smith and Tyler[2] point out that all teachers must develop the ability to put into effect the evidence gained about the student through the appraisal program. Unless the teacher is able to interpret and use these data effectively in improving the instruction and adjustment of individual boys and girls, the program will not accomplish its purpose.

Preservice preparation in measurements.-- The literature evidences some concern regarding the inadequate preparation and competence of teachers in measurement. Noll[3] surveyed the measurement courses in eighty colleges and concluded that teachers were not receiving adequate training in this field. This view is held by Coleman[4] and other writers. A recent analysis of requirements for teacher certification showed that only five states require a course in measurements.[5] The

1/D. H. Wilkinson, "Measurement and Evaluation: A Significant Area in Planning Programs of In-Service Education," The Bulletin of the National Association of Secondary School Principals (February, 1957), 41:90.

2/Smith and Tyler, op. cit., pp. 454-455.

3/Victor H. Noll, "Requirements in Educational Measurement for Prospective Teachers," School and Society (September 17, 1955), 82:88-90.

4/William Coleman, "Assisting Teachers in Using Test Results," Personnel and Guidance Journal (September, 1950), 36:12.

5/Victor H. Noll, Introduction to Educational Measurement, op. cit., p. ix.

lack of preparation or "educational lag" is not only the concern of writers but also of teachers. Wilkinson[1/] contends that there is evidence to support the statement that testing and evaluation are concerns of teachers with classroom experience and are ranked high by them when listing their problems. Mayo[2/] concludes that although there is a widespread interest and effort to educate test users in a better understanding of tests, not nearly enough is being done to educate prospective teachers in measurements.

Means for in-service training.-- The foregoing points to the need of a continuous in-service educational program for faculty members in the interpretation and use of test results. Wrightstone, Justman, and Robbins[3/] make the following comment concerning this fact:

> "The formulation of an evaluation program must provide opportunity for discussion and in-service education of teachers on newer trends in evaluation and measurement, especially the application of results of measurement to the instructional program. This is part of the responsibility of the principal or supervisor who is leading the program."

Traxler[4/] recommends that this program should center around the actual measurement and evaluation techniques used in the system.

Many means of conducting the in-service program are noted in the literature. Durost[5/] recommends short courses, the workshop on dis-

---

1/Wilkinson, op. cit., p. 94.

2/Mayo, op. cit., pp. 5-6.

3/Wrightstone, Justman, and Robbins, op. cit., p. 61.

4/Arthur E. Traxler, "Evaluation of Aptitude and Achievement in a Guidance Program," Educational and Psychological Measurements (Spring, 1946), 6:14.

5/Durost, op. cit., p. 5.

missed school time, regularly scheduled staff and faculty meeting, and the like. King[1] suggests short conferences with small groups dealing with the theory on special problems of the particular school system. Kvaraceus[2] advocates extension courses given in the school system under the direction of a University specialist in measurement and evaluation. These courses should be geared to meet aspects of measurement faced daily by the teacher and keyed to the investigation and solution of problems of local concern. Traxler[3] lists the following ways of educating the faculty in fundamentals of measurement:

"1. Have the testing program undertaken by the school explained fully and frankly by the school head.

2. Invite a test specialist who is not too technical to speak to the faculty and answer their questions.

3. Carry on a series of 'workshop' meetings centered around examination of specific tests and study of the practical problems of testing.

4. Send faculty members who are to have main responsibility for the testing program to do summer study in measurement and statistics.

5. Carry on case conferences with faculty members about individual students and to use test scores as one of the main kinds of data about the individual.

6. Make available and use as bases of discussion books and articles written especially to help schools do a better measurement job."

1/King, op. cit., p. 59.

2/W. C. Kvaraceus, In-Service Training in Educational Measurement, Test Service Bulletin No. 64, World Book Company, Tarrytown-on-Hudson, New York.

3/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 361.

Smallenberg[1] recommends grade level meetings and the maintenance of a central file or library of measuring devices and techniques.

In studies of current practices, Traxler[2] reported that 75 per cent of the large city systems have in-service programs in measurement, while North[3] reported 58 per cent of the systems have such programs in operation.

7. The Testing Program

Basic considerations.-- Instruction and guidance must be based upon a thorough pupil appraisal program extending from kindergarten through high school. The total appraisal program, while leaning heavily on testing, must make full use of cumulative records, teacher observations, school marks, conferences, and other nontesting devices. To be truly comprehensive, the program must appraise achievement in all areas, include all the pupils in the school system, and make use of the techniques which perform the most useful service.[4] It must provide the school with the following accurate information:[5] (1) careful measure-

1/Harry Smallenburg, "Developing a Program for Evaluation of Pupil Growth in the Burbank Schools," Educational Administration and Supervision (October, 1942), 28:552.

2/Arthur E. Traxler, "The Status of Measurement and Appraisal Programs of Large City School Systems," 1953 Achievement Testing Program in Independent Schools and Supplementary Studies, Educational Records Bureau, Bulletin No. 61, Educational Records Bureau, New York, 1953, p. 86.

3/North, op. cit., p. 90.

4/American Association of School Administrators, op. cit., p. 323.

5/California Test Bureau, How Tests Can Improve Your Schools, Educational Bulletin No. 1, California Test Bureau, Los Angeles, California, 1945.

ments of each pupil's learning abilities and disabilities, (2) semester or quarterly evaluation of each pupil's mastery of the basic learning skills, (3) measurement and analysis of each pupil's personality characteristics and adjustment, and (4) an inventory of interests, hobbies, and recreations of each pupil.

The literature reveals that the test approach continues to be the most widely used method for pupil appraisal, and represents the most practical means of obtaining valid and reliable information about the pupil. There are many indications noted in the literature that more and more emphases are being placed upon this approach. Among them are the provisions for guidance, counseling, and testing in the National Defense Act of 1958; the increased use and influence of the College Entrance Examination Board tests and the National Merit Scholarship Qualifying Test; and the widespread expansions of state, regional, and national testing programs. Writings reveal that testing is usually the beginning phase of any developing appraisal program and can be broadened later to include other techniques of appraisal. The testing program thus is given prime consideration in this review.

The literature cites that the testing program is an integral and continuing part of the instructional program itself and includes the ordinary informal tests and examinations made by the teacher, locally constructed tests by staff for district-wide use, and standardized tests. All writings reveal that there is a legitimate place for all kinds of tests, but no one test is equally good for all purposes.[1/]

1/Ross and Stanley, op. cit., p. 217.

Ross and Stanley[1] note that the particular combination of measuring techniques required in any given situation will depend upon the specific purpose to be served. The literature reveals that local school conditions will dictate the extent of testing. Since the schools aim to educate the child at his normal learning rate, it is necessary to obtain complete data about the individual pupil and his school progress. Writers agree that the most effective way of accomplishing this is by means of an integrated and comprehensive testing program utilizing a variety of tests.

This review is concerned with major aspects of the system-wide testing program and excludes practices carried on by individual classroom teachers. The reader should recognize, however, that standardized tests merely supplement rather than supplant the ordinary tests made by the classroom teacher.[2] Lennon[3] emphasizes this axiom in discussing achievement tests when he states, "Standardized achievement testing should by no means be thought of as a substitute for the use of teacher-made examinations, but as complementary to their use." He notes that standardized achievement tests provide the teacher with an independent yardstick that is less likely to reflect his own special bias in course coverage and provides an additional frame of reference for evaluating the performance of the class. He further states that the optimum approach for comprehensive evaluation is a combination of

---

1/Ross and Stanley, op. cit., p. 216.

2/Ibid., p. 210.

3/Lennon, Testing in the Secondary School, op. cit., p. 1.

teacher-made tests and carefully selected standardized tests as well as other methods of appraisal. Although standardized tests represent the major concern of this study, the means of evaluating the results of instruction are not limited solely to these tests. The literature clearly indicates that no one type of test by itself represents the means of obtaining a comprehensive appraisal of student achievement.

The objective of the testing program is to furnish accurate and dependable information about important characteristics of pupils and the success being realized in learning. In order to attain this, the testing program should provide the following:[1/]

> "1. Provide teachers and administrators with a measure of the effectiveness of the instructional program.
>
> 2. Provide teachers and parents with the machinery for measuring the educational progress and mental growth of the pupils.
>
> 3. Provide information that will increase the effectiveness of educational and vocational counseling.
>
> 4. And most important--provide a yardstick for comparing a child's achievement with his aptitudes and abilities."

The tests employed in the program should describe collectively all important aspects of the pupil's educational development. The information thus collected may be put to many uses. These are reviewed in a later section of this study. Recent writings urge the schools to increase their efforts toward identifying, as early as possible, the gifted individual, with the aim of developing his talents to the max-

---

1/H. F. Grimes, "A Testing Program's Excuse for Being," Clearing House (September, 1954), 1:33-34.

imum.[1] Full utilization of a comprehensive testing program can aid in accomplishing this task.[2]

Writers cite that the testing program provides the basic means of securing information for curriculum evaluation and pupil guidance. It must be recognized that tests are means to an end--purposeful use--and not an end in themselves. The obvious limitations of tests must also be recognized in providing data.[3] To attain maximum effectiveness, the program should possess the following features:[4]

> "(1) Comprehensive and well-balanced in terms of both the learner and the curriculum
>
> (2) Continuous and well-articulated from kindergarten through college
>
> (3) Functional and practical
>
> (4) Use integrated and scientific measuring instruments."

Scope.-- The literature reveals a large number and variety of tests available for various purposes. The Fourth Mental Measurements Yearbook[5] lists 64 classifications of tests under the following 13 headings: (1) Achievement Batteries, (2) Character and Personality,

---

1/John L. Holmes, Guidance Testing and the Identification of Pupil Characteristics, Educational Bulletin No. 21, California Test Bureau, Los Angeles, California, 1958, p. 15.

2/Roger T. Lennon and Harold F. Bligh, Finding Mathematics and Science Talent in the Junior High Schools, Test Service Bulletin No. 91, World Book Company, Tarrytown-on-Hudson, New York, p. 1.

3/American Association of School Administrators, op. cit., p. 330.

4/Clark, op. cit., p. 1.

5/Oscar K. Buros (Editor), The Fourth Mental Measurements Yearbook, Gryphon Press, Highland Park, New Jersey, 1953, p. vii.

(3) English, (4) Fine Arts, (5) Foreign Languages, (6) Intelligence, (7) Mathematics, (8) Miscellaneous--including business education, handwriting, health, industrial arts, etc., (9) Reading, (10) Science, (11) Sensory-Motor, (12) Social Studies, and (13) Vocations. Traxler,[1] in giving a practical discussion of the planning and administration of the testing program, divides tests into two broad categories: (1) group tests of intelligence, and achievement tests in the major subject-matter areas; and (2) individual intelligence tests, special aptitude tests, personality tests, and tests of vocational interest. The Kent Area Study Council[2] lists tests in the following categories: School Ability Tests, Achievement Tests, Aptitude Tests, Interest Inventories, Scholarship Tests, Personal and Social Maturity Inventories, Readiness and Maturation, Physical Health, and Other Tests. Holmes[3] states that a basic standardized testing program should consist of tests in the following areas: (1) intelligence or mental maturity, (2) achievement in basic skills, (3) interests, (4) special abilities or aptitudes, (5) personal and social adjustment, and (6) attitudes and study skills.

Any classification of the wide variety of standardized tests is purely arbitrary and made, in many instances, merely for convenience. Writers state that the most useful and practical classification of tests is based upon the purpose of the test. Using this procedure,

---

1/Arthur E. Traxler, "Planning and Administering a Testing Program," School Review (April, 1940), 48:253-267.

2/Kent Area Study Council, op. cit., pp. 7-10.

3/Holmes, op. cit., pp. 2-3.

this study considers the following four main types of standardized measures:

1. Aptitudes
   a. General, including mental ability and scholastic aptitude
   b. Special, including mechanical aptitude, clerical aptitude, musical aptitude, differential aptitude batteries, etc.
   c. Readiness
2. Achievement in subject fields
   a. Survey
   b. Diagnostic
3. Interests
   a. General
   b. Educational
   c. Vocational
4. Personal qualities or personal-social adjustment.

The literature reveals that aptitude and achievement are the hard core of the testing program, with measurement of interest and personality somewhat less dependable. This review considers all four areas. A detailed review and presentation of specific tests involving a comprehensive enumeration of test titles and other information is beyond the scope of this study. This review presents essential features and considerations in these four main areas of the testing program.

8. Appraising Aptitudes

Measuring academic aptitude or intelligence.-- The appraisal of academic aptitude or intelligence is a necessary part of any testing

program. A test of mental maturity is the best general predictor of the educational level at which each pupil is able to achieve.[1/] The two main patterns of academic aptitude tests are group tests and individual tests. In discussing these tests, Thorndike and Hagen[2/] note that group tests are much more economical to use and are satisfactory for many purposes with normal groups. They also state that individual tests have a number of advantages and are particularly useful with (1) young children, (2) emotionally disturbed pupils, and (3) pupils with special educational disabilities.

Current practices reveal that most schools administer one or more group tests to pupils during their school career. Many schools obtain the results of several such tests for each individual pupil. These tests differ greatly in content, presenting a variety of problem situations, including verbal, nonverbal, and performance items. Some tests give single scores, while other tests of the factor-type variety give multiple scores.[3/] In discussing test varieties, Durost makes this comment pertaining to these variances:

> "Some group intelligence tests are of the self-administering type, while others maintain the pattern of separate sub-tests yielding separate measures which are combined into some kind of average. Some tests emphasize the speed factor, while others have generous time limits."

1/Holmes, op. cit., p. 2.

2/Robert L. Thorndike and Elizabeth Hagen, Measurement and Evaluation in Psychology and Education, John Wiley and Sons, Inc., New York, 1955, pp. 240-241.

3/Durost, op. cit., p. 2.

The variety of available intelligence tests dictates the necessity of the test administrator being cognizant of the various types of instruments and their use in a specific school situation. The literature treats these problems.

Durost[1] states that for general school use the single score group test of verbal mental ability is the most useful instrument and should be considered basic to the minimal pupil-centered testing program. He feels that this measure may be considered a scholastic aptitude test rather than a measure of the ability of individuals to apply their intelligence to the solution of all types of life situations. Thorndike and Hagen[2] corroborate this view by indicating that in the usual school situation, aptitude for learning can be tested more efficiently by tasks that involve language than by those that do not. Durost[3] recommends that the nonlanguage test or the individual test be considered as supplementary and not as a measure to be applied generally to all pupils. In discussing the use of these tests, he makes the following comment:

> "The non-language test of mental ability may be used to supplement the verbal test, to spot pupils whose language abilities have not been developed to the point where they can make a fair showing on a verbal test of intelligence. Children with specific disability in reading or who are hard of hearing (whose verbal environment has, therefore, been limited) are illustrations.
>
> The non-language test or the individual test may be considered for the purposes of our minimal pupil-centered program

1/Durost, op. cit., p. 2.

2/Thorndike and Hagen, op. cit., p. 215.

3/Durost, op. cit., p. 2.

> as supplementary tests to be used only as indicated and not as instruments to be applied generally to all pupils."

King[1] states that a minimum program for testing intelligence will probably include one test which affords one score, a composite of three intelligences--verbal, reasoning, and numerical. Interpreting this test offers several difficulties since the contributions of these components are unknown. He notes this difficulty:

> "If the test program is confined to a single measure, the test of general intelligence is undoubtedly the answer. However, the educator must consider this as only a beginning, realizing the restriction that a single-score intelligence test places on an adequate program of guidance and supplementing this score by other information in planning the education of a student."

The literature discloses the existence of a number of relatively independent factors in intelligence such as memory, spacial relations, logical reasoning, numerical reasoning, and verbal abilities.[2] It notes that this discovery has resulted in group-factor tests, lying perhaps midway between general intelligence and special aptitude tests in specificity, which attempt to measure the more specific mental abilities or factors. Typical of these types of tests are the Chicago Test of Primary Mental Abilities[3] and the California Test of Mental Matur-

---

1/King, op. cit., p. 34.

2/Ernest W. Tiegs, The Proper Use of Intelligence Tests, Educational Bulletin No. 4, California Test Bureau, Los Angeles, California, 1951, p. 2.

3/Chicago Test of Primary Mental Abilities, by L. L. Thurston and Thelma Gwinn Thurston, Science Research Associates, Chicago, Illinois, 1947.

ity.[1] These tests give from two to six subtests in addition to the total score. Durost[2] notes that these tests do not vary greatly in type of item from the pattern established over the past 30 years, but do recognize each ability as a more or less independent mental attribute susceptible to independent measurement. Greene, Jorgensen, and Gerberich[3] caution that scores from their use should be interpreted with care until such time as their validities for various purposes have become well established. In spite of these cautions, writings indicate that these instruments present a more comprehensive picture of the individual than the one-score test.

Writers cite that schools interested chiefly in an over-all estimate of general intelligence will find the single-score test satisfactory. Such tests are also cheaper and require less time to administer. The factor-type test, yielding a profile rather than a global score will offer a fuller understanding of the pupil's special capacities and potentialities and will be more meaningful and useful for education and guidance. Strang[4] corroborates this point of view and states that language and nonlanguage scores represent significant types

---

1/California Test of Mental Maturity, by Elizabeth T. Sullivan, Willis W. Clark, and Ernest W. Tiegs, California Test Bureau, Los Angeles, California, 1936-1951.

2/Durost, op. cit., p. 2.

3/Greene, Jorgensen, and Gerberich, op. cit., p. 273.

4/Ruth Strang, "Relationships Between Certain Aspects of Intelligence and Certain Aspects of Reading," Educational and Psychological Measurement (Winter, 1943), 3:355-360.

of abilities. Traxler[1] concludes that, other things being equal, a test yielding, in addition to a gross score, separate scores for at least linguistic aptitude and quantitative aptitude is preferred to a single score. The Kent Area Study Council[2] states that the determining factor pertaining to the use of the global or factor-type test is the manner in which the results are to be used. They present the following usage guide for these tests:

"The single-score or global test might be used:

1. Where simplicity of administration, scoring, etc., is of unusual importance.

2. When it is important that the testing time be short.

3. When the highest possible correlation between a single test score and some index of academic achievement is desired.

The factor-type would be needed:

1. When it has been demonstrated that certain of the subtest scores have an exceptional relation with success in certain courses and can therefore be used for course selection.

2. When skilled teachers can use subtest patterns to suggest ways in which instruction can be modified to fit varying types of learning needs.

3. When a test which minimizes the influence of reading skill is desired.

4. When the tests are to be interpreted by a clinically trained person who can draw more inferences from them than from the global test.

5. Under any conditions in which a more highly specific

1/Traxler, Techniques of Guidance, op. cit., p. 21.

2/Kent Area Study Council, op. cit., pp. 7-8.

picture of pupils' mental strengths and weaknesses is desired than can be obtained from the global test."

Bacon[1] makes the following comment and sounds a warning about certain practices in interpreting intelligence test scores:

> "In thinking of distortions it should be indicated too, that an imbalance can be created by too large an emphasis upon testing for general intelligence. Far too much attention, for example, has been placed on the single I.Q. score. The I.Q. is not constant for all pupils. Pupils themselves change; and a single test score may be more than inaccurate. It may be harmfully misleading. It is now well established that an interpretation of mental or academic ability based on less than three I.Q. scores offers too much uncertainty. Moreover, a series of scores do not constitute, in themselves, sufficient evidence, as general intelligence is made up of a number of component abilities which are basic to the learning process although they are more or less independent of one another."

Durost[2] emphasizes the unreliability of a single test score, stating that variations will occur in results from one testing to another due to the operation of chance factors. It is thus strongly recommended that intelligence tests be given with reasonable frequency in order that the pupil will not be overrated or underrated by one testing. Stoddard[3] concurs with this view, stating, "Intelligence is analogous to health; any estimate of it should be rechecked close to the making of an important decision." Ross and Stanley[4] indicate that although it is not necessary to give pupils intelligence tests every

---

1/Francis L. Bacon, "Using Tests in the Secondary Schools," The Bulletin of the National Association of Secondary School Principals (April, 1943), 33:45.

2/Durost, op. cit., p. 2.

3/George D. Stoddard, The Meaning of Intelligence, The Macmillan Company, New York, 1943, p. 94.

4/Ross and Stanley, op. cit., p. 225.

year, fluctuations on group tests warrant giving such tests more than once. They further reveal that these fluctuations are most serious in the primary grades. Writers generally agree that intelligence tests should be administered at transitional points in the pupil's school history. The New York State Department of Education[1/] recommends administering a test in the primary grades and then repeating at least once at each of the intermediate, junior high school, and senior high school levels. Jordan[2/] strongly urges that an accurate appraisal of the pupil's intelligence be made in the first grade. He feels that this measure, together with reading readiness test results, is necessary for predicting subsequent success in reading. The advantage of giving intelligence tests about the same time as achievement tests in order to increase the comparability of the intelligence and achievement test results is also advocated by Durost.[3/]

As a result of a study of the literature, it is concluded that intelligence tests should be administered at least four or five times during the pupil's school career in grades 1, 3, 6, 8, and 10 or 11. It is further desirable to have retested, in some cases by individual tests, the following students: those who test below 80, those who test above 130, and those whose scores are out of line with the judgment of teacher.

1/New York State Education Department, op. cit., p. 15.

2/Jordan, op. cit., p. 391.

3/Durost, op. cit., p. 2.

The limitation of intelligence tests should be recognized and caution attached to the careless and indiscriminate use of the results. Greene, Jorgensen, and Gerberich[1] warn about the danger that may arise through giving publicity to the results and through careless use of such tests. The literature also reveals that physical illness and emotional blocks may seriously affect the results. The words and concepts used in some tests may also act as a barrier to clear communication. All these factors must be considered when using and interpreting intelligence tests.

Surveys reviewed in this study[2] reveal that practically all school districts administer group intelligence tests at regular intervals. Michaelis and Howard[3] reported that the majority of schools give tests at five different levels (grades 1, 3, 5, 7 or 8, and 10). Traxler[4] observed that the over-all intelligence test continued to be widely used in the regular program, but noted a trend toward the tests providing differential measures. Connecticut[5] reported that the large majority of districts administered three or more tests annually. Leonard and Tucker[6] reported 87 per cent of the schools used intelli-

---

1/Greene, Jorgensen, and Gerberich, op. cit., p. 271.

2/The reader is referred to Chapter II, supra.

3/Michaelis and Howard, op. cit., pp. 250-260.

4/Traxler, "The Status of Measurement and Appraisal Programs of Large City School Systems," op. cit., pp. 75-86.

5/Connecticut State Department of Education, Testing Programs in the Public Elementary and Secondary Schools of Connecticut, Bureau of Research and Statistics, Connecticut State Department of Education, Hartford, 1955.

6/Leonard and Tucker, op. cit.

gence tests, with 20 per cent administering three or more annually.

Measuring special aptitudes.-- Many pupils manifest special abilities and talents. Test authors have attempted to group together basic intelligence related to these abilities and talents into special aptitude tests and, more recently, aptitude batteries. Typical of the former are clerical tests which draw their content from the perceptual, numerical, and verbal areas, and the mechanical tests which sample visualization, coordination, perceptual and mechanical information. Typical of the latter type is the differential aptitude battery which is designed chiefly for use at the secondary level. Many other batteries which sample aspects of basic intelligence and measure skill and specific aspects of basic abilities have been devised for prognosis in such fields as teaching, music, art, scientific, social intelligence, medical, nursing, engineering, and physical science.[1/]

Although aptitude tests represent a major method of evaluation and diagnosis of aptitudes, they are only one means of obtaining data required for guiding the pupil. Wrightstone, Justman, and Robbins[2/] note that such methods as the interview, rating scale, anecdotal record, and cumulative record should also be used in arriving at the total picture. Ahman and Glock[3/] support this view with the following comment:

> "It should be emphasized that testing is only one means of determining the aptitudes of pupils. The results of any one test should never be the sole basis for judging ability.

1/Donald E. Super, Appraising Vocational Fitness, Harper and Brothers, New York, 1949, pp. 60-65.

2/Wrightstone, Justman, and Robbins, op. cit., p. 315.

3/Ahman and Glock, op. cit., p. 421.

> Observation of the pupil's behavior, analysis of his learning procedures and products, and the direction of his interests and attitudes must also be considered in any evaluation. However, aptitude tests will often reveal abilities that have been overlooked by the classroom teacher. They provide evidence that should be carefully considered in the total evaluation picture."

Writings reveal that aptitude tests are needed as a supplement to other not too effective methods of analyzing human abilities, interests, and achievements. Traxler[1/] states that these measures are less satisfactory for their purpose than general academic aptitude tests and recommends that subjective criteria be considered in making a valued judgment pertaining to a pupil. He further suggests that aptitude tests have a place in the guidance program if used with limitations. Greene, Jorgensen, and Gerberich[2/] indicate that major uses for such measures appear to be in the areas of individual diagnosis and vocational and educational guidance.

Super[3/] states that aptitude tests are dependent upon the following assumptions: (1) individuals differ in the extent to which they possess any given aptitude; (2) there are a number of special aptitudes; and (3) there are important differences in the amounts of these various aptitudes possessed by given individuals. Knowledge of these abilities can aid in educational and vocational guidance and supplement general intelligence measures which provide only a limited amount of

---

1/Traxler, Techniques of Guidance, op. cit., p. 49.

2/Greene, Jorgensen, and Gerberich, op. cit., p. 274.

3/Donald E. Super, The Place of Aptitude Testing in the Public Schools, Test Service Bulletin No. 49, World Book Company, Tarrytown-on-Hudson, New York, 1943, p. 2.

information about an individual. In attempting to assess and diagnose aptitudes, it should be recognized that most tests are limited to measurement of the intellectual and manipulative abilities which correspond with achievement in a study or work situation.[1/] In discussing this point, Bingham[2/] makes the following comment:

> "Aptitude tests do not directly measure future accomplishment. They make no such pretense. They measure present performance. Then, in so far as behavior, past or present, is known to be symptomatic of future potentialities, the estimate is necessarily in terms of probabilities only."

The writings note that successful achievement in a total situation involves many factors including attitudes, interest, and general adaptability. These are not measured in most tests and must be considered in test interpretation. Ahman and Glock[3/] discuss these factors in singling out mechanical aptitude tests:

> "Single tests used to measure mechanical aptitude do not sample all of these factors. Also, jobs requiring mechanical skill differ in their demands of these specific aptitudes. Although certain tasks require a great deal of manual dexterity, others demand perceptual and spatial abilities. This explains in some degree why in general the existing tests of mechanical aptitude cannot claim high predictive validity. The factors included in the test as well as on the job must both be known. It is important that other means of evaluation, such as observation of performance in shops and a careful study of the record of experience, be utilized in counseling pupils on matters pertaining to mechanical aptitude."

This is equally true in other special aptitude tests. Caution must

---

1/Wrightstone, Justman, and Robbins, op. cit., p. 316.

2/Walter Van Dyke Bingham, Aptitude and Aptitude Testing, Harper and Brothers, New York, 1937, p. 22.

3/Ahman and Glock, op. cit., p. 406.

thus be used in test interpretation. Super[1] succinctly emphasizes this point in the following comment:

> "Such uses of tests imply the existence of two basic conditions: tests which are thoroughly standardized, and test users who know their tools. To administer and to score most tests is relatively easy. To interpret them wisely requires great skill, considerably specialized knowledge, and profound wisdom ripened by experience."

The literature reveals that special aptitude tests pose some difficulty in use since the standardization population is different for each test and the units of measurement are not comparable. This will not allow the construction of a meaningful profile for each examinee. Authors have solved this problem in recent years by constructing batteries of tests all standardized on the same population. Such a battery is the Differential Aptitude Tests.[2] These batteries afford a comparable measure of the pupil's important cognitive, perceptual, and sensorimotor aptitudes, and can result in a profile for each student facilitating both counseling and instruction.[3] Writings indicate that a differentiation must be made between aptitude tests designed to predict success in single occupations and those made up of batteries intended to be valid for a variety of occupations and standardized on the same population.

North[4] reports that counselors are making increased use of multi-

---

1/Super, The Place of Aptitude Testing in the Public Schools, op. cit., p. 10.

2/G. K. Bennett, H. G. Seashore, and A. G. Wesman, Differential Aptitude Tests, The Psychological Corporation, New York, 1947.

3/Ahman and Glock, op. cit., p. 411.

4/Robert D. North, "The Use of Multi-Factor Aptitude Tests in School

factor tests for differential predictions of academic and vocational success. Stanley[1/] notes that the traditional group intelligence measures are being threatened by differential aptitude batteries. The general intelligence test, however, is still useful in counseling at the elementary and high school levels, where it can predict scholastic achievement with a fair degree of success. Traxler[2/] states that the multiscore academic aptitude test may be used to predict success in many areas and makes the following comment regarding its use:

> "For the purpose of predicting success in the academic subjects, a test which provides verbal and numerical scores is a happy compromise between the need for valid measurement of aptitude and the desire to base the appraisal upon a test which can be given and scored within a reasonable time. Beyond this point the law of diminishing returns sets in and the increased predictive value doesn't justify the added cost."

The literature indicates that the multifactor test will become more useful as the reliability, validity, norms, and the framework of scores are better determined. At present, it recommends the use of multifactor test results only in conjunction with other data of well-established validity.

The staff of the Psychological Corporation[3/] recommends that an

---

Counseling," Proceedings of the 1955 Invitational Conference on Testing Problems, Educational Testing Service, Princeton, New Jersey, 1956, pp. 11-15.

1/Julian C. Stanley, Jr., "The Use of Multi-Factor Aptitude Tests in School Counseling," Review of Educational Research (February, 1953), 23:11-32.

2/Traxler, "Evaluation of Aptitude and Achievement in a Guidance Program," op. cit., p. 47.

3/The Psychological Corporation, Human Resources and the Aptitude Inventory, Test Service Bulletin No. 4, The Psychological Corporation, New York, May, 1951, p. 9.

aptitude inventory be given annually to all eighth and ninth grade pupils to assist in curriculum planning, in identification of individual pupil patterns of ability, and in a positive program of personal counseling. The California Test Bureau[1] recommends that aptitude tests be given whenever pupils have an option of selecting courses or when they encounter difficulties which appear to be due to lack of specific ability or aptitude. Super[2] suggests that these measures "may well be administered at different ages and stages, as life's decisions make those data desirable."

Froehlich and Benson[3] recommend that special aptitude tests are not necessary for all pupils and that group administration is justified only in the mechanical and clerical fields. Durost[4] suggests that the number of aptitude tests usable is rather limited and those commonly used are not too highly satisfactory. He restricts the field to music and art and perhaps one or two other areas. It is further noted that tests of musical and mechanical abilities are useful mainly in identifying those who are markedly deficient and not those who are

---

1/California Test Bureau, How to Select Tests, Educational Bulletin No. 2, California Test Bureau, Los Angeles, California, 1945, p. 2.

2/Super, The Place of Aptitude Testing in the Public Schools, op. cit., p. 1.

3/Froehlich and Benson, op. cit., p. 77.

4/Walter N. Durost, Tests and the Junior High School Guidance Counselor, Test Service Notebook No. 2, World Book Company, Tarrytown-on-Hudson, New York, pp. 4-5.

gifted.[1]

The literature reveals several varying opinions regarding the use of aptitude tests. The issue appears to be whether the general mental ability test provides all the information required for adequate guidance, or whether additional information provided by multifactor or special aptitude tests is needed. Writers agree that the answer lies in the function to be served. A functional program should make use of both mental ability and aptitude tests. Aptitude tests may well be used to supplement mental ability measures when the pupil enters high school and shortly before he leaves high school to enter a college or vocational field. Of the various measures available, writers recommend the use of the multiscore battery. In addition to these measures, they cite that subjective methods of analyzing human abilities and achievement should be brought into play in order to obtain a complete picture of the individual.

Surveys of current practices reviewed in this study[2] reveal that, in the majority of cases, aptitude tests are available and used by schools for special needs or purposes at the secondary level. Belanger[3] and Larson and McCreary[4] found that aptitude tests were used by one

---

1/Educational Policies Commission of the National Education Association, Education of the Gifted, The National Educational Association, Washington, D. C., 1950, Chap. III.

2/The reader is referred to Chap. II, supra.

3/Belanger, op. cit., pp. 108-111.

4/Carl A. Larson and William H. McCreary, "Testing Programs and Practices in California Public Secondary Schools," California Journal of Secondary Education (November, 1956), 31:389-402.

half of the school districts. Findings indicate that schools are making greater use of differential measures. Larson and McCreary,[1/] in particular, reported that the Differential Aptitude Tests had far greater use than all other aptitude tests. Limited surveys made in the field show that schools make limited use of aptitude tests as a supplement to the regular intelligence and achievement testing program.

9. Measuring Achievement

Overview.-- One of the major responsibilities of the schools is to teach reading, mathematics, mastery of all the language arts, and basic skills and understandings in many subsidiary subject content areas. An organized and systematic achievement testing program is invaluable in assuring the pupil's performance or achievement in these areas. The literature indicates that, with the possible exception of mental maturity, achievement is the most important area of appraisal in an evaluation program. Well-constructed standardized achievement tests serve many purposes. Lennon[2/] notes that they may be used for improving instruction, for diagnosis of individual achievement, and for consideration of curriculum revision. Holmes[3/] notes the following purposes: (1) to provide a picture of class and individual achievement levels that is necessary to individualize instruction, (2) to aid in identification of class and individual weaknesses, (3) to aid in

1/Larson and McCreary, op. cit.

2/Lennon, Testing in the Secondary School, op. cit., p. 2.

3/Holmes, op. cit., p. 3.

prognosing future success in a given subject, and (4) to aid in identifying the more able pupil early in his schooling. Traxler[1/] makes the following comment in discussing the import of achievement testing:

> "Achievement is probably, with the exception of general scholastic aptitude, the most important type of appraisal for a guidance program. Scores on achievement tests are excellent bases for the prediction of the future educational success of individuals in the subjects covered by the tests, afford very helpful clues for purposes of vocational guidance, and are significantly correlated with aptitude and with interests. Thus, tests of achievement help to provide information concerning the general academic aptitude and the functioning interests of each individual, and probably should form the core of the systematic testing program of every school which hopes to do a thorough and objective job of guidance."

There are many standardized achievement and diagnostic tests available. They attempt to measure a wide variety of areas including such skills as critical thinking, reading, arithmetic, listening, and study skills. There are also tests in a wide variety of subject areas. The choice of an appropriate test merits serious consideration. Although the specifics of test selection were treated in Chapter III, Durost[2/] notes certain factors bearing on content which are somewhat unique to the standardized achievement test:

> "1. Were the items selected on the basis of a careful examination of the typical curriculum in each subject-matter area?
>
> 2. Were the retained items in the final test refined by adequate tryout and subsequent statistical treatment?
>
> 3. Are several forms available and are these forms known

1/Traxler, Techniques of Guidance, op. cit., p. 68.

2/Durost, What Constitutes a Minimal Testing Program for Elementary and Junior High School, op. cit., p. 4.

to be equal in difficulty?

4. Are various types of norms provided and are these norms based upon an adequate and representative population?

5. Does the publisher provide adequate manuals, keys, class records, etc., for utilizing the test results?

6. Does the test suit reasonably well the content of instruction in the local situation, making it practical to compare the results of local instruction with the norms provided?"

Jordan[1] cites that the standardized testing program usually begins with a test battery, proceeds with a more complete coverage of a single area, and then may include diagnostic test coverage. In discussing the essentials of a minimal achievement testing, Durost[2] makes the following comment:

> "Any minimal program of pupil-centered testing should include achievement tests in the basic skills. Most schools feel that their basic responsibility is the teaching of reading, arithmetic, spelling, and mastery of the English language. Tests in the information areas and tests of other subsidiary skills are important too, but are secondary to tests of basic skills."

In the general area of standardized achievement testing, the Commission on American School Curriculum[3] notes that "conspicuous progress has been made in securing objectivity and validity in measuring achievement in such basic abilities as reading, arithmetic, and correctness of written English and in measuring specific items of knowledge." The

---

1/Jordan, op. cit., p. 114.

2/Durost, What Constitutes a Minimal Testing Program for Elementary and Junior High School, op. cit., p. 4.

3/American Association of School Administrators, op. cit., p. 324.

literature reveals, however, that only limited confidence may be placed in standardized measures in such areas as music, social adjustment, and the visual arts.

The matter of content validity also receives considerable emphasis. Wrightstone, Justman, and Robbins[1] cite that standardized subject-matter tests may not measure fairly the objectives of the curriculum in a specific school. The wide variation of the social studies content from community to community is cited as an example. Tests of general educational development do not present this problem since they purport to measure the ability of the pupil to apply facts he has learned rather than the specific content of the local school's curriculum. The literature reveals that schools are making greater use of this type of measure in an attempt to assess the depth and power of functional learning skills in a continuous way from the elementary level through high school. The STEP series[2] is an example of this type of measure.

Standardized achievement tests represent only one means of evaluating pupil achievement. Wrightstone[3] succinctly points this out in the following comment:

> "Achievement should be evaluated not only by standardized tests and techniques, but also by informal teacher-made tests and techniques as well as by the pupils' self-evaluation.

1/Wrightstone, Justman, and Robbins, op. cit., pp. 269-270.

2/Sequential Tests of Educational Progress, Cooperative Test Division, Educational Testing Service, Princeton, New Jersey, 1958.

3/J. Wayne Wrightstone, "Techniques of Appraisal," in Readings in General Education, Will S. Gray (Editor), American Council on Education, Washington, D. C., 1940, p. 253.

> Standardized tests are valuable for periodic appraisal of the growth and progress of pupils."

The selection of a specific type of achievement test is dependent upon the purpose it is to serve. Whether the school uses separate subject matter tests, a standardized achievement battery, a general educational development test, or a combination of various measures will depend upon the philosophy of the school and the aim of the testing program. Regardless of the type of achievement test used, the administration should not lose sight of the major purpose of the standardized achievement test which is stated by Lindquist:[1]

> ". . . the major purpose of a comprehensive standardized achievement test is to describe reliably and in comparable terms how far each pupil has progressed (not how far he should have progressed) in his development of each of the skills tested. In other words, its purpose is to enable the teachers to construct an educational 'profile' for the pupil that will indicate his true relative status in each of the areas measured."

Measuring achievement in the elementary grades.-- The literature advocates achievement tests in the basic skills. Noll[2] recommends the use of survey batteries when an over-all measure of achievement in the common branches or subjects is desired for the purpose of grade placement, promotion, or grouping.[3] Wrightstone, Justman, and Robbins also cite these batteries as valuable aids in evaluation of pupil

---

1/E. F. Lindquist, "Standardized Achievement Tests and Their Relation to Curriculum Content," in Appraising the Elementary-School Program, Sixteenth Yearbook, Department of Elementary School Principals, National Education Association, Vol. XVI, No. 6, July, 1937, Washington, D. C., p. 481.

2/Noll, Introduction to Educational Measurement, op. cit., p. 156.

3/Wrightstone, Justman, and Robbins, op. cit., p. 245.

growth. Since scores are comparable from subject to subject and among different forms, they note that such batteries permit an analysis of the pupil's strength and weakness in the basic skills and provide a profile of the progress of the pupil. The major batteries do not lack comprehensiveness. A study of the commonly used batteries reveals that they sample most of the more formal, defined outcomes of the elementary school. A review of the batteries reveals the following achievement areas being measured:[1] reading comprehension, vocabulary, fundamentals of arithmetic, reasoning, language, literature, social studies, science, spelling, and study skills.

Durost[2] recommends that "ideally, achievement tests should be given annually in every grade; in a minimal program a standardized battery should be given at least each time the curriculum changes noticeably." This implies at the end of the third grade, sixth grade, and possibly grade eight. Traxler[3] notes that achievement tests are usually either one of two types: (1) tests in which one battery serves throughout the whole range of the elementary grades; and (2) tests consisting of overlapping batteries for different grade levels. He does not feel that the one battery type is too well suited to the lowest and highest grades that it is designed to serve, and also questions the equating of the batteries of some of the overlapping battery-type tests.

1/Ahman and Glock, op. cit., p. 361.

2/Durost, What Constitutes a Minimal Testing Program for Elementary and Junior High School, op. cit., p. 4.

3/Traxler, Techniques of Guidance, op. cit., pp. 68-69.

Authors generally agree that appraisal of reading must receive top priority and that next to an intelligence test the reading test has the most general usefulness. Traxler[1/] states that no conclusions regarding intelligence, based on the usual group test of mental maturity, are valid unless the reading ability of the pupil is known. Martin[2/] corroborates this viewpoint and recommends that the elementary school have a continual reading testing program, including reading readiness, survey-type tests, diagnostic reading tests, measurements of eye movements, use of the opthalmograph and the metronoscope, and instruments for the identification of reading difficulties. Tiegs[3/] makes the following comment pertaining to the reading appraisal program:

> "Tests of reading readiness, mental maturity, and reading vocabulary and comprehension are essential aids in the development of competency in reading as diagnostic and evaluative instruments. Standardized diagnostic tests in reading should be given at least once a year, so that the teacher will be able to appraise the progress of her pupils, to determine reading difficulties, and to plan an appropriate developmental program."

Hildreth[4/] discusses determination of a beginning pupil's readiness status and recommends the use of an intelligence test and a gen-

---

1/Traxler, "Planning and Administering a Testing Program," op. cit., pp. 257-258.

2/C. W. Martin, "The Elementary School Testing Program," The National Elementary Principal (February, 1946), 25:8.

3/Ernest W. Tiegs, Diagnosis in the Reading Program, Educational Bulletin No. 10, California Test Bureau, Los Angeles, California, 1951, p. 8.

4/Gertrude Hildreth, Using Readiness Test Results, Test Service Notebook No. 10, World Book Company, Tarrytown-on-Hudson, New York, 1950, p. 1.

eral and composite readiness test to determine pupil limitations which instruction may overcome.

Clark[1] notes that since arithmetic is one of the most essential basic skills, it should receive particular attention. He urges that standardized diagnostic tests in arithmetic reasoning and arithmetic fundamentals be given to all pupils above the primary grades at reasonable intervals, preferably at least once a year. The value of standardized achievements tests in providing a diagnosis of specific learning difficulties for the individual pupil as well as survey data is pointed out by the California Test Bureau.[2]

Measuring achievement in the secondary school.-- Traxler[3] and King[4] recommend that achievement testing be continued at the secondary level and that broad field tests and the single achievement test in appropriate subjects be administered each year. The Division of Test Research and Service of the World Book Company[5] recommends that test batteries be administered at least at two points: (1) early in the program prior to course decisions and when provisions are made for developing special talent and for aiding pupils with special disabilities;

1/Willis W. Clark, Identifying Difficulties in Learning Arithmetic, Educational Bulletin No. 9, California Test Bureau, Los Angeles, California, 1951, p. 7.

2/California Test Bureau, How to Select Tests, op. cit., p. 2.

3/Traxler, Techniques of Guidance, op. cit., pp. 21-22.

4/King, op. cit., pp. 46-47.

5/World Book Company, Planning the Testing Program--Testing Program Organizational Chart, op. cit., p. 6.

and (2) toward the end of the program when decisions are made for the college-bound pupils and those going into jobs. It is further recommended that in grades 7-9 the minimum program should include measures of reading, vocabulary, language, arithmetic reasoning, and arithmetic computation.[1] In grades 10, 11, and 12 the program should include separate measures of achievement in mathematics, science, social studies, and language and literature.[2] The California Test Bureau[3] recommends the use of separate subject achievement tests at the secondary level and suggests that diagnostic tests in the basic skills be given shortly after the pupil's entrance into high school and each year thereafter.

Achievement testing in the high school poses greater problems than those encountered in the elementary school. Noll[4] cites the following limiting factors in the development of appropriate standardized achievement tests: wide variety of individual pupil study patterns, variety of subject offerings, and diversity of subject content and emphases. Ahman and Glock[5] cite the problem of developing meaningful norms for the content subjects. Lennon[6] also discusses this problem and comments

---

1/World Book Company, Planning the Testing Program--Testing Program Organizational Chart, op. cit., p. 2.

2/Ibid.

3/California Test Bureau, Conducting High School Guidance Programs, Educational Bulletin No. 7, California Test Bureau, Los Angeles, California, 1945, p. 2.

4/Noll, Introduction to Educational Measurement, op. cit., pp. 204-205.

5/Ahman and Glock, op. cit., p. 365.

6/Lennon, Testing in the Secondary School, op. cit., p. 1.

on this subject:

> "Care must be exercised, however, to ascertain that the standardized test considered for such use actually covers adequately the outcomes of the particular course--in other words, the user must consider the content validity of the test for his group. Not all biology courses, or general mathematics courses, or world history courses are identical; and a test appropriate for one course would not necessarily be equally valid for another course in the same subject."

The Commission on the American School Curriculum[1] evidences concern in the situation when the standardized test itself lacks reliance either to what is taught or should be taught in a specific situation. It is urged that such measures be used with caution and with full knowledge of their limitations.

The importance of certain basic skills, particularly reading and work-study, is cited by writers. It is generally agreed that competency in these skills is a controlling factor in school success. Lennon[2] recommends the use of reading and study skills tests for early identification of pupils requiring special assistance in these areas and for measurement of their progress in remedial work.

Noll[3] recommends the survey battery as a useful measure for determining the pupil's basic orientation in the broad fields and as a useful tool for guidance in assisting the pupil in making decisions about courses and curriculum. A review of many of the newer batteries reveals that they now combine to some extent the functions of the survey

---

1/American Association of School Administrators, op. cit., p. 321.

2/Lennon, Testing in the Secondary School, op. cit., p. 3.

3/Noll, Introduction to Educational Measurement, op. cit., pp. 206-207.

and diagnostic instruments by providing total scores indicating total level of achievement in broad subject areas plus subscores indicating specific strengths and weaknesses in various components of the total test. Many of the newer batteries aim to measure functional knowledge rather than specific subject content. The Sequential Tests of Educational Progress[1] provide such an approach in measuring achievement in the content fields. The Bureau of Examinations and Testing of the New York State Department of Education[2] makes the following comment regarding these measures:

> "Somewhat different orientation in achievement testing is measurement of understanding in a broad area rather than the specific content in a given subject. Since an important purpose of education is teaching the pupil to apply facts he has learned, emphasis is placed on general educational development. The typical areas sampled by such tests are skill in quantitative thinking, understanding of social concepts, interpretation of material in the physical sciences, and correctness and effectiveness of expression."

Jordan[3] notes that tests in music, art, and related areas are not as well developed as those in the basic academic areas and suggests that in the arts, real achievement consists primarily of products which lend themselves chiefly to rating. Caution against indiscriminate use of these tests is thus urged. Seashore[4] urges a comprehensive

---

1/Sequential Tests of Educational Progress, op. cit.

2/New York State Education Department, op. cit., p. 13.

3/Jordan, op. cit., p. 329.

4/Harold G. Seashore, "How May Tests Be Used to Obtain Better Articulation of the Total Educational Program?" The Bulletin of the National Association of Secondary School Principals (April, 1954), 38:285.

achievement program which will facilitate good decision making. He advocates that measures be integrated longitudinally from the elementary grades through high school and cross sectionally so that the results may be simultaneously multipurposeful.

Frequency of testing.-- The literature reveals that, ideally, achievement tests should be given annually in all grades. Many schools, however, for budgetary reasons, must be satisfied with an achievement program less than ideal. Ross and Stanley[1/] recommend a general test battery in all subjects about every three years, with intensive testing limited to one subject in each of the intervening years. Traxler[2/] notes that "when tests in only certain grades are possible, preference should be given to those points when major decisions are made about a pupil's school placement and his educational and vocational plans." Most writers agree that tests should be administered regularly at the following critical transition periods in the pupil's school career:

1. transition from home to school
2. transition from primary unit to intermediate unit
3. transition from intermediate unit to junior high unit
4. transition from junior high unit to senior high unit
5. transition from senior high unit to post-high unit.

In addition, diagnostic measures and separate subject measures should be administered as the need arises and the local situation dictates.

1/Ross and Stanley, op. cit., p. 210.

2/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 358.

Summary.-- Standardized achievement tests are invaluable measures in determining the achievement level of groups as well as individual pupils, and form the core of any systematic testing program. Their purpose and limitations must be thoroughly understood and their indiscriminate use discouraged. These measures should be used primarily as a means of evaluating achievement and not as a means of rating pupils and subtly judging the teacher's competence.[1/]

The school should recognize that there are many ways of appraising achievement. Standardized testing should be supplemented with other means of appraisal. Locally constructed achievement tests represent an integral part of this program. Engelhart[2/] notes that these measures are more closely related to instructional objectives and their use can be a definite factor in the improvement of instruction and in the improvement of evaluation.

Current practices surveys reviewed in this study[3/] reveal that achievement testing is the hard core of all standardized testing. Michaelis and Howard[4/] report that all schools administer standardized tests at regular intervals, with reading measured usually at grades 1,

---

1/Harry N. Rivlin, The Teacher's Role in Achievement Testing, Test Service Notebook No. 9, World Book Company, Tarrytown-on-Hudson, New York, 1950, pp. 1-2.

2/Max D. Engelhart, "Making Testing Meaningful to Teachers Through Local Test Construction and Analysis of Test Data," 1953 Invitational Conference on Testing Problems, Educational Testing Service, Princeton, New Jersey, 1954, pp. 55-60.

3/The reader is referred to Chapter III, supra.

4/Michaelis and Howard, op. cit.

2, 3 or 2, 3, and 6, and general achievement usually at grades 4, 5, and 8. All other studies report wide use of achievement tests. English and reading are most frequently tested areas. Wide usage is reported of the general achievement battery at the elementary level and separate subject testing at the secondary level.

10. Appraising Interests

Place of interests in school program.-- Interest is an important factor in the process of course selection and career planning. The literature indicates that interests are relatively unstable during the upper elementary and junior high school years, but notes that a study of interests at the junior high and senior high level may aid in assisting students in developing clearer aims and plans. Traxler[1/] cites two kinds of information that the school should obtain:

> "1. It should record the individual pupil's activities as an indication of functioning interest.
>
> 2. It should summarize the interest of the pupil on the basis of careful observation and scores on standardized interest questionnaires."

Measures of interest are useful in many ways. Wrightstone, Justman, and Robbins[2/] list six reasons why the school should be concerned with the interests of pupils:

> "1. To promote desirable interests
> 2. To foster new interests
> 3. To discover undesirable interests
> 4. To develop teacher-pupil rapport

1/Traxler, "Evaluation of Aptitude and Achievement in a Guidance Program," op. cit., p. 7.

2/Wrightstone, Justman, and Robbins, op. cit., pp. 296-297.

5. To enliven the curriculum
6. To provide educational and vocational guidance."

Holmes[1] notes the following use of measures of interest: (1) in identifying unknown or unrecognized interests, (2) in confirming stated interests, (3) in identifying potential or actual conflict between stated and tested interests, (4) in identifying discrepancies between interests and aptitudes, (5) in verifying the absence of well-defined interest patterns, and (6) in identifying educational or vocational maladjustments due to inappropriate interests.

Means of evaluating interests.-- Clues to pupil interest come from many sources. Greene, Jorgensen, and Gerberich[2] list many informal methods such as questioning pupils about their interests, having pupils write themes about their interests, noting reading interest in books, interest in radio, movies, and television, and interest in school subjects. Wrightstone, Justman, and Robbins[3] add such means as interviews, student talks, class discussion, anecdotal records, and logs. Jordan[4] notes three basic techniques in discovering interests: (1) direct questioning, (2) observations, and (3) objective tests of information.

Informational tests of interest represent a more objective means. Jordan[5] cites that scores on interest tests are not clear-cut measures

1/Holmes, op. cit., p. 8.

2/Greene, Jorgensen, and Gerberich, op. cit., p. 289.

3/Wrightstone, Justman, and Robbins, op. cit., p. 306.

4/Jordan, op. cit., p. 444.

5/Ibid., p. 423.

of interest and indicates that they are not used as widely as subjective techniques. He further notes that the problem in measurement of interests is one of discovery for guidance and not of evaluation. Wide use is made of interest inventories to assess interest patterns. Noll[1] indicates that inventories are useful tools in assisting the pupil in making educational and occupational choices. He questions their use below the senior high school level because of the following considerations: (1) inventories are not aptitude tests; (2) interests may change in the early years, making the results of junior high school pupils very tentative; and (3) few pupils have the breadth of knowledge and experience to make valid choices among the wide range of activities presented in the inventories.

Froehlich and Benson[2] feel that interest inventories should not be administered to pupils under 17 years of age. Durost[3] questions the real significance of the interest inventory at the junior high school level, but notes that the measure may have some value as a rough and unreliable indicator of patterns of interest. He urges caution in its use and interpretation, noting the danger of relating the scores to an achievement measure for a particular type of occupation:

> "It is quite possible that a person would have all of the native talent in the world for a particular occupation and

---

1/Noll, Introduction to Educational Measurement, op. cit., pp. 286-290.

2/Froehlich and Benson, op. cit., pp. 35-40.

3/Durost, Tests and the Junior High School Guidance Counselor, op. cit., p. 5.

> not reveal it on an interest inventory because he had not had the experience prior to the time the interest inventory was taken to permit him to make a valid judgment of his own in the area."

King[1] notes that there are two types of score information given by current interest inventories, scores for specific occupations and scores for interest areas. He suggests that inventories following the occupational area scoring are more applicable to high school situations since the student is usually choosing a vocational field rather than a specific career. He also warns that interest inventories are rather difficult to score and require a trained person. Greene, Jorgensen, and Gerberich[2] make special note of this fact.

Conclusions.-- Writings reveal that interests play an important part in the pupil's school adjustment and vocational plans. The school should employ all means possible in determining interest patterns. Objective measurement of interests is best obtained by the interest inventory. Although test publishers recommend use at both the junior and senior high school levels, its use below the senior high school is questionable. Inventories should be used with care and caution. Interest should not be confused with aptitude.

Current practices surveys reported in Chapter II of this study indicate that interest tests or inventories are used to some extent in the schools. Studies did not concern themselves to any great extent with this area, and in the majority of cases indicated that interest

---

1/King, op. cit., p. 40.

2/Greene, Jorgensen, and Gerberich, op. cit., p. 290.

inventories were available for special needs or purposes at the secondary level. From findings reported in these studies, it is evident that interest inventories do not have extensive use in the schools.

11. Appraising Personal-Social Adjustment

Overview.-- The literature reveals that the area of personal-social adjustment is complex and not susceptible to ready measurement. The measurement of personality offers many difficulties not encountered in other areas. Jordan[1/] notes that the techniques of measurement have not developed to the desired stage of validity. Traxler[2/] notes that there is no general agreement concerning a definition of personality or the number and nature of the traits of which it is composed. King[3/] notes that psychologists have not as yet isolated the basic components of emotional adjustment. He further adds that there appears to be little statistical or subjective agreement as to what the traits are that the various instruments of adjustment and personality are attempting to measure. Krugman[4/] succinctly indicates the state of the field with the following comment:

> "It seems no exaggeration to state that, in spite of thirty years of extensive experimentation and in spite of

---

1/Jordan, op. cit., p. 467.

2/Arthur E. Traxler, "The Use of Tests and Rating Devices in the Appraisal of Personality," Educational Records Bulletin No. 23, Educational Records Bureau, New York, 1938, p. 56.

3/King, op. cit., p. 42.

4/Morris Krugman, "Changing Methods of Appraising Personality," 1956 Invitational Conference on Testing Problems, Educational Testing Service, Princeton, New Jersey, 1957, pp. 55-56.

> a vast array of instruments, projective and non-projective, that are available today, we do not now have instruments for personality appraisal that are generally acceptable to psychologists. There seems to be a growing belief among many psychologists that we have, perhaps, been on the wrong track in seeking such instruments; that personality is so complex and personality factors, at best, so unstable, changing rapidly from situation to situation, that simple approaches to personality appraisal can yield little that is valid."

The literature abounds with writings regarding various aspects and techniques of personality appraisal. Recent writings pertaining to structured tests are contained in the review of Coleman and Collett.[1/] Heimann and Rothney[2/] review projective techniques. An extensive review of the many specialized methods of studying personality in a field that appears to be somewhat chaotic and disorganized is beyond the scope of this review. The present study presents a brief review of appraisal approaches that are most pertinent to the public school situation.

A variety of many types of instruments and techniques is available for evaluating personal-social adjustment. Ahman and Glock[3/] note that "some are based on informal observations, others on individual instruments designed primarily for use by clinicians, whereas still others are paper-pencil inventories to be administered to groups of pupils."

1/William Coleman and Dorothy Manley Collett, "Development and Applications of Structured Tests of Personality," Review of Educational Research (February, 1959), 29:57-72.

2/Robert A. Heimann and John W. M. Rothney, "Development and Applications of Projective Techniques," Review of Educational Research (February, 1959), 29:73-83.

3/Ahman and Glock, op. cit., p. 430.

Wrightstone, Justman, and Robbins[1] list the following techniques: (1) self-descriptive inventories or personal reports, (2) rating scales of personal and social conduct, (3) observational and anecdotal records, (4) free association and projective method, (5) autobiographies, (6) interviews, (7) sociometric techniques, and (8) situational tests. Some of the categories of appraisal methods comprise a variety of techniques and variations of the same technique.

Noll[2] cites two approaches to the measurement of personality: (1) the self-report approach, and (2) the observational approach. He notes that the first approach includes tests and other devices which yield scores and include personality inventories, projective tests, and attitude scales. Observational techniques, employing information supplied by sources other than the pupil being studied, include rating scales, anecdotal records, and sociograms. Writers agree that rating scales, anecdotal records, sociograms, and inventories are best adapted to serve the classroom teacher as evaluative instruments.

Interpretation based upon the use of these techniques must be tentative and contingent.[3] To apply and employ such techniques intelligently, the educator needs preparation in child psychology and mental hygiene. Wrightstone, Justman, and Robbins[4] strongly urge that the school should rely on a combination of techniques for corroborative

---

1/Wrightstone, Justman, and Robbins, op. cit., pp. 340-341.

2/Noll, Introduction to Educational Measurement, op. cit., p. 278.

3/Jordan, op. cit., pp. 467-468.

4/Wrightstone, Justman, and Robbins, op. cit., p. 352.

evidence. All writers agree that proper precautions must be exercised in measuring personal-social adjustment. Varying amounts of specialized knowledge are required to interpret these data. The literature reveals that various techniques employed in this area depend for their efficacy on the interpersonal relationship established between the examinee and the individual administering the test. Easy and quick conclusions cannot be drawn in interpreting data.

The literature is emphatic in stating that projective techniques are primarily clinical tools and should be confined to clinical studies. Wrightstone, Justman, and Robbins[1] note that they require considerable training and specialized skill for interpretation. In reviewing the research on projective techniques, Heimann and Rothney[2] note that very little progress has been made during the past few years in this area and make the following comment regarding the last six-year period of research:

> "At that time it was stated that research in this field was needed to separate what could be demonstrated from what was claimed. The present reviewers agree wholeheartedly with this statement. It would seem that the time has come to unify the piecemeal research and to concentrate on a co-operative effort at deriving meaning from the morass of casual empiricism that currently typifies much of the research with projective techniques."

Because of the specialized nature of projective techniques, writers cite that schools should not consider these measures unless there is a skilled psychologist on the staff.

1/Wrightstone, Justman, and Robbins, op. cit., p. 339.

2/Heimann and Rothney, op. cit., pp. 80-81.

Since the literature reveals that the personality inventory or adjustment questionnaire, observation, anecdotal records, and rating scales probably represent the chief means of personality appraisal used in the schools, they are reviewed briefly in this study.

Personality tests and inventory.-- Durost[1/] states that personality tests are largely in the developmental stage and should be used with extreme caution and sound judgment. This view prevails throughout the literature with but few exceptions. Noticeable among these are views expressed in bulletins distributed by publishers of these tests. A typical recommendation is found in Education Bulletin No. 11,[2/] which states that such tests "be utilized [periodically] in the primary, elementary, and secondary grades as a professional instrument for controlled observation and analysis of the status of personality development so that intelligent and constructive plans may be made for each pupil."

Wrightstone[3/] makes the following comment regarding personality tests:

> "The tests of personality are still in an experimental stage. Personality tests should be used with great caution and by only those persons who have adequate training for interpreting the results. Unless one has had some training in mental hygiene, it may be unwise to use self-description tests such as the California Test of Personality and other person-

1/Durost, Tests and the Junior High School Guidance Counselor, op. cit., p. 5.

2/Louis P. Thorpe, Appraising Personality and Social Adjustment, Educational Bulletin No. 11, California Test Bureau, Los Angeles, California, 1953, p. 11.

3/Wrightstone, op. cit., p. 257.

> ality tests. When these personality inventories are used, it is wise to supplement the rating by making systematic observations and anecdotal records of his emotional and social behavior."

Traxler[1] believes that the personality test has chiefly two uses for school purposes: to stimulate the pupils to evaluate critically their own personality characteristics, and to serve as a point of departure in conferences between counselors and the individual pupil. He goes on to make the following comment about the personality test:

> "Since it is generally conceded that the reliability of a test should be 0.9 if it is to be used as individual diagnosis, most of the personality tests and scales are not very satisfactory for the study of the individual. On the basis of reliability alone, the general conclusion may be drawn that scores on personality tests have at present doubtful value for use in cumulative office records of the different pupils. For the purpose of records of growth, anecdotal records and behavior descriptions seem preferable to existing personality tests. Likewise a few of the rating scales in which the categories are carefully defined and explained in very concrete terms probably provide more valuable records than most of the personality tests."

Because of the difficulty of securing adequate criteria of validity and because so little can be measured directly and so much must be measured by inference, writings urge the educator to proceed with caution in developing a program in this area.

The problem checklists are not tests in the usual sense and do not yield scores. Used individually, they provide clues which can guide the counselor in determining and providing individualized assistance.[2]

1/Traxler, The Use of Tests and Rating Devices in the Appraisal of Personality, op. cit., p. 56.

2/Kent Area Study Council, op. cit., p. 10.

Mooney[1] outlines five ways in which problem checklists may be used:

"1. to make group surveys which include finding out what youth are thinking about in their personal lives and the location of students who want and need counseling or other personal aid.

2. to provide a basis for group guidance, orientation, and personnel programs.

3. to increase teacher understanding in regular classroom teaching.

4. to facilitate guidance interviews.

5. to conduct research in the problems of youth."

It was also noted that such instruments may be used as a basis for guidance and orientation programs. Great care is urged in their use. Durost[2] feels that the adjustment questionnaire may be used as a diagnostic instrument in determining those pupils whose achievement in the social situation requires attention. He advises that it can be used, with extreme caution, to locate the extremely maladjusted child. Traxler[3] believes that, although designed to measure various traits, its chief use should be to locate extroversion-introversion. He further states that the questionnaire is mainly usable at the senior high school level and should not be used in the elementary grades. Ellis[4] makes

---

1/R. L. Mooney and M. A. Price, Manual to accompany Ross L. Mooney's Problem Check List, College Form, Bureau of Educational Research, Ohio State University, Columbus, Ohio, 1948.

2/Durost, Tests and the Junior High School Guidance Counselor, op. cit., p. 5.

3/Traxler, The Use of Tests and Rating Devices in the Appraisal of Personality, op. cit., p. 19.

4/Albert Ellis, "Personality Questionnaire," Review of Educational Research (February, 1947), 17:59.

the following comment regarding these instruments:

> "While experimenters continue to report satisfactory reliabilities for most of the tests employed, variability studies bring forth many unsatisfactory and highly questionable results. Authors of tests tend to find their instruments quite 'valid' but other observers do not corroborate these findings.
>
> The validity of the personality questionnaire seems to be much higher for some uses than for others. For purposes of distinguishing between good or bad students or teachers these tests are woefully inadequate.
>
> There is a continued punicious tendency on the part of many experimenters to employ personality questionnaires whose validity is still very much in doubt and, on the basis of scores on these tests, naively divide their subjects into 'neurotic' and 'normal' or 'introverted' or some dichotomous groupings.
>
> There can be no doubt whatever that a great deal remains to be done in the construction, validation, and application of personality inventories. Further research designed to increase test validity is still the crying need in this area."

Noll[1/] makes the following statement in summing up the use of personality inventories:

> "As to the question of whether personality inventories should be used in schools, it may be said, first of all, that they should be used very conservatively. Whereas the average classroom teacher can handle standardized tests of achievement and intelligence with some help and guidance from the counselor or school psychologist, this is seldom the case with personality tests. The use and interpretation of these require more training and experience than most teachers possess. As a rule, the personality test should be given on an individual basis rather than on a school-wide basis."

Observation.-- Techniques such as the rating scale, anecdotal records, and sociometry involve observation. Since these, as well as other methods of evaluating human behavior, are dependent upon efficient pro-

---

1/Noll, Introduction to Educational Measurement, op. cit., p. 282.

cedures for making and recording such observations, writers have given some consideration to the observation technique. Schwartz and Tiedman[1] distinguish between seeing and observing and define observation as "purposeful seeing directed toward the obtaining of useful facts or information about a specific person, place, event, object, situation, or condition." Wrightstone, Justman, and Robbins[2] define observational techniques "as systematic methods of analyzing and recording behavior by directly perceiving the individual or group." They further classify these techniques under those involving structured or controlled observation and those involving unstructured observation. Use of this method by the elementary teacher is constantly stressed. All writers emphasize the importance of training and skill in observation by the classroom teacher and the evaluator recording student behavior on such forms as charts, checklists, and anecdotal records.

The literature indicates that the types of situations which may be explored through observation are extensive. All writers emphasize the value of observation at the kindergarten and primary grade levels. The necessity of recording observations is stressed. Thorndike and Hagen[3] give illustrative studies using direct observation. Wrightstone, Justman, and Robbins[4] note the following devices for recording child behavior: anecdotal record, interview, checklist, rating scale, log,

---

1/Schwartz and Tiedman, op. cit., p. 193.

2/Wrightstone, Justman, and Robbins, op. cit., p. 33.

3/Thorndike and Hagen, op. cit., pp. 315-320.

4/Wrightstone, Justman, and Robbins, op. cit., pp. 363-374.

motion picture, and recording. At the upper grade level they list three general approaches for observing: (1) techniques which represent direct observational reporting of actual daily behavior, (2) approaches in which a stimulus is presented to students in order to secure their responses, and (3) use of make-believe dramatic situations or socio-dramatic techniques.

Noll[1] notes that much of the information about pupils and activities is based upon observations by teachers. Emphasizing this point of view, much of the literature is devoted to setting forth principles and procedures pertaining to reliable and accurate observation. Thorndike and Hagen[2] cover this aspect in detail and set forth the following steps to improve observational procedures:

> "1. Select certain aspects or categories of behavior to be observed.
>
> 2. Define the behaviors that fall within the category.
>
> 3. Train observers by practice session, etc.
>
> 4. Select and use a form of quantification, breaking the observation period up into short segments.
>
> 5. Provide some technique for immediate and efficient recording of the events that are observed."

Advantages and limitations of observation as a technique of evaluating pupil behavior are treated to some extent in the literature. Thorndike and Hagen[3] note the following significant points which con-

---

1/Noll, Introduction to Educational Measurement, op. cit., p. 304.

2/Thorndike and Hagen, op. cit., pp. 312-315.

3/Ibid., pp. 320-321.

tribute to the attractiveness of direct observation as a technique for studying the individual:

"1. Observation provides a record of actual behavior.

2. Observational techniques can be applied to the naturally occurring situations in life.

3. Observation is possible with young children and others for whom verbal communication is difficult."

Observation is by no means the answer to all measurement problems and has serious limitations. The limitations set forth by Schwartz and Tiedman[1/] summarize drawbacks noted in the writings:

"1. Reveals only overt behavior

2. Is cross-sectional in nature

3. Is difficult to interpret what is observed

4. May be difficult to record accurately what is observed, thus limiting the usefulness of the information

5. Depends for its value upon the purpose of the observer

6. Is influenced tremendously by the inferences, attitudes, biases, and prejudices of the observer."

Writings indicate that observation as a technique is one of the most useful means of evaluating pupil behavior. To be of value it must be done purposefully and used with understanding. Too often what is referred to as observation is, in reality, only seeing. Observation is flexible in its adaptability to a wide variety of purposes and situations, requires a considerable amount of skill, and is time-consuming. Its use, particularly as a research tool, and its limitations must be

1/Schwartz and Tiedman, op. cit., p. 194.

thoroughly understood by the educator. If this technique becomes an integral part of the evaluation process, the importance of training and skill in observation on the part of the staff can scarcely be overemphasized.

Anecdotal records.-- The literature indicates that anecdotal records provide an informal and largely quantitative picture of certain aspects of an individual's behavior and represent a practical approach to personality appraisals. Requiring less technical knowledge on the part of the teacher and not requiring the formulization and standardization of some of the other appraisal techniques, Wrightstone[1/] and Traxler[2/] cite that it is applicable to all students and to varied school situations. Wrightstone, Justman, and Robbins[3/] state that, to be of maximum value, this technique should concentrate on a few major aspects of child growth for which information cannot be obtained except by observation. They cite as examples the areas of social adjustment and growth, personal and emotional adjustment and growth, as well as other factors which have major influences on a particular pupil's adjustment. Thorndike and Hagen,[4/] as well as other writers, hold this view and note that these records should not be used for aspects of behavior that can be appraised by more objective and accurate methods.

Various features and use of the anecdotal record are covered exten-

1/Wrightstone, op. cit., p. 257.

2/Traxler, The Use of Tests and Rating Devices in the Appraisal of Personality, op. cit., p. 6.

3/Wrightstone, Justman, and Robbins, op. cit., p. 126.

4/Thorndike and Hagen, op. cit., p. 326.

sively by Krugman and Wrightstone[1] and Traxler.[2] The following characteristics of the anecdotal record are well stated by Thorndike and Hagen:[3]

"1. It provides an accurate description of a specific event.

2. It describes the setting sufficiently to give the event meaning.

3. If it includes interpretation or evaluation by the recorder, this interpretation is separated from the description and its different status is clearly identified.

4. The event it describes is one that relates to the child's personal development or social interactions.

5. The event it describes is either representative of the typical behavior of the child or significant because it is strikingly different from his usual form of behavior. If it is unusual behavior for the child, that fact is noted."

Several of the writings develop criteria for recording anecdotes and note various recording forms. Writers agree that the important element is not the form but the content of the record. The appropriate form for keeping records is dependent upon the primary purpose for which they are being kept. In discussing this matter, Schwartz and Tiedman[4] set forth certain characteristics of the anecdotal record which are significant:

1/Judith I. Krugman and J. Wayne Wrightstone, A Guide to the Use of Anecdotal Records, Educational Research Bulletin No. 11, Board of Education, New York, 1949.

2/Arthur E. Traxler, The Nature and Use of Anecdotal Records, Educational Records Bureau Supplementary Bulletin D, Revised, Educational Records Bureau, New York, 1949.

3/Thorndike and Hagen, op. cit., p. 327.

4/Schwartz and Tiedman, op. cit., p. 147.

"1. It is factual, recording only the actual event, incident, or observation, uncolored by the feelings, interpretations, or biases of the observer.

2. It is a record of only one incident.

3. It is a record of an incident which is considered important and significant in the growth and/or development of the pupil."

Varied suggestions and principles pertaining to the use of anecdotal records appear in the literature. Traxler[1/] notes some practical suggestions pertaining to teacher use that appear frequently in the writings:

"1. Train teachers to observe and record properly.

2. Teachers must learn to observe and make records for the shy, introverted, colorless child as often as they do for the most forceful personalities in the group.

3. Teachers must learn to see individuals rather than groups, to observe and remember, and then record later in the day, making their recordings brief and definite.

4. The plan calls for a well-organized system of recordings in the central office and much free time on the part of some staff member for summarizing and interpreting the records."

Writings note that evaluations based on anecdotal records often lack reliability and thus recommend their use with caution, particularly if the evaluations are recorded in the pupil's permanent record. The problem of errors in sampling is noted by the Commission on the American School Curriculum[2/] in discussing limitations of this technique.

1/Traxler, Use of Tests and Rating Devices in the Appraisal of Personality, op. cit., pp. 29-30.

2/American Association of School Administrators, op. cit., p. 334.

Traxler[1] notes the following shortcomings which may seriously affect reliability:

"1. Most teachers are not trained observers.

2. Under group methods of instruction, many teachers feel that they do not have much time to observe the behavior of individuals and still record it.

3. There will be a tendency for most of the anecdotes to pertain to the pupils whose overt behavior is outstanding, while other pupils will be neglected entirely.

4. Records may become so unwieldy that the system will break down of its own weight."

Anecdotal records represent a fairly reliable means of evaluating pupil behavior and deserve a place in the pupil appraisal program. If their use is to attain maximum value, the cumulative aspect of the program must be stressed. Full cooperation of the teaching staff must be enlisted and observers must have a clear understanding of what is to be expected. A summary evaluation of this technique is provided by Thorndike and Hagen.[2] This statement succinctly presents the point of view of the literature reviewed and shows the place of the anecdotal record in the school-wide testing and evaluation program:

> "An anecdotal record provides a medium for recording the observation of a significant item of pupil behavior. When teachers have developed skill in selecting incidents and in describing them objectively, when the mechanics of record-keeping and summarizing are kept within reasonable bounds, and when the records are available for use by those whose concern it is to understand the individual pupil, such records can be a significant aid to working with children."

1/Traxler, The Use of Tests and Rating Devices in the Appraisal of Personality, op. cit., p. 29.

2/Thorndike and Hagen, op. cit., p. 331.

Behavior rating scales.-- Rating scales are used rather extensively in the schools to make and record observations and as a means of obtaining a systematic and objective sampling of opinion on certain characteristics of an individual. Wrightstone, Justman, and Robbins[1/] list five major types of scales: (1) the descriptive rating scale, (2) the graphic rating scale, (3) the forced choice technique, (4))the rank order method, and (5) the paired-comparison method. Schwartz and Tiedman[2/] also classify scales according to the individual rating, as follows: (1) self-rating scale, (2) scale for rating others, and (3) scales that do both. Regardless of the type, writings reveal that their main purpose is to translate observation of people into quantitative terms.

There are many standardized rating scales of various types on the market, the graphic predominating. In addition to these, the literature notes the use of locally constructed scales to meet specific needs. Wrightstone[3/] and Traxler[4/] note that these measures meet a real need and, together with anecdotal records, represent a practical approach to personality appraisal. Writers generally agree, however, that they should be used to appraise only those qualities for which no valid objective measures are available. These scales should supplement, not

---

1/Wrightstone, Justman, and Robbins, op. cit., pp. 163-167.

2/Schwartz and Tiedman, op. cit., p. 166.

3/Wrightstone, op. cit., p. 357.

4/Traxler, The Use of Tests and Rating Devices in the Appraisal of Personality, op. cit., p. 6.

supplant, more objective means. Rating scales are tools in the hands of the user. The literature thus emphasizes considerations in making proper use of these devices.

In discussing the reliability of rating scales, Symonds[1/] notes some basic considerations in their use:

"1. Ratings should be made in a systematic way.

2. An extended period of observation should precede rating.

3. More attention should be paid to defining the qualities or traits to be rated and more extensive definitions should be introduced.

4. For experimental purposes all ratings should be disregarded except those which are at the extreme ends of the rating scale and those on which the raters are sure of their judgment.

5. Single ratings should not be used in the rating of human qualities. Sufficient reliability may be obtained only when a compromise is made of the independent judgment of from five to ten observers.

6. Traits for rating should be selected which experience shows yield better than average reliability.

7. So far as possible bias should be eliminated from ratings. Individuals should not be expected to give fair ratings when judging themselves, friends, old acquaintances or persons whom they much like or dislike, admire, or despise."

In discussing refinements, Thorndike and Hagen[2/] note that researchers have attempted to obtain greater uniformity of meaning in traits included, and are basing ratings more closely upon observable

---

1/P. M. Symonds, Diagnosing Personality and Conduct, D. Appleton-Century Company, New York, 1931, p. 106.

2/Thorndike and Hagen, op. cit., pp. 348-349.

behavior. They state that this movement has effected improvement in the following ways: (1) trait names have been defined, (2) trait names have been replaced by several more concrete and limited phrases, and (3) each trait name has been replaced by a substantial number of descriptions of specific behaviors. The rules of good scale construction noted by Jordan[1] are in agreement with principles noted in the literature:

> "1. Not more than seven divisions of the line
> 2. Divisions reinforced by careful verbal descriptions
> 3. A continuous line
> 4. Simplicity of administration
> 5. Extremes not so far distant from the mean that nobody will use them
> 6. Descriptive terms easily understood by the rater."

Noll[2] notes the necessity of obtaining judges who can express accurate and dependable opinions and the need to clearly define traits in terms of behavior rather than vague observations. Eurich and Wrenn[3] note that scales should call for judgments only on those behavior traits which can be observed objectively by the rater, and make the following recommendations:

> "1. The traits to be rated on the scale should be small in number and not overlap--from three to five traits is considered to be the accepted maximum.
>
> 2. The number of raters should be from three to five

1/Jordan, op. cit., p. 437.

2/Noll, Introduction to Educational Measurement, op. cit., pp. 298-299.

3/Alvin C. Eurich and C. Gilbert Wrenn, "Appraisal of Student Characteristics and Needs," Guidance in Educational Institutions, Thirty-seventh Yearbook, Part I, National Society for the Study of Education, Public School Publishing Company, Bloomington, Illinois, 1938, pp. 59-62.

for the best composite judgment.

3. The graphic scale is the most useful for school purposes."

The literature notes several limitations, hazards, and pitfalls in rating procedures. Thorndike and Hagen[1/] cite two major factors in obtaining valid appraisals of an individual: (1) factors that limit the rater's willingness to rate honestly and conscientiously according to the instructions, and (2) factors that limit the rater's ability to rate consistently and correctly. They further state that these effects "show up in certain pervasive distortions of the ratings, in relatively low reliabilities, and in doubt as to the basic validity of rating procedures."[2/] Hahn and MacLean[3/] list several generalizations based upon research evidence concerning errors obtained with scales:

"1. Self-ratings tend to be high on desirable traits and to be low on undesirable ones.

2. One tends to rate his own sex higher than the opposite sex on desirable traits, the reverse being true on undesirable traits.

3. Men are more lenient in their ratings than women.

4. In self-ratings, superior individuals underestimate themselves and inferior individuals overrate themselves, the latter having the greatest error.

5. Parents overrate their children as a rule, but they underestimate superior children. (Although not supported by research evidence it is the author's opinion

---

1/Thorndike and Hagen, op. cit., pp. 338-344.

2/Ibid., p. 344.

3/Milton E. Hahn and Malcolm S. MacLean, General Clinical Counseling in Educational Institutions, McGraw-Hill Book Company, Inc., New York, 1950, p. 163.

that the same error is made by teachers who tend, in general, to overrate their own pupils, but underestimate superior pupils, even in their own class or group.)

6. Two ratings by the same judge are no more valid than one."

Despite their limitations, rating scales represent a useful means of evaluating certain aspects of personality. Present findings indicate that they will continue to have extensive use in the schools. Schools using these means should do so with full awareness of the limitations of the instrument and attempt to minimize these limitations. Thorndike and Hagen[1/] sum up means of obtaining maximum accuracy with these instruments:

"1. Appraisal is limited to those qualities that appear overtly in interpersonal relations.

2. The qualities to be appraised are analyzed into concrete and relatively specific aspects of behavior, and judgments are made of these behaviors.

3. A rating form is developed that forces the rater to discriminate and/or that has controls for rater differences in judging standards.

4. Raters are used who have had the most opportunity to observe the individual in situations in which he would display the qualities to be rated.

5. Raters are 'sold' on the value of the ratings and trained in the use of the rating instrument.

6. Independent ratings of several raters are pooled when there are several persons qualified to carry out ratings."

Other means.-- A complete review of the many and varied techniques of evaluating the personal-social adjustment of the pupil is beyond the

---

1/Thorndike and Hagen, op. cit., p. 367.

scope of this study. Because personal-social adjustment is a complex process, it demands extensive and varied approaches in evaluation. The literature contains voluminous references to a variety of means of evaluating personality. Two, in particular, receive considerable attention and deserve mention in this review.

The case study is taking on ever-increasing importance as a primary tool of evaluation, since it permits the synthesis of all pertinent facts about a pupil in one framework. Wrightstone, Justman, and Robbins[1/] note that its purpose is twofold: diagnosis and treatment. The literature reveals that an increasing number of schools are turning to the case study method. It thus deserves a place in the appraisal program.

Knowledge of social relationships is recognized as important by the school. In recent years considerable attention has been directed toward systematic information about interpersonal relations. Sociometry and the sociodrama represent two methods used to obtain evidence of the pupil's development of behaviors in satisfactory human relationships. Jennings[2/] and others treat this area of appraisal. Evidence clearly indicates that sociometric techniques are gaining prominence and should receive consideration in a testing and evaluation program.

Summary.-- The importance of personal-social adjustment cannot be overestimated. In order to provide effective learning situations, the

---

1/Wrightstone, Justman, and Robbins, op. cit., p. 215.

2/Helen Hall Jennings, Sociometry in Group Relations, American Council on Education, Washington, D. C., 1948.

school must have an insight into the nature and causes of emotional problems. Evaluation of personal-social adjustment plays an important role in the total educational program. Personal-social adjustment is complex. Evaluation requires extensive and varied approaches, and is not susceptible to ready measurement. The school should move cautiously until valid and reliable measures are available. Appraisal in this field should be entrusted primarily to persons who have specific and highly specialized training. The findings and conclusions found in the literature may be summed up in the statement of King:[1]

> "The guidance program is thus given little assistance by current measures in the area of emotional adjustment. This situation is highly lamentable, because of the importance of the emotional component of personality in school, career, and other situations which the student must face. Much research remains to be carried out in this area before the educator has practical tools with which to work. The basic emotional traits are only partially known at present; few valid predictors of adjustment to life situations have been isolated; and the reliability of many of the methods currently being used is questionable. Again it is stated that, unless a school has a complete testing program in the areas of intelligence, interests, and achievement, the emotional area should not be covered."

Current practices surveys reviewed[2] show that personality tests have only infrequent use in the schools. Traxler[3] and North[4] reported use for only special cases. Leonard and Tucker[5] found that

---

1/King, op. cit., p. 45.

2/The reader is referred to Chapter III, supra.

3/Traxler, "The Status of Measurement and Appraisal Programs of Large City School Systems," op. cit.

4/North, "Testing Programs of Public School Members of the Educational Records Bureau--Report of a Questionnaire Survey," op. cit.

5/Leonard and Tucker, op. cit.

approximately 33-1/3 per cent of the schools made use of such measures. In the California area[1] it was reported that approximately 44 per cent of the schools use such measures for special purposes. Michaelis and Howard[2] noted that very little is being done to evaluate social attitude, emotional adjustment, social adjustment, and related needs of the learner. They further conclude that the greatest need for improvement is in the area of personal-social development. It appears from these studies that schools are moving rather slowly in this area. It is apparent that more advances are being made in the larger systems where psychologists and specialists are employed.

12. Testing Patterns

Overview.-- Although several illustrative and recommended testing programs are found in the writings, the literature, generally speaking, does not provide a particular pattern for a testing program. This is due to several factors. Local conditions dictate variations from time to time. The testing program may serve different functions in different schools. Wide variations exist in financial and professional resources available in each community.

Suggested and illustrated programs are noted in several textbooks on measurements. Test publishers have also prepared typical programs in chart form as guides for school systems developing testing programs. In many cases these appear in the form of minimal and maximal or expanded

1/Larson and McCreary, op. cit., p. 391.

2/Michaelis and Howard, op. cit.

programs. A selected bibliography of the sources noting testing programs or patterns is included in the Appendix.[1/] Certain generalizations and principles pertaining to the testing program are presented by the writer. These result from a synthesis of principles, practices, and recommendations reviewed in the literature. The reader is referred to earlier sections of Chapters II and III, as well as the selected bibliography on testing patterns and programs included in Appendix I. A suggested practical testing program or pattern for different levels is also included. This program is concerned entirely with group tests. It represents views of writers appearing in the literature.

1. A minimal program in the elementary and secondary schools should include periodic measures of mental ability and a regular schedule of achievement testing.
2. The hard core of the testing program includes two broad areas: (1) mental ability or scholastic aptitude, and (2) achievement in basic skills and content fields. Any school embarking upon a program should start in these two areas.
3. The clues to what should be measured in any school are the objectives and the curriculum. These should largely determine what to measure in the field of achievement.
4. Under certain circumstances repeat testing may be required. When special disabilities are noted, diagnostic tests in subject-matter areas may be used.
5. In a maximum or optimum program, achievement tests should be

1/The reader is referred to Appendix F.

given annually, and capacity measures (readiness, intelligence, prognostic and aptitude tests) should be included at transition periods. Achievement testing should include tests of a basically developmental nature at selected points in the student's school career.

6. Interest measures are rather difficult to interpret. If specialized personnel is available for counseling, an interest inventory may be administered in the high school.
7. Adjustment inventories and personality measures should be used with caution and then only when a special need is indicated. Qualified personnel should be available to administer and interpret the results of these measures.
8. It is assumed that the latter two measures are used only in an expanded or optimum program.

Minimal group testing program.-- The following recommendations are made pertaining to the minimal group testing program:

1. A reading readiness test is administered in late kindergarten or early first grade.
2. An intelligence or scholastic aptitude test is administered four or five times from the time the pupil enters school until he completes high school. The suggested sequence is at the end of grades 1, 3, 6, 8, and 10. A test including nonlanguage as well as language factors is included in the upper grades.
3. A complete achievement battery is given at the end of grades 3, 6, 8, and 10. In the upper grades at least one administration

includes general educational development measures.

4. Reading is measured in grades 2, 4, 6, 8, and 10.
5. A study skills test is given at the beginning of grade 7.
6. Separate subject achievement tests are given in English, mathematics, science, and social studies in grades 10-12 when needed.
7. Interest inventories are administered at the end of grade 9 provided guidance personnel is available for counseling.

Expanded or optimum group testing program.-- The expanded group testing program includes all measures noted in the minimal program plus additional measures used when they are needed to obtain objective information concerning the pupils.

1. Achievement tests are given annually in all grades, including fundamental skills.
2. Aptitude tests are given for educational and vocational guidance at the end of grade 9 or the beginning of grade 10.
3. College aptitude or vocational aptitude tests are given at the beginning of grade 12.
4. Interest inventories are given at the beginning of grade 11.
5. Personality measures are given when needed by qualified personnel and administered with extreme caution.

13. Organizing and Reporting Test Results

Overview.-- Writings point to the fact that the most important phase of the testing and evaluation program is the interpretation and use of test results. The basic assumption must exist that interpretation

and usage are fully as important as test selection, administration, and scoring.[1] Wellman[2] notes that if the testing program is to function, test results must be made available in a form understandable to staff, pupil, and parents.[3] Thorndike and Hagen emphasize the fact that an effective testing program implies an effective system of records.[4] Durost and Traxler[5] indicate the necessity or organizing test results for use by school functionaries as a preliminary step and an integral part of the process of interpeting and using the results. Traxler[6] succinctly emphasizes this phase of the over-all program in the following comment:

> "The value of testing in a guidance program is almost wholly dependent upon the effectiveness with which the results are used by the faculty of the school. Intelligent and efficient use of the data calls for a definite plan of testing and of organizing and reporting the results and rendering them understandable to teachers, many of whom have had little training in psychology, measurements, or statistical procedures."

Durost[7] notes that "to obtain full benefit from the pupil-centered testing program, the administration should also provide for recording

---

1/Kent Area Study Council, op. cit., p. 19.

2/Wellman, op. cit., p. 20.

3/Thorndike and Hagen, op. cit., p. 454.

4/Durost, What Constitutes a Minimal Testing Program for Elementary and Junior High School, op. cit., p. 1.

5/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 360.

6/Traxler, "Evaluation of Aptitude and Achievement in a Guidance Program," op. cit., p. 13.

7/Durost, What Constitutes a Minimal Testing Program for Elementary and Junior High School, op. cit.

the test data on a cumulative basis." The literature corroborates this view and reveals that, in addition to providing group data in understandable form to the school staff, a systematic plan for organizing pertinent information about the individual pupil on a cumulative basis is essential. Writings reveal that this can be best accomplished by an individual pupil cumulative record. Lee and Lee[1/] state that if the school accepts as a criteria of evaluation that the child's progress must be measured in terms of his growth, a cumulative record is essential. They further note that this record should begin when the child enters school and should follow him during his entire school career. Traxler[2/] notes that test data should be recorded in a meaningful and comparable basis in this record. Segel[3/] emphasizes the diagnostic and predictive value of the cumulative record because of the span of entries and notes its use by the teacher for any aspect of guidance and instruction.

The literature contains voluminous references to the cumulative record. A complete review is beyond the scope of this study. Any review of this phase of the testing program dictates its inclusion, since the literature reveals it is general practice to record test scores in each pupil's cumulative folder. Noll[4/] indicates that results must be made a part of the permanent records of the pupil as well as the school.

1/Lee and Lee, op. cit., p. 670.

2/Traxler, Techniques of Guidance, op. cit., p. 163.

3/Segel, op. cit., p. 35.

4/Noll, Introduction to Educational Measurement, op. cit., p. 312.

Each new test score entered in the cumulative record increases and stabilizes the knowledge about the pupil as an individual and adds to continuity in the use of results in studying the growth and development of the pupil over the years. The individual pupil cumulative records become an integral part of the process of organizing and reporting test results.

Writers cite that the results obtained from a test require interpretation. Test results possess added meaning when compared with those obtained from different populations to whom the tests have been administered. Norms provide a means of aiding in interpreting results and must receive consideration in any planned program of organizing and reporting results.

The present study reviews the literature pertaining to recording data obtained in the testing and evaluation program for staff use. It also considers derived scores used in test interpretation and circularizing results. The review does not treat statistical methods nor delve into the specifics of test analyzation and procedural aspects. It reviews general principles and practices pertaining to these selected phases of organizing and recording test results.

Recording test results.-- The writings reveal that an analysis of test scores is essential for intelligent interpretation and utilization of results. The use of such results depends upon the manner in which they are made available to staff members. The nature of the report and record will vary somewhat according to the local situation, facilities for interpretation and use, and the purpose it is to serve. Its main

function is succinctly stated by Stenquist:[1] "To present test results and related information in such a meaningful way as to arouse interest and action on the part of teachers, principals, supervisors, directors of special divisions, and superintendents." Ross and Stanley[2] give extensive coverage to this topic.

The Kent Area Study Council[3] notes that certain reports and records are essential to the success of the testing program and makes the following generalizations regarding records and reports:

"1. A plan for recording test results must be considered an integral part of test interpretation and usage.

2. System-wide procedures for recording test data should be established by the test director.

3. Recording should be done in consistent form.

4. Summary sheets should be made in triplicate for multiple use."

Roeber[4] indicates that if results are to be meaningful to the teacher, the following minimum information must be included in the summary sheet:

"1. Full name of test
2. Level and form of test used
3. Raw score made on test
4. Derived score

1/John L. Stenquist, "The Administration of a Program of Diagnosis and Remedial Instruction," Educational Diagnosis, Thirty-fourth Yearbook of the National Society for the Study of Education, Public School Publishing Company, Bloomington, Illinois, 1935, p. 518.

2/Ross and Stanley, op. cit., Chaps. VIII, IX, X.

3/Kent Area Study Council, op. cit., p. 5.

4/Edward C. Roeber, "A Meaningful Record of Tests," Educational and Psychological Measurement (Autumn, 1948), 8:397.

5. Norm group or basis of comparison
6. Probable error of test or parts of test
7. Test administrator
8. Date of test administration
9. Unusual testing or administering conditions."

Ross and Stanley[1] and Traxler[2] recommend the use of the class record and class analysis chart as a means of identifying strengths and weaknesses. These usually are provided with commercially published tests. They further note the value of developing distributions for class, buildings, and entire groups for use by administrators. Stenquist[3] recommends duplicating analysis charts for each class for use by teachers, principals, supervisors, and superintendents. The summarization of district-wide test data by class groups within each school and for grades as a whole is noted by Fisher.[4] Lewerenz[5] summarizes the procedure followed by the Los Angeles Public Schools to bring test data to teachers, supervisors, and administrators. An item analysis chart for studying results is suggested by Ahman and Glock.[6] In most cases they state that this is most applicable to systems equipped with scoring and tabulating machines.

1/Ross and Stanley, op. cit., pp. 237-240.

2/Traxler, "Ten Essential Steps in a Testing Program," op. cit., p. 360.

3/John L. Stenquist, "Devices for Testing," The Nation's Schools (September, 1937), 20:30-33.

4/Marie R. Fisher, Using Test Results to Improve the Education of the Individual Child, Test Service Bulletin No. 76, World Book Company, Tarrytown-on-Hudson, New York, p. 4.

5/Alfred Speir Lewerenz, "New Developments in Evaluating Achievement in the Public Schools of Los Angeles," Education (December, 1950), 71:237-244.

6/Ahman and Glock, op. cit., p. 518.

Graphical portrayal is recommended by writers. King[1] suggests the use of individual pupil profile sheets for collecting individual pupil scores and composite profiles for groups to students to indicate standings of given rooms, class, school, or district. Cook[2] cites the advantage of cumulatively and graphically portraying test results and concludes that this means may best serve as a stimulus in effecting improvements in curriculum emphasis, teaching procedures, and methods of study.

Stodola[3] describes the procedures followed in Lebanon, Pennsylvania in recording test scores for teacher use, and lists the following criteria which are pertinent to this review:

"1. Only data deemed essential should be included.

2. All data should be recorded on one sheet for ease of reference.

3. Local norms should be used so that pupils are compared only with those youngsters within their own community.

4. As far as possible, all scores should be expressed in the same form so that comparisons can be made among them.

5. From a practical point of view, emphasis should be placed on identifying gross differences in the recorded scores."

1/King, op. cit., p. 62.

2/Walter W. Cook, "Use of Tests in the Supervisory Program," in Appraising the Elementary School Program, Sixteenth Yearbook, National Elementary Principal, Bulletin of the Department of Elementary School Principals, National Education Association, Vol. XVI, No. 6, July, 1937, Washington, D. C., p. 470.

3/Quetin Stodola, How One School System Records and Interprets Test Scores: A Do-It-Yourself Kit for Teachers, Test Service Bulletin No. 89, World Book Company, Tarrytown-on-Hudson, New York, 1958, p. 6.

Using derived scores.-- Test scores of individuals acquire meaning when they can be compared with the scores of well-identified groups of students. Several systems for deriving meaningful standard scores are noted in the literature. All the textbooks in measurement reviewed in this study give extensive treatment to norms. Bulletins distributed by test publishers also treat the subject in detail. Typical of these are Methods of Expressing Test Scores[1/] and The Characteristics, Use, and Computation of Stanines.[2/]

Thorndike and Hagen[3/] list age, grade, percentile, and standard score as the most frequently used types of parometers in the schools today. Lennon[4/] notes that the appropriateness of a particular unit in which to express scores depends upon: (1) the nature of the ability or function being measured, and (2) the use to be made of the score. The utility, limitations, and technical considerations in developing norms are covered in detail in the literature and are beyond the scope of this review.

The literature also advocated the use of local norms as supple-

1/Psychological Corporation, Methods of Expressing Test Scores, Test Service Bulletin No. 48, January, 1955, Psychological Corporation, New York, 1956.

2/Walter N. Durost, The Characteristics, Use, and Computation of Stanines, Test Service Notebook No. 23, World Book Company, Tarrytown-on-Hudson, New York, 1959.

3/Thorndike and Hagen, op. cit., p. 153.

4/Lennon, "Discussion of the School Administrators' Problems," op. cit., p. 98.

mentary to the national norms provided by the test publishers.[1] Jordan and the Psychological Corporation,[2] in particular, urge that local norms be constructed for appropriate groupings of cases and note that they are essential for good interpretations of test scores.

Wrightstone, Justman, and Robbins[3] note recommendations of various committees which are attempting to establish standards for tests. These are pertinent to this review:

> "1. Tables or scales used to report scores should be designed to permit easy and accurate interpretation by the test user.
>
> 2. Norms should refer to clearly defined and described populations, or reference groups, such as grade, age, curriculum or occupational group. These populations should be the groups to whom users of the test will usually wish to compare the individual or group tested.
>
> 3. Norms should be based upon a representative cross section or sample, of the defined population or reference group.
>
> 4. Although norms may be reported in terms of grade or age groups, it is desirable also, to provide percentile equivalents or standard scores into which raw scores may be converted.
>
> 5. For some uses of tests, local norms are more appropriate and important than national, regional, or other group norms. In such cases, the test manual should suggest appropriate use of local norms."

Circularizing the results.-- Writers agree that four groups have an interest in knowing the results of the testing program: (1) super-

---

1/Jordan, op. cit., p. 9.

2/Psychological Corporation, Norms Must Be Relevant, Test Service Bulletin No. 39, May, 1950, Psychological Corporation, New York, 1950, p. 4.

3/Wrightstone, Justman, and Robbins, op. cit., p. 53.

visory and administrative staff, (2) teachers, (3) parents or public, and (4) pupils.

All writers advocate that school functionaires have access to test results that are of concern to them. Ross and Stanley,[1] in particular, note the need for each teacher to have the results for his class as quickly as possible after testing. The literature reviewed corroborates this view. General principles pertaining to this aspect are reviewed in an earlier section of the present study.

The literature reveals that it is the school's responsibility to tell its story to the people in understandable language both graphically and verbally. Farley[2] states the problem in this manner: "Parents wish to know what their children are being taught, how they are being taught, and what results are being achieved." Dobbins[3] states that there are three kinds of information that should be provided to parents: (1) purpose for which the test is used, (2) general characteristics of the test, and (3) nature of test scores. He further notes that the problem is one of how to instruct and inform the parents in the bare essentials of test understanding and score interpretation. Since there is a real possibility that the lay person may misinterpret and misuse the information, writers cite that certain cautions must be exercised.

1/Ross and Stanley, op. cit., p. 238.

2/Belmont Farley, What to Tell the People About the Schools, Bureau of Publications, Teachers College, Columbia University, New York, 1929, p. 19.

3/John E. Dobbins, "What Parents Need to Know About Tests and Testing," National Elementary Principal (September, 1954), 34:152-160.

Thorndike and Hagen[1] cite the need for sound interpretation in terms of educational or vocational significance. They further state that exact scores should be minimized and the "results of testing a particular individual should be interpreted to the lay public, and especially to the parent or pupil, by statements of general level or range of ability."

The literature generally concurs with the above view and recommends the presentation of test data to parents and public with appropriate explanation and interpretation. Ross and Stanley,[2] as well as other writers, feel that group results need to be organized effectively in graphs and charts illustrating the nature of tests, with analysis and interpretation centering around significant comparisons. Thorndike and Hagen[3] cite three qualities of an interpretation of an individual pupil's score that has significance:

> "1. The interpretation should be set in the frame of reference of the particular pupil.
>
> 2. The interpretation should be directed toward positive and constructive action.
>
> 3. It should be factual and dispassionate, rather than appearing to pass judgment on the individual."

Writers agree that test data should function in the lives of pupils and aid them in personal assessment of their aptitudes, interests, traits, strengths, and weaknesses. Thorndike[4] noted this fact

---

1/Thorndike and Hagen, op. cit., p. 452.

2/Ross and Stanley, op. cit., p. 245.

3/Thorndike and Hagen, op. cit., p. 452.

4/Edward L. Thorndike, "Tests and Their Uses," Teachers College Record (October, 1924), 26:93-94.

many years ago in the following statement:

> "The final justification for every testing regime rests in Mary Jones and John Smith, and it therefore behooves all persons who are making and giving tests to take them into partnership as soon and as completely as is feasible."

In discussing evaluation and tests in the Denver Public Schools, Leake[1] makes the following comment pertaining to the necessity of pupils understanding the testing program:

> "Pupils, too, need to have greater knowledge about their own learning advancement; they should understand some things about their own limitations and capacities; they should definitely be familiar with the purposes for having evaluative instruments, and should have a good understanding of the meaning of scores on such tests as they relate to individual expectancy."

Cautions and limitations regarding interpreting results to pupils are cited throughout the literature. The danger of presenting intelligence test results are noted by Ross and Stanley.[2] Dobbins[3] and others cite that the most useful application of test results is that application made by and with the student. Allen[4] recommends that the results be used in a program of pupil self-appraisal and interpreted by means of self-appraisal charts so that pupils may appraise their own

---

1/James D. Leake, Evaluation in the Denver Public Schools, Test Service Bulletin No. 83, World Book Company, Tarrytown-on-Hudson, New York, 1958, p. 2.

2/Ross and Stanley, op. cit., p. 237.

3/John E. Dobbins, "How Can the Results of a Testing Program Be Used Most Effectively?" Bulletin of the National Association of Secondary School Principals (April, 1958), 42:66.

4/Richard D. Allen, Tests for Instruction and Guidance, Test Service Bulletin No. 46, World Book Company, Tarrytown-on-Hudson, New York,

abilities, achievements, interests, and adjustments objectively and realistically. The literature generally holds this view and encourages schools to instruct pupils in studying their own traits so they may see their own competence and limitations.

Summary.-- Test interpretation and usage are integral parts of the total program. Ultimately, test interpretation must reach those who work directly with pupils, the pupils themselves, the parents, and those who make decisions affecting the curriculum and guidance activities. A plan for recording and circularizing test results must be adopted by all schools. This plan must provide for presenting test results and related information in a meaningful manner to school functionaires, parents, and pupils. Data should be recorded on a cumulative basis, and records format should be developed to present a high "glance" value.

The most frequently mentioned derived scores are age, grade, percentile, and standard scores. Very little mention is made of stanines. This is due to their relative recency of use. The use of local norms is strongly advocated as a means of increasing the meaningfulness of scores. The literature advocates giving results to all interested parties, but urges caution in interpretation and use. Writers cite that the importance of educating parents and pupils in test interpretation and usage cannot be overemphasized.

Current practices surveys reviewed in this study[1/] reveal that the majority of schools provide individual pupil cumulative records.

1/The reader is referred to Chapter II, supra.

Belanger[1] reported the following percentage use of derived scores by the schools: (1) grade equivalent, 85 per cent; (2) age equivalent, 66 per cent; (3) percentile, 55 per cent; (4) raw scores, 49 per cent. He further noted that 27 per cent of the schools reported the development of pupil profiles. North[2] reported that the single type norm is used exclusively. The findings reveal that local norms are practically nonexistent. Michaelis and Howard[3] reported that test interpretation and records system represent one of the major problems of the schools.

Although the above studies give minimum attention to organizing and reporting results, findings evidence a need for schools to give more concern to this phase of the program. Writers agree that a greater effort needs to be exerted to organize results in a meaningful fashion which will be conducive to ready use by all school functionaires. Teacher as well as pupil education is necessary to accomplish this.

14. Using the Results of the Testing and Evaluation Program

Overview.-- The literature emphasizes the point that the crucial aspect of any testing and evaluation program is the use that is made of the measures of that program. Obtaining data about individual pupils is simply a means to an end. The real test of the program is the manner in which the results are used. No program can be justified unless the

---

1/Belanger, op. cit.

2/North, op. cit.

3/Michaelis and Howard, op. cit.

results are used to assist the individual in making the maximum school adjustment and development in accordance with his potential, and as a basis for improving the educational program. Traxler[1/] states, "The value of a good testing program--when it is intelligently used by the entire staff in combination with comprehensive records--can hardly be overrated."

Testing and evaluation may serve many purposes. Standardized tests are not the exclusive tool of any one branch of the profession. The following statement by the Bureau of Examinations and Testing of the New York State Education Department[2/] succinctly emphasizes this truism:

> "Standardized tests are not the exclusive tool of any one branch of the educational profession. The administrator faced with problems of broad planning, the supervisor concerned with the professional development of the teaching staff, the counselor seeking to help the child adjust to school and vocation and the classroom teacher directly and actively engaged in guiding the progress and growth of the child--all can utilize the results of a sound testing program to tremendous advantage. Nor should we overlook the benefits likely to be derived by the child himself when he is given insight into and understanding of his own needs, potentialities and achievements."

Wrightstone, Justman, and Robbins[3/] classify the various measures according to use by school functionaires as follows: administrative uses, instructional uses, educational and vocational guidance, and research uses. It is noted that several of these purposes or uses may

---

1/Arthur E. Traxler, "Standardized Tests," National Education Association Journal (November, 1959), 48:18.

2/New York State Education Department, op. cit., p. 9.

3/Wrightstone, Justman, and Robbins, op. cit., p. 38.

be satisfied by the same data. Thorndike and Hagen[1] list the following three functions of a testing program: (1) classroom functions, (2) guidance functions, and (3) administrative functions. Noll,[2] Traxler,[3] and others cite major uses of measurement. Considerable duplication regarding usage occurs in the writings. The literature reveals the following main purposes for which tests and other instruments are used in the school:

1. Curriculum evaluation and planning
2. Classification of pupils
3. Grouping of pupils
4. Diagnosis and remedial work
5. Counseling and guidance
6. Motivation of pupils and staff
7. Identification and study of exceptional pupils
8. Interpreting the schools to the public
9. Evaluating and improving instruction
10. Educational research
11. Giving psychological security to, and improving, the school staff.

The literature cites ways in which the classroom teacher, guidance

1/Thorndike and Hagen, op. cit., p. 423.

2/Noll, Introduction to Educational Measurement, op. cit., Chap. XIV.

3/Arthur E. Traxler, Robert Jacobs, Margaret Selover, and Agatha Townsend, Introduction to Testing and the Use of Test Results in Public Schools, Harper and Brothers, New York, 1953, pp. 92-95.

counselor, and other staff members may make use of standardized test results. It also notes uses that the school administrator may make of test data. The administrative uses listed by the California Test Bureau[1] are typical of suggested uses included in other writings:

"1. To refute ill-founded charges that school achievement is below reasonable expectations, when test results show achievement to be satisfactory.

2. To determine whether differences in achievement between succeeding grades are satisfactory.

3. To determine whether the objectives of the curriculum are being achieved.

4. To determine whether marking practices in various schools reflect the true performances of the students as revealed by test results.

5. To determine whether the proportion of student 'failures' (where students are failed) reflects the true performance of the student as revealed by the test results.

6. To use as a basis for developing policies on ability grouping of students for instructional purposes.

7. To determine whether the achievement test results are reasonable and satisfactory in light of the intelligence of the students and other related factors."

The literature also cites the value of tests in the supervisory process. Cook[2] reports varying uses in this field. Lazar and Aronow[3] note the valuable contributions of tests in connection with research and experimentation in upgrading the educational program.

1/California Test Bureau, Administrative Uses of Test Results, Educational Bulletin No. 4, California Test Bureau, Los Angeles, California, 1951, p. 4.

2/Cook, "Use of Tests in the Supervisory Program," op. cit., pp. 470-473.

3/Mary Lazar and Miriam Aronow, "Research Uses of Testing Program," Education (March, 1957), 77:395-398.

The number of uses to which test results can be put is great. The nature and extent of test utilization depend upon the local situation, the organization, and the competencies of the school staff. New approaches are being developed constantly as personnel apply imagination and new-found methods to the task. Many different techniques are available to staff members in making maximum utilization of measurement data.

The literature spells out the specifics of varied uses and techniques. Detailed resource material may be found in the following:

1. Standard texts on measurement and guidance
2. Testing bulletins, manuals, and materials distributed by test publishers
3. Research reports by various groups and organizations
4. Articles in professional journals
5. Local, state, and national curriculum bulletins.

A complete review involving all the specifics of utilization is beyond the scope of the present study. As stated in the Delimitation of the Study,[1/] this investigation is concerned primarily with the following areas of pupil appraisal: (1) academic aptitude or intelligence, (2) achievement, (3) special aptitude, (4) interest, and (5) personality. The utility of measures in these areas is covered extensively in the literature. A critical review of this material reveals many uses of major import. Significant administrative and supervisory uses cited in the literature follow.

1/Supra, p. 6.

Intelligence tests.-- The major administrative and supervisory uses of intelligence tests are:

1. To determine how different groups compare in ability levels from school to school and within schools
2. To aid in the placement of pupils in grades
3. To aid in grouping within grades
4. To aid in determining the expectancy level in achievement within the class group and schools
5. To assist in planning special programs to meet the needs of individual pupils and class groups
6. To identify those pupils achieving significantly below their capacity to achieve
7. To study development of the abilities of individuals and groups over a period of years
8. To identify pupils who are significantly advanced or retarded
9. To compare pupil abilities with achievement in school courses.

Standardized achievement tests.-- The main administrative and supervisory uses of standardized achievement tests are:

1. To reveal the extent to which instructional objectives are being attained
2. To identify instructional areas needing special emphasis
3. To evaluate different teaching methods and materials of instruction
4. To rate the proficiency of different teachers
5. To determine how the achievement of local school groups compares

with that of comparable groups in other systems

6. To compare the achievement of different class groups within the system
7. To aid in the placement of pupils in grades
8. To aid in grouping within the grade or class group
9. To aid in curriculum revision and evaluation
10. To interpret the work of the school to the public
11. To aid in the counseling of individuals with respect to educational and vocational planning
12. To aid in planning special programs to meet pupil needs
13. To reveal exact points at which the tools of learning are inadequate or inoperative.

Special aptitude tests.-- The main administrative and supervisory uses made of special aptitude tests are:

1. To determine the need for special school programs
2. To aid in the placement of pupils in grades and for grouping within the grade or class
3. To place pupils in remedial or corrective classes
4. To suggest levels of instruction appropriate to the needs of individual students or groups of students
5. To conduct educational research
6. To guide pupils toward appropriate academic goals and courses
7. To group students of roughly similar levels of ability for specific instructional purposes

8. To note types of learning difficulty that might occur in different classes and schools.

Interest measures.-- The main administrative and supervisory uses made of interest measures are:

1. To provide a partial basis for pupil-teacher and/or counselor conferences on educational and vocational planning
2. To provide a basis for an occupational unit in various courses
3. To identify special interest areas coinciding with special or high level general ability
4. To note general fields of interests in groups
5. To note whether the level and types of work offered by the school meet the interests and needs of the student population.

Personality measures.-- The main administrative and supervisory uses made of personality measures are:

1. To determine the personal and mental health problems which characterize class groups, schools, or school districts
2. To determine the need for special attention to certain areas of the curriculum
3. To determine class groups or schools that need special assistance in the social school situation.

Limitations to the use of tests.-- Writings cite many limitations which restrict the extent and precision of the use of test results, and reveal that they are likely to be used erroneously unless these limitations are taken into consideration. The concern of school administrators in general is well expressed by Resolution Number 13, adopted

by the American Association of School Administrators, meeting in convention in Atlantic City, New Jersey, on February 18, 1959:

> "The importance and usefulness of tests in evaluation and teaching cannot be overestimated. There is a great need for improved and more adequate instruments for evaluation of many kinds of educational growth. The misuse of tests and the misinterpretation of test data continue to be a glaring danger to good educational programs. Any test instrument should be selected and used in terms of what a particular school had intended to teach. To judge a school solely on the basis of data derived from any battery of examinations is an invalid and dangerous venture. State or national examination results, used without due regard for the educational objectives of the school and the nature of the student body, are likely to be misleading."

Traxler[1/] mentions several limitations which seriously affect the use of test results:

> "1. Many tests do not provide valid measurement of the curriculum with which they are used.
>
> 2. No test is perfectly reliable.
>
> 3. The names of many tests are misleading.
>
> 4. The relationship of the test to future success in various fields is unknown."

MacMinn[2/] advises that tests must be used judiciously and calls attention to the following considerations:

> "1. Tests in and of themselves are far from being perfect measuring instruments.
>
> 2. The attitude and motivation of the pupil taking a test

1/Traxler, Techniques of Guidance, op. cit., pp. 198-199.

2/Paul MacMinn, "A Few Basic Facts About Tests," in Understanding Tests, United States Department of Health, Education, and Welfare, Office of Education, United States Government Printing Office, Washington, D. C., 1960, p. 3.

and the conditions under which he takes it influence the results.

3. The accuracy with which the results are interpreted determine to a great extent the usefulness of tests."

Bridges[1] states that the tendency for the administrator to infer the average score obtained in class on a standardized achievement test is a direct index of the effectiveness of the instruction in that class, places a serious limitation upon the interpretation of results of these measures. He further points out that the achievement level of a pupil is the result of a highly complex pattern of factors other than instruction that have been interacting for many years. He indicates that the achievement of students is not determined by the quality of instruction alone.

The literature states that testing is not the complete answer to pupil appraisal. Test data should never be used as the sole basis for action. Test results must be supplemented by additional information, and become meaningful only when interpreted with all other possible data. Only when tempered with good judgment and placed in the proper perspective will they provide a valuable means of understanding the individual pupil and aid in improving the educational program.

Summary.-- All writers agree that the crucial aspect of the testing program is the use made of test results. Proper interpretation and utilization of results are likely to be the weakest link in the whole

---

1/Claude F. Bridges, Some Basic Considerations in Determining the Significance of Achievement Test Results, Test Service Bulletin No. 66, World Book Company, Tarrytown-on-Hudson, New York, pp. 2-3.

program. Without them the whole program is of questionable value. No program is justified unless proper and full utilization is made of the results.

Tests are multipurposeful. Teachers, counselors, supervisors, and administrators use the measurement data in varied ways in working with individual pupils and in the instructional and curriculum improvement program. Although the value and usefulness of tests cannot be overestimated, their misuse and the misinterpretation of their data continue to be a glaring danger to good educational procedures. Proper utilization requires that the educator be fully cognizant of their limitations. Writers agree that tests are merely tools, the value of which depends upon the insight and skill of the user. There is no magic in mere use.

The findings pertaining to test utilization in current practices surveys[1/] are somewhat limited. In studying large city systems, Traxler[2/] reports that the primary use made of test results is for guidance, with instruction a close second. He also notes that considerable use of results is made by administrators and supervisors. Little use for public relations purposes was found. Leonard and Tucker[3/] found that major uses are for guidance and pupil classification. Belanger[4/] reports

---

1/The reader is referred to Chapter III, supra.

2/Traxler, "The Status of Measurement and Appraisal Programs of Large City School Systems," op. cit., pp. 75-86.

3/Leonard and Tucker, op. cit.

4/Belanger, op. cit.

the following uses, arranged in decreasing order of importance: sectioning and placement, individual counseling, appraisal of the individual, diagnosis, individualization of instruction, and as a check on the efficiency of teachers and the curriculum.

CHAPTER IV

METHOD OF PROCEDURE

1. Preliminary Considerations

Sources of data.-- Many and varied sources of data were utilized to gain background information prior to formalizing the study. During the progress of the study these sources were tapped continually to obtain pertinent data. General sources are noted below:

1. Conferences were held with the writer's major advisor and with members of a measurements seminar and workshops in testing and evaluation problems at Boston University.
2. The writer attended national and regional meetings of many professional organizations. Meetings of the National Council on Measurements Used in Education and Invitational Conferences on Testing Problems sponsored by the Educational Testing Service proved to be of particular significance. Research reports of the latter two were utilized.
3. Conferences were held with nationally recognized measurement authorities and with consultants and editors of major test publishers.
4. Conferences were held and correspondence was carried on with directors of guidance, testing, and research in various school systems throughout the East.

5. Several superintendents of schools in the New England area were consulted.
6. Correspondence was carried on with testing, guidance, and research personnel in State Departments of Education in all of the New England states.
7. Conferences were held with the Executive Secretary and members of the Executive Committee of the New England School Development Council to determine the feasibility of the study and possible endorsement by this group.
8. The writer's personal experience as a teacher, secondary school administrator, college instructor, director of curriculum, and Assistant Superintendent of Schools was utilized.

Review of the literature.-- A thorough review of the research and literature was made in order to obtain a general background for the study and to compile all possible information pertaining to testing and evaluation programs. Particular attention was directed to previous surveys and to general administrative aspects of a pupil-centered testing and evaluation program. Chapters II and III review the related literature and research. Primary sources in this project follow:

1. Card catalogues in the following libraries:
   a. School of Education, Boston University
   b. Graduate School of Education, Harvard University
   c. Teachers College, Columbia University
   d. University of Connecticut
   e. Boston Public Library

2. Issues of the Review of Educational Research[1] pertaining to measurement
3. Issues of the Education Index[2] covering the last ten-year period
4. Bulletins and reports issued by major test publishers covering important aspects of testing
5. Standard texts on measurement, guidance, and testing
6. Test manuals
7. Leading journals covering the field of measurement and evaluation
8. The Mental Measurements Yearbooks[3], [4]
9. The Encyclopedia of Educational Research[5]
10. Educational doctoral dissertations in progress and doctoral dissertations in education published since 1940.[6]

Selection of the geographical area.-- An early decision was made to limit the study to the New England public schools. A thorough review

---

1/American Educational Research Association, Review of Educational Research (February, 1947), 17:1; (February, 1950), 20:1; (February, 1953), 23:1; (February, 1956), 25:1; (February, 1959), 29:1.

2/Educational Index, H. W. Wilson Co., New York, 1950-1960.

3/Oscar K. Buros (Editor), The Fourth Mental Measurements Yearbook, Gryphon Press, Highland Park, New Jersey, 1953.

4/Oscar K. Buros (Editor), The Fifth Mental Measurements Yearbook, Gryphon Press, Highland Park, New Jersey, 1959.

5/Walter S. Monroe (Editor), Encyclopedia of Educational Research, The Macmillan Company, New York, 1952.

6/Stanley B. Brown, Mary Louise Lyda, and Carter V. Good, Research Studies in Education, A Subject Index, Phi Delta Kappa, Bloomington, Indiana, 1940-1959.

of the literature and research revealed that no thorough investigation had been made of testing and evaluation practices in this region. Many administrators in the area evidenced an interest in the project and indicated that the study met a definite need. The New England School Development Council also indicated a willingness to participate in the study. It was further felt that a thorough study on a regional basis would have more direct effect upon the local scene than one of national scope requiring limited sampling of local school districts. Because of the investigator's long association with the New England public schools, a high percentage of school participation was anticipated. For these purposes, the New England states were selected.

Selection of the type of instrument.-- Two techniques were considered as means of gathering data pertinent to the study: the personal interview and the inquiry form.

Use of the personal interview would require the investigator to make personal contact with administrators and/or supervisory personnel in 627 systems in the six-state region. Further investigation revealed that these individuals were scattered over the 66,608 square mile area of New England.[1/] Prohibitive costs of travel and time ruled out the personal-interview technique and made the inquiry form the most practical and feasible means of gathering date required to solve the problem presented in this study.

1/Dan Golenpaul Associates, Information Please Almanac, 1959, The Macmillan Company, New York, 1959, pp. 366-387.

Eells[1] defends this technique of investigation and notes the following regarding its use:

> "Over forty percent of the surveys depend more or less extensively upon questionnaires. Questionnaires in educational investigations have aroused much adverse criticism, some of it valid, because of their misuse. For gathering certain types of information the questionnaire is the only feasible means. When properly safeguarded, it is entirely satisfactory. Due attention should be given to its careful formulation, to the determination of those who are to receive it, to conditions of distribution, to the explanation of its purpose, and to the number and percentage of replies received."

A carefully devised form[2] was constructed as the information-gathering instrument for the study. Careful consideration was given to the points enumerated by Eells. A description of the instrument, as well as the procedure followed in its development, is found in the next part of this study.

2. Developing the Inquiry Form

Criteria.-- Careful attention was paid to the criteria suggested by Scates and Yeomans[3] in constructing the instrument:

> "1. It must be short enough so as not to take too much time and so that the respondent will not reject it completely.
>
> 2. It must be of sufficient interest and have enough

1/Walter Crosby Eells, Surveys of Higher American Education, The Carnegie Foundation for the Advancement of Teaching, New York, 1937, p. 113.

2/See Appendix C.

3/Douglas E. Scates and Alice V. Yeomans, The Effect of Questionnaire Form on Course Requests of Employed Adults, American Council on Education, Washington, D. C., 1950, pp. 2-4.

face appeal so that the respondent will be inclined to respond to it and to complete it.

3. The questionnaire should obtain some depth to the response in order to avoid superficial replies.

4. The ideal questionnaire must not be too suggestive or too unstimulating, particularly with reference to choices.

5. The questionnaire should elicit responses that are definite but not mechanically forced.

6. Questions must be asked in such a way that the responses will not be embarrassing to the individual.

7. Questions must be asked in such a manner as to allay suspicion on the part of the respondent concerning hidden purposes in the questionnaire.

8. The questionnaire must not be too narrow, restrictive, or limited in its scope or philosophy.

9. The responses to the questionnaire must be valid, and the entire body of data taken as a whole must answer the basic question for which the questionnaire was designed."

In addition to these criteria, the investigator employed the following guides in planning and developing items, mechanical features, and general format of the inquiry form:

1. Items are definite, concrete, and significant to the problem.
2. Items, when necessary, are open-ended in order to invite free responses.
3. Items elicit concise responses and avoid unnecessary details.
4. Items are phrased in a manner that facilitates tabulation and summarization of the responses.
5. The instrument solicits all required information and does not include extraneous elements.

6. The instrument is pertinent to the situation of the respondent.

7. The format of the instrument provides a minimum of difficulty in passing from one item to another in responding.

8. The instrument is precoded in order to facilitate analysis by the Punch Card Method.

A review of the inquiry form and study reveals that it meets each of the criteria developed by Koos,[1] which are documented as follows:

(a) "Is the questionnaire adequately sponsored?"

The New England School Development Council and the Test Usage Committee of the American Educational Research Association endorsed this study. In addition, the investigator's status as a school administrator, former officer of the State School Committee Association, and doctoral candidate provided adequate sponsorship.

(b) "Is the purpose of the study frankly stated, and is it one which calls for a reply under the policies dealing with questionnaires?"

The covering letter[2] accompanying the questionnaire clearly states the purpose of the study and requests participation of the school system. It further notes that a compilation of the data will contribute materially to a better understanding of what the schools are doing and will afford a measuring stick for comparing school practices.

(c) "Is the questionnaire on a worthy educational topic?"

Increased attention to student development in basic instructional areas, current charges of academic shortcomings in public education, and increased emphasis being placed upon the pupil appraisal program in order to secure valid data regarding student development in all areas of learning make the study a worthy educational document. The impact

---

1/Leonard V. Koos, The Questionnaire in Education, A Critique and Manual, The Macmillan Company, New York, 1928.

2/See Appendix D.

of the National Defense Education Act of 1958 further increases the import and timeliness of the study.

(d) "Is the questionnaire well organized?"

The inquiry form[1] was carefully constructed and follows a well-developed outline. It was painstakingly organized after a thorough review of the research. Measurement authorities who reviewed the instrument stated that it was well organized.

(e) "Are the questions clearly and briefly worded?"

A study of the inquiry form[2] will elicit an affirmative answer. The investigator received many favorable comments from respondents and jury members regarding this point.

(f) "Can most of the questions be answered briefly with a check mark of a factor figure and is the number of questions requiring extensive projective replies kept to a minimum?"

A review of the inquiry form[3] will provide an affirmative answer to the above question.

(g) "Is the information requested not available elsewhere and obtainable only through the questionnaire?"

A review of the research revealed that no previous study of testing and evaluation practices had been made in the New England area. Information provided by this study is not available through any other source.

(h) "Is the questionnaire set up in proper mechanical form?"

The design of the inquiry form[4] requires a minimum of effort by the respondent in passing from one item to another. It is furthermore precoded to facilitate tabulation and summarization by the punch card method.

(i) "Are the demands of the questionnaire reasonable?"

Approximately one half hour is required to complete the inquiry form. Information requested should be readily

---

1/See Appendix C.

2/Ibid.

3/Ibid.

4/Ibid.

available to the respondent.

(j) "Is the summary of results or other proper return promised respondents?"

The covering letter states that a summary of the results will be made available when the study has been completed. Arrangements have been made to print a brochure under the auspices of the New England School Development Council for distribution.

Validation of items in the inquiry form.-- After a careful review of the literature and research, an outline was developed covering the important aspects of a pupil-centered appraisal program. The outline was built around the following four major categories: (1) program organization and administration, (2) measuring devices used in the program, (3) organizing and reporting test results, and (4) using the results of tests and evaluation. Items covering important phases of the problem under investigation were then developed on separate cards and arranged systematically according to the outline under the four main categories. These items were submitted to the writer's major advisor and members of a Measurements and Evaluation Seminar at the School of Education, Boston University. The items were critically reviewed for clarity, pertinence to the problem, coverage of content, and appropriateness of terms. As a result of this review, certain ambiguities were eliminated, inaccuracies corrected, and only items significant to the problem retained.

A preliminary inquiry form[1/] of 47 items, divided into four major categories, was developed. Careful attention was paid to the criteria

1/See Appendix A.

previously presented in this study. This preliminary form was sent to a jury of 17 educators[1] in order to secure expert criticism of its make-up and content. The jury was composed of measurement authorities,[2] guidance directors, and selected superintendents of schools. Jury members were asked to review the inquiry form, considering the following points:

1. general format
2. clarity of directions
3. arrangement and clarity of items
4. relevance of items to the problem
5. general scope.

In addition to correspondence, the investigator held personal conferences with the majority of the jury members to obtain their valued judgment regarding the above points.

Revision of the inquiry form.-- Every member of the jury completed the inquiry form and made suggestions pertaining to general make-up and selected items. All suggestions were thoroughly considered. As a result of the pretest, certain refinements were made in the inquiry form. These refinements also reflected the recommendations of the writer's doctoral committee. Table 1 notes changes made in the instrument.

1/See Appendix E for list of individuals participating in inquiry form revision and tryout.

2/A measurement authority was considered to be one of the following: a testing director or specialist in the public schools or college; a testing specialist in a State Department of Education; a director of a testing division of a leading publisher; a director of a testing and advisement center; a testing author or editor.

Table 1. Number of Changes Made in Preliminary Inquiry Form as a Result of Jury Evaluation

Changes	Inquiry-Form Category				
	General Information	Part I	Part II	Part III	Part IV
(1)	(2)	(3)	(4)	(5)	(6)
Number of items deleted.....	1			2	
Number of items added.....	1				
Number of items revised.....	1	9	1	1	

A copy of the revised inquiry form may be found in Appendix B.

Pilot study for further refinement of the inquiry form.-- The revised inquiry form was further tested for validation in terms of practical use by a pilot study. Twenty forms were sent to educators,[1/] who were requested to complete the form, evaluate the instrument as a research tool, and make suggestions for improving the instrument. Conferences were held with several of these individuals, at which time the instrument was carefully evaluated. Particular attention was paid to clarity of directions and pertinence to the situation of the respondent.

Responses on the tryout were tabulated to determine if the data could be tabulated satisfactorily and if they warranted conclusions which were significant for the intended purpose of the study. Very few suggestions were received for improving the instrument. As a result of this pilot study, only minor revisions were made. These are noted in Table 2.

1/See Appendix E.

Table 2. Number of Changes Made in Revised Inquiry Form as a Result of the Pilot Study

Changes	Inquiry-Form Category				
	General Information	Part I	Part II	Part III	Part IV
(1)	(2)	(3)	(4)	(5)	(6)
Number of items deleted.....				2	
Number of items added.....					
Number of items revised.....		1			2

Printing of inquiry form.-- The final draft of the instrument was printed in booklet form. A copy of the finalized form may be found in Appendix C.

3. Distributing the Inquiry Form

Selecting the sample.-- As a preliminary to the task of distributing the inquiry form, educational directories and other data pertaining to school organization were secured from the State Departments of Education in each of the New England states. These data provided the investigator with a complete school organizational pattern for the 1957-1958 school year. Included in this material were directories of superintendents, listings of communities included in school unions, listings of separate community school districts, and complete enrollment data for every school system. The 1957-1958 directory of the New England Association of School Superintendents[1/] was also used as source

1/New England Association of School Superintendents, 1957-58 Directory of the New England Association of School Superintendents.

material.

Inquiry forms were sent to every single school district community (not to members of regional organizations or supervisory unions) and to a representative community chosen from each supervisory union or regional school district in New England. The following procedure was used to select the representative community for each union or regional organization. Communities in a supervisory union or region were arranged alphabetically. A number was assigned to each community consecutively. Corresponding numbers were written on chips. These chips were then placed in a receptacle, mixed thoroughly, and one chip was drawn. The community whose number corresponded to the drawn chip was selected as the representative community for that supervisory union or region. This procedure was followed for each supervisory union or region in New England. Every community in a union or regional organization thus had an equal chance of being selected, with no bias.

The alternate method of dealing with the regional or union problem would be to send the inquiry form to the superintendent and suggest that he complete the form for the town in his district which he considered to be the most representative. The investigator felt that bias might enter the selection and the town chosen would not be typical but would represent what the superintendent considered to be the best. For this reason, this method of sampling was rejected.

Six hundred twenty-seven communities were selected to participate in this study. Represented in this group were 347 communities in single school districts and 280 communities that were members of a regional or

supervisory union organization. School systems represented in the sample enrolled an aggregate number of 1,120,121 pupils. The small community, as well as the large city system, was represented in the study.

Distributing the inquiry form.-- Six hundred twenty-seven printed inquiry forms[1] were sent to superintendents throughout New England. Forms were addressed to the superintendent, since it was felt that he was in the best position to supply the data required for the study. Covering letters, requesting cooperation in supplying the requested data and setting forth the purpose of the study, accompanied each inquiry form. Two covering letters were devised: (1) for the superintendent of the single community district,[2] and (2) for the superintendent of the school union or regional district.[3] The latter requested the superintendent to supply information for the specific town in the union chosen at random for the study. The name of this town was written on the inquiry form. Stamped, self-addressed printed envelopes were included. Copies of the inquiry form and covering letters are included in the appendix.

Mailing of the inquiry form was started on April 15, 1958. This date was chosen because it was felt that April or October represent the most opportune time for the recipients to provide the information requested.

Preparation and distribution of the first follow-up letter.--

---

1/See Appendix C.

2/See Appendix D.

3/Ibid.

Approximately three weeks after the inquiry forms were sent, follow-up letters[1] were sent to those who had failed to return the inquiry form. This letter requested cooperation in the study.

Preparation and distribution of the second follow-up letter.-- Approximately two weeks after mailing the first follow-up letter, final follow-up letters[2] were sent to those who still had not returned inquiry forms. This letter noted the value of the study and announced the cut-off date. It further offered to supply an additional inquiry form to the superintendent in case the original form had been lost or misplaced. Window envelopes were utilized in the mailings of both inquiry forms.

Other methods of follow-up.-- Personal letters were sent to many superintendents and key educational leaders requesting their cooperation by alerting colleagues in the field to the desirability of participation. State Departments of Education personnel were also contacted and informed that the study was in progress. The study was also announced at several professional meetings.

Additional data.-- A large percentage of those with whom contact was made concerning this study were most cooperative. Many respondents sent letters, in addition to answering the inquiry forms, describing the testing and evaluation programs. Brochures were received from 20 school systems describing their testing program. Many letters of explanation concerning responses were also received.

1/See Appendix D.

2/Ibid.

Inquiry form returns.-- Six hundred twenty-seven forms were mailed to superintendents of schools in the New England states. Completed forms were received from 66.8 per cent of the communities in single districts and from 58.1 per cent of towns who were members of a supervisory school union or regional district. Table 3 gives the disposition of the 627 inquiry forms.

Table 3. Disposition of 627 Inquiry Forms Mailed to Superintendents of Schools in the New England States

State	Total Number Sent		Percentage of Towns Returning Completed Inquiry Form		
	Towns in Single District	Towns in Supv. Union	Towns in Single District	Towns in Supv. Union	Total
(1)	(2)	(3)	(4)	(5)	(6)
Mass.	181	52	67.2	57.6	65.6
Conn.	98	15	69.3	100.0	73.4
R.I.	37	0	56.7	0	56.7
Maine	12	131	50.0	54.1	53.8
N.H.	10	38	70.0	60.5	62.5
Vt.	9	44	77.7	54.5	58.4
Total	347	280	66.8	58.1	62.9

The large community is usually a single district and not a member of a supervisory union or regional district. These communities are thus generally included in column (4) in the table above. The small community is generally included in column (5). A study of Table 3 reveals that an adequate sampling has been obtained from both types of school systems. The total percentage of returns obtained in the study was 62.9 per cent.

School systems in all enrollment categories are represented in the study as shown in Table 4.

Table 4. Distribution of School Systems Represented in Study According to Enrollment Categories

State	School Systems According to Enrollment Categories						
	I Under 500	II 500-999	III 1000-2999	IV 3000-4999	V 5000-9999	VI Above 10,000	VII Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mass.	24	23	58	23	16	9	153
Conn.	7	12	32	11	13	8	83
R.I.	4	1	9	1	3	3	21
Maine	41	18	15	1	1	1	77
N.H.	13	6	5	3	3	0	30
Vt.	12	10	7	2	0	0	31
Total	101	70	126	41	36	21	395

Of the 395 school systems included in the study, 101 school systems enrolled less than 500 pupils, 70 school systems enrolled 500-999 pupils, 126 school systems enrolled 1000-2999 pupils, 41 school systems enrolled 3000-4999 pupils, 36 school systems enrolled 5000-9999 pupils, and 21 school systems enrolled more than 10,000 pupils.

Table 5. Pupil Enrollment Represented in Study

State	Enrollment in Systems Participating in Study	Total Public School Enrollment	Percentage of Total Enrollment Represented in Study
(1)	(2)	(3)	(4)
Mass.	564,987	780,893	72.3
Conn.	323,317	421,817	76.6
R.I.	91,273	121,000	75.4
Maine	80,153	186,318	43.0
N.H.	34,904	94,306	37.0
Vt.	25,487	69,717	36.5
Total	1,120,121	1,674,051[a/]	66.9

a/Compiled from enrollment data provided by State Departments of Education for 1957-1958 school year.

Over one million pupils were enrolled in the school systems participating in the study, as shown in Table 5. This represents 66.9 per cent of the total number of pupils enrolled in the New England public schools.

4. Reliability

General considerations.-- Although reliability generally refers to tests, this quality is equally important in inquiry forms. Consistency stems from two sources:[1/] (1) the situation in which the instrument is used, and (2) the instrument itself.

The instrument.-- The quality of the individual items and the length of the instrument are the principal factors which affect the

1/Victor H. Noll, Introduction to Educational Measurement, Houghton Mifflin Company, Boston, 1957, p. 69.

reliability of the inquiry form.[1/] Great care was exercised in constructing individual items and the over-all instrument. Through try-outs and other means described in a preceding section, a high degree of clarity in terms, items, and directions was obtained. The inquiry form consists of thirteen pages, including 43 items and 619 possible responses, assuring comprehensive coverage of the universe of questions which might be asked about the measurement and evaluation program. The avoidance of ambiguity in terms and the length of the form contribute materially to its reliability.

Inquiry form returns.-- Two methods were used to determine the consistency of responses on the returned inquiry form, and the extent to which the respondent supplied the information called for: (1) the form was filled in twice by a selected sample of respondents, and (2) a personal follow-up was made of another selected sample of respondents.

The second inquiry form was sent to sixteen selected school systems two weeks after the original forms had been returned. Four of the respondents receiving these forms were requested to complete Part I; four, Part II; four, Part III, and four, Part IV. These responses were then compared with the responses provided by these individuals on the original inquiry form. With the exception of 27 responses, where a slight variance was noted, all items completed by the sixteen respondents were the same as those received in the original inquiry form. This reliability study involved receiving an aggregate total of 172 items containing 2476 possible responses. All parts of the inquiry

1/Noll, op. cit., p. 69.

form revealed the same degree of consistency.

In the personal follow-up, the investigator sat down with six respondents and discussed their testing and evaluation program, using items taken at random from the inquiry form. The investigator noted these verbal responses on an inquiry form. These were later compared with the responses received on the original inquiry form. All responses were identical.

Conclusions.-- The inquiry form used in the study seems to be a reliable and valid data-gathering device.

5. Treatment of Data

Tabulating inquiry forms.-- Prior to printing, the inquiry form was coded for IBM analysis. Each returned inquiry form was hand-coded by the investigator. Forms were then forwarded to the IBM Laboratory at Boston University. This facility punched IBM tabulating cards for each inquiry form and provided a statistical analysis of the raw data. Items calling for free responses were hand-tabulated by the investigator.

Analysis of data.-- All data were tabulated in terms of raw scores and then converted to percentages. For the purpose of analysis and interpretation, the results are reported in tabular form. Each table is accompanied by a summary and an interpretation of the data. For the sake of comparison and analysis, the responses are reported according to enrollment categories as follows:

(1) school systems enrolling under 500 pupils

(2) school systems enrolling 500-999 pupils

(3) school systems enrolling 1000-2999 pupils

(4) school systems enrolling 3000-4999 pupils

(5) school systems enrolling 5000-9999 pupils

(6) school systems enrolling over 10,000 pupils

(7) responses for the total New England sample.

Analysis of data may be found in Chapter V.

CHAPTER V

TESTING AND EVALUATION PRACTICES IN THE NEW ENGLAND PUBLIC SCHOOLS

1. Introduction

Preliminary considerations.-- This study is concerned with the status of the pupil-centered testing and evaluation program in the New England public schools for the 1957-1958 school year. It analyzes, by means of an inquiry form,[1/] the following aspects of system-wide practices within the local school system: (1) organizing and administering the program, (2) tests and other measuring devices used in the program, (3) organizing and reporting the results of the program, and (4) administrative and supervisory uses made of test results. The following areas of pupil behavior are considered in the appraisal program: (1) aptitudes, (2) achievement, (3) interest, and (4) personal-social adjustment. The study is not concerned with evaluation practices carried on by individual classroom teachers nor with the technical aspects or complexities of test construction or nontesting devices.

It is the purpose of this chapter to report in detail the status of the testing and evaluation programs in the New England public schools as recorded in the responses provided in the inquiry form received from 395 school systems. The reader is referred to Inquiry

1/See Appendix C.

Form Returns[1/] for a complete breakdown of the participating schools.

For the purpose of analysis and interpretation, the results are reported in tabular form. Each table is accompanied by a summary and interpretation of the data reported. Responses are also reported in the tables according to enrollment categories to allow comparisons and further analysis. Tables including data which depict the testing and evaluation practices represent the major part of the present chapter.

Use of tables.-- Factors concerned with the testing and evaluation program were tabulated in terms of raw scores and then converted to percentages. All the responses in the inquiry form have been treated in this manner. For the purpose of analysis and comparison, the school systems responding have been grouped according to enrollment categories. In addition to enrollment categories, the responses for the total New England sample are also included in the tables. Tabular material is organized in this chapter according to the column headings noted below:

Column Heading	School Category
I	School systems enrolling less than 500 pupils
II	School systems enrolling 500-999 pupils
III	School systems enrolling 1000-2999 pupils
IV	School systems enrolling 3000-4999 pupils
V	School systems enrolling 5000-9999 pupils

1/Supra, Chapter IV, p. 196.

Column Heading	School Category
VI	School systems enrolling over 10,000 pupils
VII	Responses from the total New England sample

In most cases, responses listed in the tables are arranged in decreasing order of frequency, as reported in the total New England sample.

Attention is called to the fact that the summation of the percentages in the columns listed in the tables may be more or less than 100 per cent. This is caused by the following; (1) some items call for more than one response to be checked; and (2) some responses were not checked by the respondant.

2. Budgetary Considerations

Provisions made for program.-- To a large extent, the comprehensiveness of a testing and evaluation program is dependent upon the financial provisions made for its operation. If the program is considered an integral part of the total educational program, definite provisions must be made in the school budget for its operation (Table 6). Unless such provisions are made, there is great danger that the program will be relegated to a position of minor importance.

Table 6. Extent to Which the School System Makes Definite Provisions in the School Budget for the Pupil Appraisal Program

Extent of Budgetary Provisions	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Yes	70.0	78.5	80.0	95.0	90.9	90.9	80.7
No	30.0	18.6	20.0	5.0	9.1	9.1	19.3

Table 6 reveals that 80.7 per cent of the school systems make definite budgetary provisions for the pupil appraisal program. This is particularly true in the large school systems, as noted in Columns IV-VI. This may be due to the more formal budgetary procedures pursued by these systems.

The method of allocating the expenditure for the pupil appraisal program is another significant factor deserving consideration. Table 7 provides a breakdown of this element.

Table 7. Method of Allocating Expenditure for the Pupil Appraisal Program in the School Budget

Method of Allocation	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Separately to elementary and secondary schools	12.0	8.6	21.5	40.0	27.3	22.7	19.2
Separately to individual school buildings	--	5.7	2.3	10.0	12.1	--	3.8
Included in budget in one lump sum	49.0	52.8	33.1	45.0	54.5	50.0	44.5
Others	12.0	17.1	10.8	2.5	3.0	18.2	11.1

This breakdown reveals that 44.5 per cent of the schools include the expenditure in one lump sum, with 19.2 per cent of the schools making allocations separately to elementary and secondary schools. Other methods of allocation reported in order of decreasing frequency are: included in instructional and pupil supplies, guidance department budget, pupil personnel services account, general school supplies account, textbook account, and miscellaneous account. There appears to be no significant difference in method of allocation in relation to the size of the school system. "Allocation separately to elementary and secondary school" and "included in one lump sum" may imply more uniform apportionment on a system-wide basis. Schools report "others" infrequently. There may be danger in these schools of burying the pupil appraisal expenditure in some other account, such as textbook or miscellaneous.

Per pupil expenditure.-- The key to the extent of the program is the amount spent per pupil. Table 8 provides a breakdown of this expenditure.

Table 8. Amount Spent Per Pupil for the Pupil Appraisal Program (Exclusive of Salaries of Personnel)

Amount Spent Per Pupil	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Under 25¢	16.0	21.4	23.1	17.5	33.3	36.4	22.0
25¢-49¢	38.0	31.4	42.3	42.5	33.3	18.2	37.2
50¢-74¢	16.0	21.4	15.4	15.0	15.1	9.1	16.2
75¢-99¢	11.0	5.7	6.9	12.5	3.0	--	7.6
1.00-1.24	4.0	2.8	3.1	--	--	--	2.5
1.25-1.49	2.0	2.8	0.8	--	--	4.5	1.5
1.50-1.74	--	1.4	1.5	--	--	--	0.8
1.75-2.00	--	--	--	--	3.0	4.5	0.2
Above 2.00	--	4.2	0.8	2.5	--	--	1.3
No response	13.0	8.6	6.1	10.0	12.1	31.8	10.6

A wide variation in expenditure exists among the schools, ranging from less than 25 cents per pupil to above $2.00 per pupil. It is noted that the smaller schools generally spend more than the larger districts. This may be due to the higher general expenditure for adequate pupil personnel services required by this type of district. A further breakdown of the table reveals that the median expenditure for the New England public schools is 40 cents per pupil, with the 25th percentile at 25 cents and the 75th percentile at 62 cents per pupil. Although adequate national figures are not available, the 40-cent median expenditure compares favorably with the national average expenditure of 20-25

cents per pupil, as estimated by the test publishing industry. At any rate, when viewed in terms of the total per pupil educational cost, the expenditure for pupil appraisal is exceedingly low.

It is significant to note that a number of schools did not respond to this item. It may be assumed that these schools had no means of estimating the amount spent, or spent such a small sum that they failed to report. Financial aid provided in the National Defense Education Act of 1958 is not reflected in these figures. Undoubtedly this aid will increase the per pupil expenditure. It would be significant to compare the expenditure reported in this study and the expenditure reported in the 1960-1961 school year to determine the impact of this Act on the pupil appraisal program.

3. Organizing and Administering the Program

Responsibility for administering the program.-- The responsibility for planning, organizing, and supervising the system-wide appraisal program is considered in Table 9.

Table 9. Individual(s) Charged with the Responsibility of Planning, Organizing, and Supervising the System-Wide Pupil Appraisal Program

Individual	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Supt. of schools	70.0	60.0	66.1	47.5	15.1	18.2	57.2
School principal	48.0	60.0	37.7	35.0	18.2	18.2	40.5
Director of guid.	4.0	24.3	52.3	60.0	66.7	27.3	35.7
Assist. Supt.	1.0	2.8	1.5	10.0	30.3	22.7	6.1
School psychologist	1.0	4.2	3.8	12.5	18.2	13.6	5.3
Dir. of testing	1.0	4.2	3.8	7.5	12.1	22.7	5.3
Testing comm.	2.0	4.2	5.4	5.0	12.1	4.5	4.8
Dir. Pupil Personnel	-	-	3.1	5.0	9.1	18.2	3.3
Others	11.0	8.0	17.7	22.5	24.2	18.2	15.9

Wide diversity of practices is noted in the schools. In the total sample, the superintendent of schools, the building principal, and the guidance director, in the order named, are most frequently mentioned as directing the pupil appraisal program. This order may be due to the influence of the small school systems that normally do not have specialized personnel. This requires the superintendent and principal to assume this responsibility. In most small systems these two individuals are best qualified to exercise this responsibility. It should be noted that in some cases the responsibility for program direction is shared jointly by several individuals. The data further reveal that this responsibility shifts to specialists in guidance, pupil personnel, and testing as the school organization increases in size. In school systems enrolling over 1000 pupils, the guidance director most often is given the responsibility of directing the program.

Other individuals mentioned by respondents as being charged with this responsibility listed in order of decreasing frequency are: elementary supervisor, helping teacher, director of special services, classroom teacher, reading supervisor, supervisor of instruction, and administrative council.

Only 4.8 per cent of the systems reported a testing committee sharing this responsibility. This practice is in sharp contrast to the recommendations found in the literature, where committee organization is strongly advocated.

Many varying titles were reported for the individual(s) charged with the responsibility of coordinating and administering the system-wide testing and evaluation program (Table 10), running the whole gamut of professional personnel.

Table 10. Title of Person Charged with the Responsibility of Coordinating and Administering the System-wide Testing and Evaluation Program

Title	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Supt. of Schools	72.0	48.5	32.3	27.5	3.0	4.5	40.7
Director of Guid.	11.0	12.8	30.0	55.0	42.4	18.2	25.1
School Principal	26.0	20.0	3.8	2.5	6.1	-	12.1
Elementary Supv.	7.0	5.7	9.2	7.5	-	4.5	6.8
Assist. Supt.	-	1.4	1.5	2.5	18.2	22.7	3.8
Dir. of Pupil Personnel Services	-	-	3.1	2.5	15.1	18.2	3.5
Director of Testing	3.0	2.8	-	2.5	6.1	13.6	2.8

In addition to the titles included in Table 10, other titles mentioned in order of decreasing frequency are: helping teacher, psychological examiner, school psychologist, remedial supervisor, and reading consultant. Findings reveal that program direction centers in the superintendent's office in the small school system and shifts to the guidance officer as enrollments increase and specialized personnel becomes available. In the large city district the title reported most frequently is the Assistant Superintendent, 22.7 per cent; the Director of Guidance, 18.2 per cent; and the Director of Pupil Personnel Services, 18.2 per cent. A Testing Director was reported by 13.6 per cent of the city districts.

The data presented indicate that the responsibility for testing direction shifts from the general administrator to special services and pupil personnel functionaires as the school system increases in size and complexity. Although 40.7 per cent of the schools reported the title Superintendent, this high frequency is due primarily to the predominance of the smaller school system in the New England scene. As the consolidation movement increases, program direction should shift to specialized functionaires with such titles as Guidance Director, Director of Pupil Personnel, and Director of Testing. The large number of titles reported in this study is in keeping with varying titles reported in other surveys reviewed in Chapter II of the present study.

The problem of planning, organizing, and coordinating the testing and evaluation program is a technical job requiring special skills and knowledge. This implies special training in the field of tests and

measurements. Table 11 notes the extent of training of the individual coordinating and administering the program, as reported by respondents.

Table 11. Extent to Which the Person Responsible for Coordinating the Evaluation Program Holds an Advanced Degree with Specialization in the Field of Tests and Measurements

Extent Advanced Degree Held	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Yes	34.0	27.1	30.8	45.0	69.7	63.6	37.5
No	66.0	47.1	33.8	27.5	15.2	18.2	44.0

Findings indicate that only 37.5 per cent of the total sample have a person with an advanced specialized degree in charge of the program. Although the larger system reported a higher degree of training than the smaller system, only 63.6 per cent of the large city systems reported that the person responsible for the program had an advanced degree, with specialization in the field of tests and measurements.

All writings recommend that the person responsible for directing the appraisal program be given time to carry out this function. In the small system, where it may not be feasible to employ a full-time specialist, the literature advocates providing released time for a functionaire to operate the program. In the large system it recommends that the director devote full time to this responsibility. Currect practices surveys reviewed in this study revealed that all the large cities reported a full-time director of the program. Table 12 reveals the status in New England of this area of concern.

Table 12. Proportion of Time of Person Responsible for Coordinating and Administering the System-Wide Appraisal Program Spent in Testing and Evaluation Activities

Proportion of Time	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Less than 25%	76.0	45.7	33.8	35.0	33.3	22.7	46.1
25% - 49%	8.0	11.4	16.4	7.5	21.2	18.2	12.9
50% - 74%	2.0	4.2	3.1	7.5	6.1	4.5	3.8
75% -100%	2.0	2.8	1.5	15.0	18.2	36.4	6.6

Approximately 46 per cent of the schools reported that the professional person directing the program spent less than 25 per cent of his time in testing and evaluation activities. Only 6.6 per cent of the schools reported a person spending full time in these activities. Only 36.4 per cent of the large city school systems, enrolling over 10,000 pupils employ a full-time person in this area. It is expected that the small school will not employ a full-time specialist but will provide some released time for a functionary to carry out the necessary activities required in directing the appraisal program. This should increase proportionately as the enrollment increases. Data presented in this study indicate that sufficient time is not provided at present for this work.

The literature recommends that a school-wide committee, representing a cross section of the total staff, assist in developing plans for the appraisal program. Committee composition is depicted in Table 13.

Table 13. Composition of the Testing and Evaluation Committee

Committee Member	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Schl. Principal	16.0	27.1	33.8	30.0	27.3	45.4	27.8
Supt. of Schools	19.0	30.0	33.8	25.0	21.2	27.3	27.1
Dir. of Guidance	1.0	7.1	27.7	27.5	39.4	27.3	18.7
Classroom Teacher	12.0	12.8	13.1	10.0	24.2	31.8	17.0
Guid. Counselor	-	4.2	6.9	17.5	24.2	27.3	8.3
Dir. of Testing	-	5.7	2.3	2.5	9.1	18.2	5.8
Asst. Supt.	-	4.2	3.1	5.0	21.2	27.3	5.6
Schl. Psychol.	-	2.8	4.6	12.5	15.1	27.3	5.6
Dept. Head	-	1.4	3.1	3.0	12.1	18.2	4.3

Table 13 indicates that the Testing and Evaluation Committee represents a fairly good cross section of the school staff. In some instances, however, the committee appears somewhat top-heavy administratively. This is particularly true in categories V and VI, where the percentage frequency is rather heavy for Directors, Principals, and Superintendents. Other personnel classifications listed as committee members in order of decreasing frequency are: elementary supervisor, reading specialist, curriculum coordinator, director of pupil personnel, helping teacher, and supervisor of secondary education. Two school systems reported parents and pupils as committee members. The literature discloses that the successful program is one in which the teachers as well as other staff members participate in carrying out many of the functions. Although committees include all staff members, added representation from the classroom teachers might aid the general effectiveness and acceptance of the testing and evaluation program.

Attention is called to the fact that the testing committee is not common to most New England schools. Only 4.8 per cent of the school systems reported a committee organization (Table 9).

Scheduling the system-wide testing program.-- Scheduling is an administrative problem requiring decision prior to test administration (Table 14).

Table 14. Person Charged with the Responsibility of Scheduling the System-Wide Testing Program

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Supt. of Schools	69.0	50.0	51.5	40.0	12.1	9.1	48.9
Schl. Principal	45.0	60.0	43.1	40.0	18.2	22.7	43.0
Dir. of Guidance	2.0	12.8	44.6	60.0	48.5	31.8	29.4
Assist. Supt.	1.0	2.8	3.1	7.5	24.2	22.7	5.8
Classroom Teacher	9.0	5.7	2.3	5.0	9.1	4.5	5.6
Dir. of Testing	-	5.7	1.5	7.5	18.2	22.7	5.1
Com. on Testing	-	-	3.1	10.0	6.1	4.5	2.3
Other	9.0	-	6.1	25.0	24.2	-	15.4

The superintendent of schools, building principal, and guidance director most frequently schedule the system-wide program. Other individuals, listed by respondents in decreasing order of frequency, who carry out this responsibility are: elementary supervisor, director of special pupil services, helping teacher, school psychologist, curriculum coordinator, supervisor of instruction, administrative council, and reading consultant. The same general pattern, noted in other areas, of shift of responsibility from the general administrative office to

specialized personnel, as the school system increases in size, is apparent from Table 14. As revealed in the table, this responsibility is generally carried out by a central office staff member(s) and/or building principal(s). This should afford more continuity on a system-wide basis. As was the case in other areas of responsibility, scheduling may be a shared responsibility.

Choosing tests.-- Tests used in the system must be most appropriate for the intended purpose. Tests must be selected with care and involve those staff members who are best qualified to assume this task. The literature advocates that selection be done as a coordinated group enterprise involving measurement specialists, curriculum specialists, administrators, and classroom teachers. Surveys of current practices revealed that guidance personnel, principals, and the classroom teacher participated in the selection process involving academic aptitude and achievement tests.

Group academic aptitude and achievement tests represent the core of the testing program. The person responsible for choosing these measures are shown in Tables 15 and 16.

Table 15. Person Responsible for Choosing Group Intelligence Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Supt. of Schools	78.0	57.0	53.1	32.5	12.1	4.5	53.2
Schl. Principal	38.0	41.4	33.8	45.0	15.1	9.1	34.4
Guid. Personnel	6.0	24.3	49.9	52.5	45.4	22.7	32.7
Elementary Supv.	4.0	12.8	15.4	20.0	18.2	9.1	12.4
Testing Comm.	3.0	4.2	9.2	10.0	18.2	18.2	8.1
Classroom Teacher	6.0	10.0	5.4	10.0	6.1	9.1	7.1
Teacher Committee	7.0	2.8	0.7	10.0	9.1	13.6	5.1
Assist. Supt.	-	2.8	2.3	5.0	24.2	13.6	4.5
Dir. of Testing	-	4.2	3.8	2.5	12.1	18.2	4.3

A study of the percentages indicates that several school functionaries share in this responsibility in most cases. The superintendent of schools, building principal, and guidance personnel, in the order named, most frequently choose group intelligence and standardized achievement tests. Others, in order of decreasing frequency, are: school psychologist, director of special pupil services, administrative council, helping teacher, curriculum director, college testing service.

Table 16. Person Responsible for Choosing Standardized Achievement Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Supt. of Schools	81.0	64.3	57.7	32.5	15.1	4.5	55.7
Schl. Principal	38.0	41.4	42.3	52.5	18.2	13.6	38.5
Guid. Specialist	5.0	21.4	44.6	50.0	39.4	36.4	30.1
Elementary Supv.	5.0	12.9	14.6	20.0	18.2	18.6	12.6
Classroom Teacher	7.0	12.9	11.5	20.0	24.2	9.1	12.4
Testing Committee	2.0	2.8	9.2	12.5	18.2	27.3	8.3
Teacher Committee	10.0	5.7	3.8	12.5	12.1	9.1	7.6
Assist. Supt.	-	2.8	1.5	7.5	27.3	13.6	4.8
Dir. of Testing	-	4.2	1.5	15.0	12.1	18.2	3.8

The superintendent and the building principal assume this responsibility in the small system, with the guidance specialist being most frequently mentioned in systems enrolling above 3000 pupils. The same general pattern of frequency responses is noted in group intelligence and standardized achievement tests, with the exception of the classroom teacher and testing committee, the former being mentioned more frequently in choosing standardized achievement tests and the latter in the case of group intelligence tests. In the larger school systems this responsibility is carried largely by special services and instructional functionaries and involves several groups. Others, in order of decreasing frequency, are: director of special pupil services, school psychologist, administrative council, and helping teacher. Attention is also called to the fact that some schools made use of college testing services in choosing group intelligence tests.

Individual intelligence tests are generally used for retesting purposes and not on a system-wide basis in most cases. Responsibility for choosing these measures is noted in Table 17.

Table 17. Person Responsible for Choosing Individual Intelligence Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Supt. of Schools	50.0	35.7	28.5	25.0	12.1	4.5	32.1
Guid. Personnel	6.0	24.3	43.8	52.5	39.4	27.3	30.4
Schl. Principal	30.0	25.7	18.5	27.5	6.1	4.5	21.8
Elementary Supv.	2.0	12.8	13.1	20.0	12.1	-	10.1
Teacher Committee	8.0	2.8	5.4	10.0	6.1	13.6	6.6
Classroom Teacher	5.0	7.1	2.3	7.5	3.0	4.5	4.6
Dir. of Testing	-	5.7	2.3	2.5	12.1	18.2	4.0
Testing Committee	2.0	2.8	3.1	5.0	9.1	9.1	3.8
Assist. Supt.	-	4.2	0.8	2.5	15.1	4.5	2.8

Approximately 32 per cent of the school systems reported the superintendent of schools responsible for selection, followed by the guidance personnel at 30.4 per cent and the building principal at 29.8 per cent. As was noted in previous tables, responsibility for this function shifts from the general administration to specialized personnel as the systems grow in size. This responsibility is most frequently handled by guidance or special service functionaries in the larger school systems. Outside specialists and state agencies also aid in this work, particularly in the small rural school district where no specially trained pupil personnel are available. Others, in

order of decreasing frequency, are: school psychologist, director of special pupil services, helping teacher, administrative council, reading consultant, outside specialist, and state agency.

Special aptitude tests, interest inventories, and personality measures are used extensively for guidance purposes and are somewhat specialized in nature. Tables 18, 19, and 20 indicate functionaries responsible for their choice.

Table 18. Person Responsible for Choosing Special Aptitude Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Guidance Special.	5.0	25.7	62.3	67.5	72.7	40.9	41.5
Schl. Principal	37.0	41.4	23.5	30.0	3.0	9.1	28.3
Supt. of Schools	34.0	31.4	16.9	25.0	9.1	4.5	23.3
Elementary Supv.	1.0	8.6	7.7	12.5	3.0	9.1	6.3
Classroom Teacher	4.0	7.1	6.1	15.0	6.1	-	6.3
Teacher Committee	7.0	2.9	0.8	7.5	3.0	4.5	3.8
Dir. of Testing	-	4.2	1.5	15.0	12.1	18.2	3.5
Testing Committee	1.0	2.9	3.1	2.5	9.1	13.6	3.5
Assist. Supt.	-	2.9	0.8	5.0	15.2	4.5	2.8

Others, in order of decreasing frequency, are: director of pupil personnel, school psychologist, helping teacher, administrative council, elementary supervisor, college testing service, and state department agency.

Table 19. Person Responsible for Choosing Interest Inventories

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Guid. Specialist	5.0	27.1	60.8	65.0	72.7	50.0	41.5
School Principal	30.0	41.4	21.5	15.0	6.0	-	24.0
Supt. of Schools	24.0	24.3	13.5	15.0	3.0	4.5	17.0
Testing Committee	-	2.9	4.6	2.5	12.1	13.6	4.0
Teacher Committee	8.0	1.4	-	7.5	3.0	4.5	3.5
Classroom Teacher	4.0	4.2	0.8	5.0	3.0	-	2.8
Assist. Supt.	-	-	0.8	5.0	18.1	4.5	2.5
Elementary Supv.	-	4.2	3.0	-	3.0	4.5	2.3
Dir. of Testing	-	-	2.3	-	9.1	13.6	2.3

Others, in order of decreasing frequency, are: director of pupil personnel, school psychologist, helping teacher, administrative council, elementary supervisor, state agency, and college testing service.

Table 20. Person Responsible for Choosing Personality Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Guid. Specialist	5.0	22.8	51.5	55.0	45.4	31.8	33.4
Schl. Principal	26.0	31.4	17.7	22.5	3.0	-	20.5
Supt. of Schools	24.0	18.6	8.5	17.5	6.1	4.5	14.7
Elementary Supv.	1.0	5.7	3.8	10.0	-	4.5	3.8
Testing Committee	-	2.9	2.3	2.5	12.1	13.6	3.3
Teacher Committee	6.0	1.4	0.8	7.5	-	4.5	3.0
Classroom Teacher	3.0	11.4	2.3	2.5	3.0	-	2.8
Dir. of Testing	-	-	1.5	2.5	12.1	13.6	2.5
Assist. Supt.	-	-	-	5.0	15.1	4.5	2.0

Others, in order of decreasing frequency, are: school psychologist,

director of pupil personnel, administrative council, helping teacher, elementary supervisor, state agency, and college specialist.

Tables 18 to 20 reveal that the guidance specialist is mentioned most frequently as the functionary responsible for choosing special aptitude, interest, and personality measures. In all three tables this individual is followed by the building principal, superintendent, and elementary supervisor, in the order named. In all cases except the small school system where guidance specialists may not be available, selection responsibility most frequently is handled by guidance functionaries. This is to be expected, since these measures have primary use in the guidance process. Use is also made of outside sources in the form of state agencies and college testing services in some school systems, particularly the small school district. Attention is also called to the similarity of pattern of responses in choosing these three measures.

Responses noted in Tables 15 to 20 reveal that test selection is usually done by a group of school functionaries rather than by one individual. This is particularly true of the system large enough to employ specialists in the area of pupil personnel and instruction. Although the general administrators are most frequently mentioned, particularly in group intelligence and standardized achievement tests, the responsibility shifts to specialists in the guidance and pupil personnel fields as the school system increases in size. Similarity in patterns of percentage responses are noted for selection of group intelligence and standardized achievement tests, and for selection of

special aptitude, interest, and personality measures. General administrators play a more important role in the former and guidance personnel in the latter case.

Administering tests.-- Tests must be carefully administered with meticulous attention to directions. The literature states that individual tests and tests for special purposes should be administered by specially trained personnel. In the general group testing program it advocates that the classroom teacher administer the tests. Tables 21 and 22 indicate practices in administering group intelligence and standardized achievement tests.

Table 21. Person Responsible for Administering Group Intelligence Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Classroom Teacher	72.0	72.8	63.1	70.0	57.6	54.5	68.3
Schl. Principal	30.0	21.4	17.7	27.5	15.1	12.0	22.3
Guid. Specialist	6.0	12.9	32.3	37.5	39.4	9.1	22.3
Elementary Supv.	1.0	8.6	13.0	10.0	3.0	-	7.3
Schl. Psychologist	-	1.4	0.8	5.0	18.2	12.0	3.5
Dir. of Testing	-	5.7	2.3	5.0	6.0	3.0	3.0
Dept. Chairman	-	-	1.5	2.5	18.2	-	1.0
Other[a/]	3.0	4.2	6.1	15.0	9.1	3.0	6.6

a/In order of decreasing frequency: superintendent of schools, outside specialist, college testing service, state department specialist, assistant principal, psychological examiner, and reading consultant.

Table 22. Person Responsible for Administering Standardized Achievement Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Classroom Teacher	77.0	90.0	82.3	70.0	84.8	45.4	82.0
Schl. Principal	25.0	17.1	15.4	32.5	9.1	22.7	19.7
Guid. Specialist	2.0	8.6	20.8	37.5	30.3	9.1	15.7
Elementary Supv.	1.0	5.7	8.5	10.0	3.0	-	5.3
Dir. of Testing	-	2.8	0.8	7.5	6.1	9.1	2.5
Schl. Psychologist	-	1.4	0.8	-	9.1	9.1	1.8
Dept. Chairman	-	1.4	0.8	5.0	3.0	4.5	1.5
Other[a]	4.0	1.4	4.6	10.0	6.1	-	4.3

a/In order of decreasing frequency: reading supervisor, superintendent of schools, college students, and college testing service.

Tables 21 and 22 reveal that the classroom teacher most frequently administers group intelligence and standardized tests. This is particularly true of the latter measure where 82 per cent of the school systems report that the classroom teacher administers these tests. The building principal and guidance specialist, reported in 22.3 per cent of the systems, as well as the elementary supervisor and school psychologist, are mentioned prominently as carrying on this responsibility. The use of the administrator or specialized personnel in administering these measures may imply large-scale testing for survey or special purposes in the systems reporting the use of these functionaries. This plan tends to increase accuracy, uniformity, and efficiency. This plan is reflected in the responses provided by the larger schools, particularly in intelligence testing (Table 21).

Individual intelligence testing should be given by specially trained personnel. This is reflected by the responses indicated in Table 23.

Table 23. Person Responsible for Administering Individual Intelligence Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Guid. Specialist	6.0	17.1	10.8	32.5	63.6	77.3	24.0
Schl. Psychologist	5.0	8.6	3.1	5.0	18.2	13.6	19.2
Classroom Teacher	26.0	17.1	0.8	-	3.0	-	15.9
Schl. Principal	20.0	17.1	19.2	2.5	3.0	-	11.6
Elementary Supv.	1.0	10.0	5.4	12.5	6.1	-	6.6
Dir. of Testing	1.0	8.6	9.2	10.0	6.1	-	5.6
Dept. Chairman	-	1.4	4.1	3.5	2.7	0.5	0.8
Other[a]	13.0	8.6	13.1	17.5	-	4.5	15.4

a/In order of decreasing frequency: reading specialist, outside specialist, superintendent of schools, helping teacher, assistant principal, state department psychologist, teachers college personnel, and school social worker.

Twenty-four per cent of the systems reported that the guidance specialists performed this task, with 19.2 per cent reporting the school psychologist. These individuals are undoubtedly highly trained in this field. Attention is called to the fact that these functionaries, together with the director of testing, dominate the scene in the larger systems (categories IV, V, VI). The classroom teacher and building principal are mentioned in the small district. It is also significant to note that outside assistance, state department psychologist, and teachers college personnel aid in this specialized work.

It is noted that guidance functionaries are most frequently mentioned in administering special aptitude tests, interest inventories, and personality measures (Tables 24, 25, 26).

Table 24. Person Responsible for Administering Special Aptitude Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Guid. Specialist	4.0	17.1	60.0	60.0	51.5	40.9	36.7
Classroom Teacher	24.0	30.0	15.4	15.0	21.2	31.8	21.5
Schl. Principal	28.0	28.6	5.4	17.5	-	4.5	15.9
Schl. Psychologist	-	5.7	1.5	17.5	24.2	27.3	6.8
Dir. of Testing	-	5.7	0.8	10.0	15.1	9.1	4.0
Elementary Supv.	-	2.8	6.1	7.5	-	-	3.3
Dept. Chairman	-	2.8	0.8	2.5	9.1	-	1.8
Other[a]	3.0	1.4	6.9	20.0	12.1	-	6.6

a/In order of decreasing frequency: college testing service, reading supervisor, state employment service, assistant principal, superintendent of schools.

Table 25. Person Responsible for Administering Interest Inventories

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Guid. Specialist	5.0	20.0	60.8	52.5	63.6	50.0	38.2
Classroom Teacher	23.0	34.3	20.0	17.5	24.2	22.7	23.5
Schl. Principal	16.0	24.3	6.1	10.0	-	-	11.4
Schl. Psychologist	-	2.8	2.3	15.0	21.2	27.3	6.1
Dir. of Testing	-	2.8	0.8	7.5	9.1	9.1	2.8
Dept. Chairman	2.0	1.4	-	2.5	6.1	-	1.5
Elementary Supv.	-	1.4	2.3	-	-	-	1.0
Other[a]	2.0	1.4	6.1	12.5	6.1	4.5	5.1

a/In order of decreasing frequency: college testing service, assistant principal, helping teacher, outside testing specialist, and superintendent of schools.

Table 26. Person Responsible for Administering Personality Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Guid. Specialist	4.0	15.7	48.4	47.5	33.3	27.8	28.9
Classroom Teacher	18.0	21.4	13.1	15.0	12.1	9.1	15.7
Schl. Psychologist	-	5.7	3.8	20.0	33.3	36.4	9.1
Schl. Principal	14.0	15.7	2.3	12.5	-	-	8.3
Dir. of Testing	-	4.2	0.8	7.5	12.1	4.5	3.0
Elementary Supv.	1.0	1.4	0.8	5.0	-	-	1.3
Dept. Chairman	1.0	1.4	-	-	6.1	-	1.0
Other[a]	2.0	1.4	5.4	15.0	9.1	9.1	5.3

a/In order of decreasing frequency: outside specialist, superintendent of schools, and helping teacher.

Although the measures referred to in Tables 24 to 26 are somewhat specialized in nature, respondents indicate that the classroom teacher

plays a prominent role in administering these measures. To administer these measures effectively, teachers must be thoroughly briefed and trained in this work. This is particularly true in the case of personality tests.

Data included in Tables 21 to 26 reveal that tests are administered by teachers, administrators, and specialists. The classroom teacher generally administers group intelligence and standardized achievement tests, while the specialist most frequently administers the individual intelligence tests, special aptitude tests, interest inventories, and personality measures. The findings appear to present two alternates: the administration of group tests should be handled by only a few specially trained personnel if a considerable amount of testing is to be done, or teachers should be given thorough training in the administration of such measures.

Scoring tests.-- Rapid and accurate scoring of tests is a prerequisite for any testing program. Findings pertaining to scoring procedures are contained in Tables 27 to 31.

Table 27. Procedure Followed in Scoring Group Intelligence Tests

Enrollment Category	Percentage of Responses According to Procedure					
	Elementary Schools			Secondary Schools		
	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person
(1)	(2)	(3)	(4)	(5)	(6)	(7)
I	90.0	1.0	5.0	42.0	1.0	5.0
II	92.8	2.8	4.2	55.7	2.8	8.6
III	83.2	2.3	18.5	53.1	3.1	37.7
IV	77.5	10.0	32.5	55.0	15.0	35.0
V	84.8	9.1	18.2	30.3	24.2	42.4
VI	77.3	13.6	27.3	27.3	18.2	22.7
VII	86.1	4.0	14.4	47.6	6.3	24.3

Table 28. Procedure Followed in Scoring Standardized Achievement Tests

Enrollment Category	Percentage of Responses According to Procedure					
	Elementary Schools			Secondary Schools		
	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person
(1)	(2)	(3)	(4)	(5)	(6)	(7)
I	91.0	1.0	2.0	41.0	-	6.0
II	85.6	4.2	7.1	48.5	2.8	15.7
III	83.2	1.5	22.3	48.4	3.8	42.3
IV	80.0	5.0	27.5	45.0	15.0	30.0
V	81.8	18.2	18.2	45.4	24.2	36.4
VI	72.7	27.3	13.6	31.8	22.7	22.7
VII	85.3	5.1	14.2	45.1	6.6	27.1

A study of Tables 27 and 28 reveals that local hand-scoring dominates the scene. This is particularly true at the elementary level. It is expected that local hand-scoring percentage responses will be high at this level because of the use of consumable hand-scored booklets

at the primary grades. Findings imply, however, that the majority of schools continue this practice at the upper grade levels. Although, in the interest of efficiency, it appears advisable to machine-score tests for the fourth grade and above, widespread use of local hand-scoring continues in the majority of schools.

Little use is made of local machine-scoring except in the case of the large city school system. The expense involved in rental may account for its infrequent use.

Some use is made of scoring services of outside agencies or persons. The larger schools report more use than the small district. Highest frequency of usage is reported by schools in categories III to V at the secondary level (Tables 27 and 28). Findings reveal that the small schools report little usage of outside scoring assistance. One would believe that these schools, because of limited specialized personnel, would make extensive use of such services. This apparently is not the case.

Special aptitude tests, interest inventories, and personality measures are not used as extensively as the measures referred to in Tables 27 and 28. In the majority of schools their use is limited to the secondary level. The procedures followed in scoring these measures are indicated in the following tables.

Table 29. Procedure Followed in Scoring Special Aptitude Tests

Enrollment Category	Percentage of Responses According to Procedure					
	Elementary Schools			Secondary Schools		
	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person
(1)	(2)	(3)	(4)	(5)	(6)	(7)
I	44.0	1.0	1.0	27.0	1.0	5.0
II	45.7	1.4	8.6	37.1	2.8	15.7
III	38.5	1.5	6.9	53.1	0.8	26.9
IV	45.0	-	17.5	57.5	5.0	25.0
V	39.4	3.0	1.0	45.4	15.1	30.3
VI	31.8	13.6	4.5	27.3	22.7	36.4
VII	41.5	2.0	6.3	42.0	4.0	21.0

Table 30. Procedure Followed in Scoring Interest Inventories

Enrollment Category	Percentage of Responses According to Procedure					
	Elementary Schools			Secondary Schools		
	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person
(1)	(2)	(3)	(4)	(5)	(6)	(7)
I	29.0	-	2.0	23.0	1.0	7.0
II	32.8	1.4	4.2	37.1	4.2	11.4
III	33.8	1.5	4.6	57.7	0.8	7.7
IV	22.5	-	7.5	57.5	5.0	12.5
V	30.3	-	-	66.7	9.1	15.1
VI	31.8	4.5	-	50.0	18.2	22.7
VII	30.9	1.0	3.5	45.6	3.5	16.4

Table 31. Procedure Followed in Scoring Personality Tests

Enrollment Category	Percentage of Responses According to Procedure					
	Elementary Schools			Secondary Schools		
	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person	Local Hand-Scoring	Local Machine Scoring	Outside Agency or Person
(1)	(2)	(3)	(4)	(5)	(6)	(7)
I	31.0	-	2.0	21.0	1.0	4.0
II	35.7	-	2.8	35.7	2.8	7.1
III	33.8	0.8	4.6	46.1	0.8	4.6
IV	27.5	-	5.0	52.5	2.5	10.0
V	24.2	3.0	-	48.5	9.1	12.1
VI	31.8	4.5	-	45.4	4.5	9.1
VII	31.9	0.8	3.0	33.7	2.3	10.1

Local hand-scoring is the procedure most frequently reported by schools for scoring special aptitude tests, interest inventories, and personality measures (Tables 29, 30, 31). Since many of these measures lend themselves more readily to hand-scoring, this is to be expected. Except for special aptitude tests, only minor use is made of outside agencies. The larger systems are the more frequent users of these agencies.

Findings reveal that the majority of tests are scored locally by hand. This is consistent with the findings of other surveys of current practices.

Since local hand-scoring is reported in the majority of schools, consideration is given to this procedural aspect. Tables 32 and 33 treat this process.

Table 32. Person Who Generally Locally Hand-Scores Group Intelligence Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Classroom Teacher	77.0	72.8	67.7	72.5	57.6	63.6	70.4
Guid. Specialist	5.0	21.4	26.9	30.0	24.2	13.6	19.7
Principal	17.0	15.7	8.5	20.0	15.1	4.5	13.4
Clerk	10.0	10.0	10.8	2.5	21.2	36.4	11.9
Teacher Committee	2.0	2.8	3.8	-	6.1	-	2.8
Others[a/]	4.0	5.7	6.9	12.5	6.1	13.6	7.1

a/In order of decreasing frequency: elementary supervisor, school psychologist, testing department, reading teacher, committee from local women's club, college seniors, and hired outside help.

Table 33. Person Who Generally Locally Hand-Scores Standardized Achievement Tests

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Classroom Teacher	81.0	82.8	83.8	82.5	84.8	63.6	81.3
Principal	16.0	11.4	13.1	12.5	6.1	4.5	8.9
Guid. Specialist	1.0	7.1	4.6	17.5	9.1	9.1	8.9
Clerk	9.0	7.1	2.3	-	15.1	22.7	7.8
Teacher Committee	3.0	2.8	8.2	2.5	3.0	-	3.3
Others[a/]	3.0	4.2	5.4	7.5	3.0	-	4.0

a/In order of decreasing frequency: testing department, elementary supervisor, hired outside help, committee from local women's club, superintendent of schools, school psychologist, remedial reading teacher, and college seniors.

Tables 32 and 33 indicate that teachers generally do the hand-scoring. Other surveys of current practices also report that the

majority of tests are scored by teachers. Attention is called to the fact that in both intelligence and standardized achievement tests, professional personnel are used far more extensively than clerks. The larger systems make more use of clerks than the small schools. Current practices reveal that teachers and professional staff members are carrying the burden in this task rather than the nonprofessional.

Schools report that a variety of people aid in this task. These individuals must be trained and properly supervised if accuracy of results are obtained. This implies an in-service program, particularly in the case of lay people, directed toward the scorers.

Tests should be scored first by machines or self-scoring stencils. If this is not possible, they should be scored by specially trained clerks. Only if neither of these is possible should it be necessary for teachers themselves to score the tests. Current practices, as revealed in the findings, almost reverse this principle.

Using outside services.-- Testing involves many technical considerations requiring specialized personnel. It is not possible for many school districts to employ such functionaries as regular staff members. Outside consultative services are available to aid these schools in testing problems. The extent to which schools use such services is indicated in Table 34.

Table 34. Extent to Which the School System Uses the Services of Outside Testing Specialists on a Consultative Basis

Extent	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Regularly	3.0	12.8	23.0	32.5	18.2	13.6	16.2
Occasionally	56.0	48.5	41.5	45.0	45.5	45.5	47.3
Never	41.0	38.6	33.8	22.5	33.3	40.9	35.7

Only 16.2 per cent of the systems report regular use of outside testing specialists. The small schools, where such services may be most needed, make only infrequent use of these services. A high percentage, 35.7, report no use of such services. Further investigation of this problem attempted to determine if school systems would use the services of a consultant if one could be made available (Table 35).

Table 35. Extent to Which the School System Would Use the Services of a Consultant if One Could Be Made Available

Response	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Yes	75.0	70.0	60.0	55.0	39.4	18.2	61.3
No	24.0	30.0	39.2	45.0	60.6	81.8	38.5

The majority of systems indicate that they would use such services if they were available. Findings reveal that the small system evidences greatest need for such services, with decreasing frequency of need noted as the system increases in enrollment and number of

specialized personnel among its staff. Superintendents indicate that budgetary limitations seriously handicap securing such services and evidence an interest in sharing consultative services with neighboring communities on a pooling basis. Need for specialized and individual testing, and assistance in research, is also evidenced.

Several implications are noted from the findings. There is a need for colleges and state departments to make additional specialists available for consultative work in the field. Several small school districts could join together and employ a specialist on a sharing basis. Greater service could be provided by field representatives of commercial test publishers. Funds available through the National Education Defense Act may be used by State Departments of Education for this work. A follow-up study could reveal the extent to which this act has aided this felt need.

The extent to which schools use an outside testing agency or service to assist in the basic testing program is considered in Table 36.

Table 36. Extent to Which the School System Makes Use of the Services of an Outside Testing Agency or Service

Response	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Yes	35.0	51.4	63.8	75.0	63.6	45.4	54.4
No	65.0	48.5	36.1	25.0	36.4	54.5	45.6

Approximately one half of the schools report the use of an outside service or testing agency to assist in the basic testing program. Greatest use of such services is made by schools in categories III to V. The greater proportion of these systems is in urban communities. Further breakdown of this service is noted in Tables 37 and 38.

Table 37. Sponsorship of Outside Testing Service

Sponsor	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
College or Univ.	19.0	27.1	45.4	65.0	42.4	22.7	35.9
Commercial Service Agency	8.0	14.3	20.0	22.5	18.2	18.2	15.9
State Dept. of Education	12.0	8.6	3.8	10.0	-	4.5	7.1
Others[a/]	10.0	10.0	11.5	12.5	21.2	-	11.1

a/In order of decreasing frequency: outside specialists on fee basis, state employment service, commercial hygiene clinic, and neighboring school department.

Table 38. Testing Services Provided by the Outside Agency or Service

Services Provided	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test Scoring	9.0	24.3	32.3	55.0	42.4	31.8	28.1
Complete "packaged" program	15.0	25.7	34.6	37.5	21.2	9.1	25.8
Consultant and Field Service	20.0	27.1	15.4	27.5	9.1	9.1	19.0
Tests provided on rental basis	4.0	4.2	16.1	20.0	18.2	22.7	11.9
Others[a/]	7.0	-	8.5	-	9.1	-	5.3

a/In order of decreasing frequency: individual intelligence testing, interpreting tests at teachers' meetings, tabulation and reporting, statistical study of test results, scoring, psychological and projective services.

Schools making use of these outside agencies report that colleges or universities most frequently sponsor this service, followed by commercial services agencies and state departments of education (Table 38). Test scoring and "packaged" programs are the services most frequently provided. Findings reveal that state departments of education are not as active in this area as is the case in other sections of the country. It is also evident that assistance in the mechanical aspects (test scoring) is more frequently used than consultative aid in dealing with major testing problems. Few systems make use of field service to aid the staff in test interpretation and usage.

State-wide and large-scale testing programs appear to be on the increase. A survey conducted in 1958 reported 31 state testing programs in the United States.[1/] In addition, the literature reveals that national and large-scale programs are on the increase. The extent to which New England schools participate in state or regional programs is indicated in Table 39.

Table 39. Extent to Which the School System Participates in State-Wide or Regional Testing Programs

Extent	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Regularly	18.0	17.1	7.7	12.5	9.1	9.1	12.7
Occasionally	31.0	45.7	52.3	37.5	63.6	36.4	44.3
Never	43.0	34.3	35.4	45.0	18.2	54.5	43.0

1/Carroll H. Miller, "Guidance and Programs of Testing," in Understanding Testing, United States Department of Health, Education, and Welfare, Office of Education, United States Government Printing Office, Washington, D. C., 1960, p. 16.

Findings reveal that state-wide or regional programs have not had the impact on New England as on other sections of the country, with only 12.7 per cent reporting regular participation. Undoubtedly, the influence of the National Defense Education Act of 1958 as well as other large-scale programs, as the National Merit Scholarship Examinations, will tend to extend these programs in New England. This will require the staff to become more informed and discriminating in regard to the character and purpose of such programs to lessen the possibility of domination by any one agency.

Personnel orientation.-- The efficacy of a test depends upon the skill with which it is administered and the results interpreted and used. All personnel using and interpreting tests and resultant data must be adequately oriented. This implies an organized in-service program in the measurements area (Table 40).

Table 40. Provisions Made for In-Service Program

Response	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Yes	45.0	51.4	60.8	77.5	75.5	72.7	58.7
No	55.0	48.5	39.2	22.5	24.2	27.3	41.3

Data provided in Table 40 represent responses to the question, "Is there an organized in-service training program to assist teachers in interpreting and using tests and other evaluation devices in carrying out their regular instructional functions?" Findings reveal that

schools are woefully lacking in this program. Approximately 41 per cent report no program. Only 45 per cent of the small school systems report an organized in-service program, while 72.7 per cent of the large city school systems report such a program. It is recalled that in small schools the teachers assume the greater degree of responsibility in testing. Because of this fact and the dearth of specialized personnel, it may be assumed that the greatest need for such a program exists in these schools. Surveys of current practices reviewed in Chapter II reported 75 per cent of the large city systems[1/] and 58 per cent of the public school members of the Educational Records Bureau[2/] as having in-service programs. Practices in New England are consistent with the national scene. Apparently this is a national problem requiring attention.

Data pertaining to certain aspects of the in-service program, reported by systems who have such programs, are contained in Tables 41 to 43.

1/Arthur E. Traxler, "The Status of Measurement and Appraisal Programs of Large City School Systems," 1953 Achievement Testing Program in Independent Schools and Supplementary Studies, Educational Records Bureau, Bulletin No. 61, Educational Records Bureau, New York, 1953, p. 86.

2/Robert D. North, "Testing Programs of Public School Members of the Educational Records Bureau--Report of a Questionnaire Survey," 1956 Achievement Testing Program in Independent Schools and Supplementary Studies, Bulletin No. 68, Educational Records Bureau, New York, 1956, p. 30.

Table 41. Responsibility for Directing and Coordinating the In-Service Training Program

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Supt. of Schools	36.0	31.4	23.1	12.5	6.1	9.1	24.5
School Principal	21.0	27.1	28.5	25.0	21.2	4.5	24.0
Dir. of Guidance	2.0	10.0	33.8	42.5	39.4	22.7	22.3
Elementary Supv.	2.0	8.6	10.0	32.5	15.1	27.3	11.1
Assist. Supt.	-	-	-	7.5	18.2	18.2	3.3
Dir. of Testing	-	1.4	2.3	2.5	6.1	27.3	3.3
Others[a]	5.0	1.4	6.9	15.0	39.4	18.2	9.6

a/In order of decreasing frequency: director of pupil services, director of psychological services, head counselor, helping teacher, college extension courses, reading consultant, curriculum coordinator, vice-principal, and director of secondary education.

Various functionaries carry out the responsibility of directing the measurement in-service program (Table 41). The superintendent of schools and building principal are mentioned most frequently. This is due to the predominance of the small school systems where the general administrators are usually best qualified to carry out this responsibility. As is the case in other aspects of the program, this responsibility shifts to specialized personnel as the system grows in size and complexity. The director of guidance assumes this responsibility in the larger systems (categories III to V). The elementary supervisor, as well as the director of testing, is active in this work in the larger system. Responses also indicate that in many cases this responsibility is shared jointly by several persons.

Table 42. Means Used in the In-Service Training Program

Means Used	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Group conferences and discussions	38.0	47.1	43.8	72.5	54.5	68.2	48.1
General teachers' meetings devoted to measurements	38.0	41.4	48.4	47.5	45.5	63.6	45.1
Grade level or subject meetings	12.0	18.6	27.7	52.5	51.5	40.9	27.3
Extension courses by University	10.0	15.7	18.5	20.0	27.3	18.2	16.7
Workshop in measurements	7.0	4.2	3.8	12.5	9.1	18.2	6.8
Others[a]	2.0	1.4	3.8	10.0	15.5	18.2	5.3
No response	41.0	38.6	32.3	12.5	12.1	18.2	31.1

a/In order of decreasing frequency: individual conferences with teachers to interpret test data, school bulletins covering various aspects of program, help from outside specialists, and P.T.A. meetings for parents.

Many and varied means of conducting the in-service training program are included in Table 42, with the following reported most frequently: group conferences and discussions, 48.1 per cent; general teachers' meetings, 45.1 per cent; and grade level or subject meetings, 27.3 per cent. Parent education is a major aspect of this program. It is significant to note that this is mentioned by only a very limited number of schools. Attention is also called to the high percentage of "no response," implying no formalized approach.

Time is required to effectively carry on the in-service program. Table 43 indicates provisions made by the schools for this work.

Table 43. Extent to Which Teachers Are Given Released Time for In-Service Work in Measurements

Extent	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Occasionally	33.0	25.7	30.8	37.5	48.5	45.5	33.4
Never	23.0	37.1	34.6	40.0	33.3	31.8	32.6
Regularly	5.0	2.8	1.5	7.5	6.1	-	3.5
No response	39.0	34.3	33.0	12.5	12.1	22.7	30.4

If the in-service program is to attain professional status and constructive work is to be accomplished, time must be provided for its operation. This is not the case in the New England schools, where only 3.5 per cent of the schools regularly provide time for this work. It may further be assumed that schools not responding to this item do not provide time for the program. This implies that 63 per cent of the schools provide no released time for this work.

As an aid to attaining accuracy and uniformity in test administration, the literature recommends that mimeographed directions, etc. be prepared locally for teacher use. These may supplement the publisher's material. Practices pertaining to this technique are noted in Table 44.

Table 44. Extent to Which Mimeographed Teacher Directions, Supplementing Those of the Test Publisher, Are Prepared for Classroom Teachers Administering Standardized Tests

Extent	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Not prepared for teachers' use	44.0	34.3	30.8	22.5	21.2	4.5	31.6
Prepared regularly for all testing	18.0	22.8	29.2	45.0	57.6	68.2	31.4
Prepared only for certain types of tests	22.0	32.8	34.6	32.5	18.2	27.3	29.1
No response	16.0	10.0	6.1	5.0	6.1	-	8.9

Table 44 indicates that it is not common practice to prepare local supplementary material to assist teachers in administering standardized tests, with only 31.4 per cent of the schools reporting this as regular practice. It is further noted that the larger school systems are more active in this area than the small schools. Previous data indicated that standardized tests are largely administered by teachers, particularly in the small schools. It is evident, therefore, that there is a real need for this material if the program is to attain a higher degree of accuracy and uniformity.

Program status.-- A key to the import and impact of testing and evaluation in relation to the total education program is its status during the past five years. Tables 45 to 47 present data pertaining to this aspect.

Table 45. Status of the Testing and Evaluation Program During the Past Five Years

Status	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Program has steadily expanded	49.0	74.3	74.6	85.0	90.9	63.6	69.9
Program has remained essentially the same	45.0	22.9	23.1	15.0	9.1	36.4	27.3
Program has diminished in scope	1.0	-	2.3	-	-	-	1.0

Table 46. Status of Budgetary Expenditure for Testing and Evaluation During the Past Five Years in Proportion to the Total Instructional Budget

Status	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Slight increase	52.0	54.3	46.9	42.5	42.4	36.4	48.1
Significant increase	14.0	27.1	31.5	45.0	51.5	45.4	30.1
No appreciable increase	28.0	11.4	16.9	12.5	3.0	13.6	20.0

It is gratifying to note that approximately 70 per cent of the schools report that the program has expanded steadily. At the same time, it is disturbing to note that the small schools evidence less program expansion than the larger schools. The key to the program expansion is budgetary expenditure (Table 46). Findings indicate that 48.1 per cent of the schools report only slight increases, and only

30 per cent report significant increases. The impact of the National Defense Education Act of 1958 should affect appreciably this aspect of the program.

It is significant to note the reasons underlying status changes (Table 47), since they may affect other aspects of the program.

Table 47. Chief Reasons for Change Made in Testing and Evaluation Program During Past Five Years

Reason	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Recognition of new needs	41.0	61.4	60.8	75.0	69.7	68.2	58.5
Dissatisfaction with previous program	22.0	21.4	14.6	20.0	15.1	9.1	18.0
Others[a/]	9.0	5.7	12.3	5.0	9.1	13.6	9.4

a/In order of decreasing frequency: change in administration, increase in enrollment caused by consolidation, organization of new guidance services, pressure from outside, addition of better trained personnel, increase in funds made available, organization of regional school district.

A variety of factors causing a change in status is noted in Table 47. Approximately 58 per cent of the systems report recognition of new needs. It is also significant to note that the regional school movement affecting consolidation has aided the program, as well as new administrators bringing in new viewpoints to the schools.

Major problems.-- Major problems encountered at the local level seriously affect the testing and evaluation program.

Table 48. Major Problems or Weaknesses in Present Program

Problem or Weakness	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Limited use being made of test results	30.0	32.8	42.3	30.0	39.4	31.8	35.4
Insufficient time and resources for full utilization	23.0	30.0	32.3	40.0	39.4	50.0	31.9
Teachers not properly trained in use of test results	30.0	28.6	25.4	17.5	27.3	22.7	26.2
No proper follow-up of testing program	21.0	27.1	24.6	25.0	15.1	22.7	23.3
Lack of trained personnel in tests and measurements	33.0	17.1	16.9	20.0	12.1	4.5	20.5
Program too limited in scope	18.0	10.0	7.7	12.5	6.1	9.1	11.1
Others[a]	-	1.4	8.5	7.5	9.1	18.2	5.6

a/In order of decreasing frequency: not enough time given to statistical analysis, teachers need more training, lack of clerical help for the program, lack of coordinated teating program at different levels, need for more effective process of scoring, need for local norms, need for more individual testing.

A variety of problems and weaknesses is shown in Table 48. It is noted that most problems center around two major areas: (1) test utilization, and (2) staff limitations. The former has implications for expansion of staff training at both the in-service and preservice levels. The latter has budgetary implications and requires additional special personnel to provide more adequate services and coverage in critical areas of the program.

4. Pupil Appraisal Program

Measuring intelligence.-- The appraisal of intelligence or academic aptitude forms the basis of most testing programs. The two main patterns of academic aptitude tests are group tests and individual tests. Practices pertaining to group tests are depicted in Tables 49 to 51.

Table 49. Extent to Which Group Intelligence Tests Are Used in the School System

Extent	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Used on systematic basis at regular intervals throughout school system	83.0	90.0	96.1	97.5	100.0	90.9	91.9
Used on one-time or sporadic basis	8.0	8.6	3.1	2.5	-	-	4.8
Used only with individuals or class groups as need arises	7.0	4.2	0.8	-	-	4.5	3.0
Not used	1.0	-	-	-	-	-	0.2

Table 50. Grade Levels at Which Group Intelligence Tests Are Generally Administered

Grade	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	12.0	7.1	9.2	12.5	27.3	22.7	12.1
I	53.0	51.4	45.4	60.0	63.6	18.2	49.9
II	40.0	28.6	20.8	35.0	12.1	13.6	27.3
III	56.0	47.1	43.8	52.5	51.5	54.5	52.1
IV	54.0	38.6	42.3	50.0	45.4	18.2	44.3
V	50.0	38.6	33.1	47.5	33.3	27.3	39.5
VI	49.0	41.4	40.8	50.0	51.5	45.4	45.1
VII	62.0	47.1	44.6	65.0	42.4	18.2	49.6
VIII	52.0	37.1	43.8	45.0	24.2	45.4	43.3
IX	23.0	47.1	36.1	52.5	48.5	13.6	36.2
X	19.0	24.3	22.3	30.0	24.2	18.2	22.5
XI	27.0	37.1	41.5	50.0	60.6	22.7	38.5
XII	14.0	11.4	10.8	20.0	12.1	4.5	14.9

Table 51. Number of Times Group Intelligence Tests Are Generally Administered to a Pupil During His School Career

Number of Times Administered	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
One	3.0	-	2.3	2.5	-	4.5	2.0
Two	7.0	8.6	2.3	-	3.0	4.5	4.5
Three	23.0	27.1	20.0	20.0	15.1	27.3	22.0
Four	25.0	25.7	32.3	20.0	27.3	31.8	27.6
Five	7.0	17.1	19.2	27.5	39.4	22.7	18.5
Six	7.0	10.0	12.3	15.0	9.1	9.0	10.4
Seven	1.0	1.4	1.5	2.5	3.0	-	1.5
Eight	8.0	1.4	0.8	2.5	3.0	-	3.0
Nine or more	14.0	7.1	6.9	10.0	-	-	8.1

A study of Table 49 reveals that approximately 92 per cent of the schools report the use of group intelligence tests on a systematic basis at regular intervals throughout the system. Data presented in Table 50 note that higher percentages of responses are in grades 1, 3, 6, or 7. This implies that the concentration of testing occurs in grades 1, 3, and 6 or 7, with appreciable amounts occurring in grades 8, 9, and 11. It is also significant to note that 27.6 per cent of the schools report that they administer three group tests to pupils during their school career, 22 per cent administer four group tests, and 18.5 per cent administer five such measures during the pupil's school career.

The literature recommends that group intelligence tests be given at reasonable frequencies to avoid underrating or overrating by one testing. Although testing every year is not necessary, administration at transitional points in the pupil's school history is advocated. This implies at least four or five times in grades 1, 3, 6, 8, and 10 or 11. Current surveys practices reviewed in Chapter II reveal that the majority of schools follow this general pattern.

The general pattern in the New England schools presented in Tables 51 to 53 reveals that the large majority of schools systematically administer group intelligence tests. Approximately 68 per cent of schools test pupils three to five times during their school career, with the greatest concentration being at grades 1, 3, and 6 or 7. Attention is called to the fact that considerable deviation occurs in this general pattern, some schools going beyond this pattern and others not meeting the general practice. The literature recommends more con-

centration in testing at the transitional points in the pupil's school career, particularly at the upper grade level where Table 51 indicates a slight decrease in frequency in comparison with the other levels.

The literature recommends retests, with individual tests if possible, of pupils who test below 80 or above 130, or of those whose scores are out of line with the judgment of the teacher. Table 52 indicates the extent to which individual testing is reported by the New England schools.

Table 52. Extent to Which Individual Intelligence Tests Are Used in the School System

Extent	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Used only with individuals or class groups as need arises	46.0	58.6	74.6	67.5	81.8	95.4	65.6
Used on systematic basis at regular intervals throughout schl. system	12.0	12.8	11.5	32.5	12.1	4.5	13.7
Used on one-time or sporadic basis	10.0	11.4	6.1	2.5	6.1	-	7.3
Not used	7.0	4.2	2.3	2.5	-	-	3.5

Responses included in Table 52 imply that approximately 73 per cent of the schools make use of individual tests to retest and check questionable cases. Although the literature points to the advantages of individual testing, the expense precludes large-scale use of these

instruments. In spite of this, it is interesting to note that 13.7 per cent of the schools report use on a systematic basis at regular intervals. Findings indicate that the majority of schools use individual tests to supplement the regular group testing program.

Measuring special aptitude.-- Surveys of current practices[1/] reveal that, in the majority of schools, aptitude tests are available and used by schools for special needs or purposes at the secondary level. Tables 53 and 54 present responses provided by the New England public schools pertaining to the use of these measures.

Table 53. Extent to Which Special Aptitude Tests Are Used in the School System

Extent	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Used on systematic basis at regular intervals throughout schl. system	16.0	47.1	44.6	50.0	60.6	45.4	39.7
Used only with individuals or class groups as need arises	23.0	25.7	23.1	35.0	33.3	36.4	26.3
Used on one-time or sporadic basis	12.0	10.0	13.8	20.0	9.1	18.2	13.2
Not used	16.0	5.7	3.8	2.5	-	4.5	6.8

1/Chapter II, supra.

Table 54. Grade Levels at Which Special Aptitude Tests Are Generally Administered

Grade	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	1.0	-	1.5	7.5	3.0	-	2.2
I	4.0	-	3.8	2.5	-	4.5	2.8
II	4.0	-	2.3	2.5	-	-	2.2
III	3.0	-	3.8	2.5	3.0	4.5	2.8
IV	6.0	-	2.3	2.5	-	-	2.6
V	2.0	1.4	3.8	2.5	-	-	2.4
VI	3.0	1.4	3.1	5.0	3.0	4.5	3.0
VII	9.0	7.1	5.4	12.5	-	-	6.6
VIII	23.0	30.0	30.0	25.0	12.1	45.4	27.1
IX	16.0	37.1	26.9	45.0	6.1	31.8	26.3
X	13.0	27.1	29.2	42.5	12.1	18.2	24.0
XI	21.0	34.3	42.3	55.0	18.2	22.7	33.7
XII	17.0	21.4	25.4	25.0	9.1	13.6	20.5

Approximately 40 per cent of the schools use special aptitude tests on a systematic basis at regular intervals throughout the system (Table 53). Forty per cent report special purpose or sporadic use. Assuming that those not responding do not include these measures in the program, 20 per cent of the schools do no special aptitude testing. A study of Table 53 reveals that the large system does more aptitude testing than the small school. This may be due to many small systems not maintaining secondary schools. Grade levels most frequently tested are grade 11, 8, 9, 10, and 12, in the order named (Table 53). A further breakdown of the data provided in the inquiry form discloses that approximately one half of the systems that do administer these measures on a systematic basis throughout the schools, give these tests once

during the pupil's school career, while the remaining one half administer these measures twice.

The literature recommends use of these measures with limitations for guidance purposes. Their use at grades 8 and 9, and at points when pupils have options of selecting courses or encounter difficulties which appear to be due to lack of specific ability or aptitude, is recommended by several writers. Current practices surveys indicate that schools make limited use of special aptitude tests as a supplement to the regular intelligence and achievement testing program. Differential measures are finding greater use for this purpose.

Findings reported in the present study reveal that the majority (80 per cent) of the New England public schools are using special aptitude tests in conformity with practices of schools in other sections of the country. Although the heavy concentration of administration is at grade 11, schools are generally administering these measures at points where pupils have options of selecting courses. The present study did not investigate the use of differential measures. This may well receive further consideration in future investigations.

Measuring interests.-- The extent to which interest measures are used in the schools is indicated in Tables 55 and 56.

Table 55. Extent to Which Interest Inventories Are Used in the School System

Extent of Use	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout school system	17.0	45.7	53.8	65.0	54.5	31.8	43.0
With individuals or class groups as need arises	18.0	17.1	20.8	15.0	27.3	50.0	21.0
One-time or sporadic basis	11.0	12.8	10.8	12.5	9.1	4.5	10.9
Not used	20.0	8.6	3.1	2.5	-	9.1	8.3

Table 56. Grade Level at Which Interest Inventories Are Generally Administered

Grade	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	-	-	-	-	-	-	-
I	-	-	-	-	-	-	-
II	-	-	-	-	-	-	-
III	-	-	-	-	-	-	-
IV	-	-	-	-	-	-	-
V	1.0	1.4	1.5	-	-	-	1.0
VI	2.0	1.4	0.8	2.5	-	-	1.3
VII	4.0	1.4	2.3	5.0	-	-	2.5
VIII	20.0	14.3	16.4	20.0	9.1	13.6	16.4
IX	15.0	25.7	34.6	27.5	48.5	31.8	28.3
X	16.0	14.3	21.5	35.0	30.3	22.7	21.0
XI	21.0	25.7	47.7	40.0	51.5	45.4	36.4
XII	16.0	12.8	19.2	27.5	21.2	22.7	18.5

Forty-three per cent of the schools report use of interest inventories on a systematic basis at regular intervals, and 32 per cent report use to some extent (Table 55). It is also noted that the small schools make rather limited use of these measures. The grade level at which interest measures are reported most frequently administered, noted in order of decreasing frequency, are: grade 11, grade 9, and grade 10.

The literature recommends the use of interest inventories to assess interest patterns with limitations. Their use below the senior high school level is questioned by many writers. Current practices surveys reveal that the majority of schools use interest inventories to some extent at the secondary level.

From the findings reported in Tables 55 and 56, it is evident that the majority of the New England public schools are making some use of interest inventories. These measures find most frequent use at the senior high school level, with grade 11 being the heaviest concentration point. Significant use is also reported in grades 9 and 8. Only minor use occurs below grade 8. The practices reported in this study appear to be typical of those findings reported in other current practices surveys. Use of these measures by a limited number of schools at the junior high school level dictates care and caution in interpretation, particularly where specially trained personnel are not available.

Measuring personality.-- Personality tests are largely in the developmental stage and should be used with extreme caution and care, and

then only by those persons who have adequate training for interpreting the results. The literature urges very conservative use of these measures. This is reflected in practices reported in previous surveys[1/] where use was reported only infrequently and then generally for special purposes. Practices in the New England schools are reflected in Tables 57 and 58.

Table 57. Extent to Which Personality Tests Are Used in the School System

Extent of Use	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individuals or class groups as need arises	17.0	24.3	33.1	35.0	51.5	54.5	30.4
Not used	24.0	18.6	19.2	7.5	12.1	18.2	18.5
Systematic basis at regular intervals throughout school system	9.0	25.7	20.8	22.5	12.1	13.6	17.7
One-time or sporadic basis	7.0	8.6	6.1	17.5	6.1	-	7.6

1/Chapter II, supra.

Table 58. Grade Level at Which Personality Tests Are Generally Administered

Grade	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	-	-	-	2.5	3.0	9.1	1.0
I	2.0	-	0.8	7.5	-	9.1	2.0
II	2.0	-	0.8	7.5	-	9.1	2.0
III	2.0	-	0.8	12.5	-	9.1	2.5
IV	3.0	-	0.8	12.5	-	9.1	2.8
V	2.0	1.4	0.8	10.0	-	9.1	2.5
VI	3.0	-	0.8	12.5	-	9.1	2.8
VII	2.0	2.8	4.6	10.0	6.1	22.7	5.3
VIII	9.0	7.1	5.4	10.0	15.1	18.2	8.6
IX	9.0	18.6	15.4	27.5	12.1	13.6	15.2
X	6.0	12.8	12.3	30.0	18.2	27.3	13.9
XI	8.0	17.1	18.5	20.0	21.2	27.3	16.4
XII	10.0	10.0	15.4	15.0	15.1	22.7	13.4

Table 57 reveals that the primary use of personality tests is for individuals or class groups as the need arises, reported by 30.4 per cent of the schools. Significant use on a systematic basis at regular intervals is noted in categories II to IV. This could well bear further investigation. Taking into account respondents who did not answer this item, it may be assumed that approximately 44 per cent of the school systems do not use these measures. Data provided in Table 58 reveal that the greatest test concentration is in grades 9 to 12. Since the major use is for individuals or class groups as the need arises, it may be assumed that test usage is spread over all grades. The consistently low percentages noted in kindergarten through grade 6 may imply individual testing for special cases, while the higher per-

centages in the upper grades, particularly grades 9 to 12, may imply the addition of group testing at these levels. The discrepancy of percentage responses in schools of categories IV and VI at the elementary level bears further investigation before drawing any conclusions regarding their responses.

Measuring achievement in kindergarten through grade 8.-- The literature indicates that, with the possible exception of mental maturity, achievement is the most important area of appraisal in the testing and evaluation program. The extent to which achievement is measured by standardized tests in the basic skills and informational areas is indicated in Tables 59 to 66.

Table 59. Extent to Which Reading Is Measured by Standardized Tests in Kindergarten Through Grade 8

Extent	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout the system	84.0	88.5	94.6	90.0	100.0	100.0	93.2
Individuals or class groups as need arises	7.0	7.1	1.5	5.0	6.1	18.2	5.3
One-time or sporadic basis	8.0	4.2	3.8	7.5	-	4.5	4.5
Not measured	-	-	-	-	-	-	-

Table 60. Extent to Which Arithmetic Is Measured by Standardized Tests in Kindergarten Through Grade 8

Extent Arithmetic Is Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	84.0	88.5	94.6	90.0	100.0	100.0	91.9
Individuals or class groups as need arises	7.0	7.1	1.5	5.0	6.1	18.2	5.5
One-time or sporadic basis	8.0	4.2	3.8	7.5	-	4.5	5.1
Not used	-	-	-	-	-	-	-

Table 61. Extent to Which Language Usage Is Measured by Standardized Tests in Kindergarten Through Grade 8

Extent Language Usage Is Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	84.0	84.3	93.0	90.0	90.9	100.0	89.1
One-time or sporadic basis	9.0	4.2	5.4	10.0	3.0	4.5	6.2
Individuals or class groups as need arises	5.0	8.6	0.8	2.5	6.1	18.2	4.8
Not measured	-	-	-	2.5	-	-	0.2

Table 62. Extent to Which Spelling Is Measured by Standardized Tests in Kindergarten Through Grade 8

Extent Spelling Is Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	80.0	84.3	91.5	85.0	87.9	100.0	86.3
One-time or sporadic basis	8.0	4.2	5.4	10.0	-	4.5	5.8
Individuals or class groups as need arises	6.0	8.6	0.8	7.5	6.1	18.2	5.6
Not measured	1.0	1.4	-	-	-	-	0.5

Data presented in Tables 59 to 62 reveal that standardized measures in reading, arithmetic, language usage, and spelling are used almost universally on a systematic basis throughout the New England area. A similarity of the general pattern of appraisal by standardized measures in these four areas is noted, particularly in reading and arithmetic. More extensive use of standardized measures on a systematic basis is also evidenced in the large school systems in comparison with the small systems. The use of standardized measures for individual or class groups as the need arises and on a sporadic basis may imply retesting to check special cases or additional diagnostic testing. This is particularly evident in arithmetic and reading in the large school systems.

Table 63. Extent to Which Social Studies Are Measured by Standardized Tests in Kindergarten Through Grade 8

Extent Social Studies Is Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	70.0	77.1	68.4	72.5	48.5	45.4	67.8
One-time or sporadic basis	11.0	4.2	8.5	12.5	9.1	13.6	9.1
Individuals or class groups as need arises	5.0	11.4	3.1	7.5	3.0	27.3	6.8
Not measured	4.0	1.4	5.4	5.0	9.1	9.1	5.1

Table 64. Extent to Which Science Is Measured by Standardized Tests in Kindergarten Through Grade 8

Extent Science Is Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	67.0	75.6	67.7	70.0	42.4	45.4	65.8
One-time or sporadic basis	10.0	4.2	8.5	12.5	12.1	18.2	9.4
Individuals or class groups as need arises	6.0	10.0	3.1	7.5	3.0	27.3	6.8
Not measured	5.0	2.8	4.6	7.5	12.0	13.6	5.8

Table 65. Extent to Which Work-Study Skills Are Measured by Standardized Tests in Kindergarten Through Grade 8

Extent Study-Skills Are Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	69.0	68.6	66.1	70.0	45.4	40.9	64.5
One-time or sporadic basis	10.0	4.2	8.5	5.0	12.1	13.6	8.3
Individuals or class groups as need arises	7.0	10.0	3.1	10.0	9.1	27.3	7.8
Not measured	2.0	2.8	6.1	7.5	9.1	18.2	5.6

Social studies, science, and work-study skills are measured regularly on a system-wide basis in the majority of schools, but not as extensively as reading, arithmetic, language usage, and spelling. This may be due to the use of partial achievement batteries rather than complete batteries by some of the schools. The wide variation in subject matter in the social studies and science at the elementary grades in various school systems may also tend to decrease general use of these measures. Tables 63 to 65 reveal that the small school systems make more extensive use of standardized measures in social studies, science, and work-study skills on a system-wide basis than the larger systems. This may be due to more system-wide use of locally constructed measures in the larger schools. Attention is also called to the significant use of standardized measures for class groups and individuals in the large school system. These facts bear further investigations.

Although study skills are measured rather extensively, it is disturbing to note that they do not receive the degree of emphasis of other areas.

Table 66 provides a summarization of the extent to which the specific areas presented in Tables 59 to 65 are measured in the total sample of New England respondents.

Table 66. Extent to Which Basic Areas Are Measured by Standardized Tests in Kindergarten Through Grade 8 in the Total New England Sample

Skill or Informational Area	Percentage of Responses			
	Systematic Basis at Regular Intervals Throughout System	One-Time or Sporadic Basis	To Individuals or Class Groups as Need Arises	Not Measured
(1)	(2)	(3)	(4)	(5)
Reading	93.2	4.6	5.3	-
Arithmetic	91.1	5.1	5.6	-
Language Usage	89.1	6.3	4.8	-
Spelling	86.8	5.8	5.6	-
Social Studies	67.8	9.1	6.8	5.1
Science	65.8	9.4	6.8	5.8
Work-Study Skills	64.6	8.4	7.8	5.6

A review of the preceding tables reveals that the basic skills and informational areas are generally being appraised by standardized measures on a systematic basis at regular intervals throughout the system in most of the New England public schools. As noted previously and in Table 66, the three "R's" are receiving the greatest emphasis. Although schools report systematic testing in these areas at approxi-

mately the 90 per cent level; the literature states that, in order to meet present-day demands, this should reach the 100 per cent level. Study skills should also attain this degree of appraisal. When the science and social studies curriculum attains a higher degree of stability and standardized measures are available to adequately assess learnings in these areas, testing should be at the same level of usage as in the other basic skills.

The literature states that in an optimum program standardized achievement tests should be given annually in every grade. If this is not possible, most writers agree that the minimal program should provide for measures to be administered at least each time the curriculum changes noticeably. This implies at the end of grades 3, 6, and possibly 8, or approximately every three years such measures will be applied. In addition to this, diagnostic and other measures are advocated to be used as the need arises. Other investigations reviewed in this study[1/] reveal that all schools administer standardized achievement tests at regular intervals, with reading measured usually at grades 1, 2, or 2, 3, and 6, and general achievement at grades 4, 5, or 8. Wide use of the achievement battery was also reported.

The grade levels at which different skills and informational areas are generally measured by standardized achievement tests in kindergarten through grade 8 are noted in Table 67. A further breakdown of these data according to enrollment categories is provided in the tables included in Appendix G.

1/Chapter II, *supra*.

Table 67. Grade Levels at Which Basic Skills or Informational Areas Are Generally Measured by Standardized Tests

Skill or Informational Area	Percentage of Responses According to Grade Level								
	K	Gr.1	Gr.2	Gr.3	Gr.4	Gr.5	Gr.6	Gr.7	Gr.8
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Reading	10.6	55.7	65.6	81.5	79.2	78.7	85.6	76.2	80.2
Arithmetic	5.3	36.2	52.4	73.4	77.5	76.7	84.3	72.4	81.5
Lang. Usage	3.3	28.3	42.5	72.1	74.7	73.2	81.0	71.6	81.3
Spelling	6.3	32.4	47.8	70.3	73.2	73.4	80.0	69.3	74.7
Soc. Studies	2.8	19.7	26.8	41.8	51.4	53.7	63.3	58.7	67.1
Science	2.3	18.0	24.8	37.2	47.1	52.6	59.7	55.4	62.0
Work-Study Skills		16.4	21.8	33.0	47.1	54.7	60.0	56.7	60.7

A review of the data presented in Table 67 reveals the following:

1. Reading achievement is measured most frequently in grades 6, 3, 8, 4, 5, and 7 in order mentioned, with 80 to 85 per cent of the school systems testing in grades 6, 3, and 8; and 75 to 80 per cent in grades 4, 5, and 7.
2. Arithmetic achievement is measured most frequently in grades 6, 8, 4, 5, 3, and 7 in the order mentioned, with 80 to 85 per cent of the school systems testing in grades 6 and 8; 75 to 80 per cent in grades 4 and 5; and 70 to 75 per cent in grades 3 and 7.
3. Language usage is measured most frequently in grades 8, 6, 4, 5, 3, and 7 in the order named, with 80 to 85 per cent of the school systems testing in grades 8 and 6; and 70 to 75 per cent in grades 4, 5, 3, and 7.
4. Spelling is measured most frequently in grades 6, 8, 5, 4, and

7 in the order named, with 80 to 85 per cent of the school systems testing in grade 6; and 70 to 75 per cent in grades 8, 5, 4, and 3.

5. Social studies achievement is measured most frequently in grades 8, 6, 7, 5, and 4 in the order named, with 65 to 70 per cent of the school systems testing in grade 8; 60 to 65 per cent in grade 6; 55 to 60 per cent in grades 5 and 7; and 50 to 55 per cent in grade 4.
6. Science achievement is measured most frequently in grades 8, 6, 7, 5, and 4 in the order named, with 60 to 65 per cent of the school systems testing in grade 8, 55 to 60 per cent in grade 6; 50 to 55 per cent in grades 7 and 5; and 45 to 50 per cent in grade 4.
7. Work-study skills achievement is measured most frequently in grades 8, 6, 7, 5, and 4 in the order named, with 60 to 65 per cent of the school systems testing in grades 8 and 6; 55 to 60 per cent in grade 7; 50 to 55 per cent in grade 5; and 45 to 50 per cent in grade 4.

A further breakdown of the data provided by the respondents reveals the number of times the basic skills and informational areas are generally measured during the pupil's school career. A general summary is provided in Table 68. A further breakdown according to enrollment categories is included in the tables in Appendix G.

Table 68. Number of Times Basic Skills and Informational Areas Are Measured During the Pupil's School Career

Skill or Informational Area	Percentage of Responses According to Number of Times Measured				
	Once	Twice	Three Times	More Than Three Times	Annually
(1)	(2)	(3)	(4)	(5)	(6)
Reading	1.8	5.1	7.1	82.9	44.5
Arithmetic	2.3	7.6	10.6	75.2	33.7
Lang. Usage	5.3	8.3	11.1	75.8	28.5
Spelling	2.0	6.3	11.1	71.9	28.1
Soc. Studies	4.3	10.6	7.3	56.5	18.5
Science	5.1	11.3	8.1	49.1	15.7
Work-Study Skills	4.5	8.9	6.6	52.4	14.4

A review of the preceding tables and data implies that: reading, arithmetic, and language usage are measured by more than 85 per cent of the school systems each time the curriculum changes noticeably (grades 3, 6, 8), with spelling being measured by approximately 80 per cent of the schools at these points. Approximately 60 per cent of the schools measure social studies, science, and language arts at these points. A significant number of schools measure these areas annually. It is evident that the New England public schools as a group are meeting minimal standards in the measurement of basic skills and informational areas. It is also worthy of note that a significant number of schools provide excellent coverage in these areas. Reading and arithmetic receive the greatest coverage, with the others following in the order listed in the preceding table. The data also appear to indicate that the smaller schools make somewhat greater use of standardized

measures, both in degree and coverage and frequency, in the basic areas than the large school district.

Measuring achievement in grades 9 to 12.-- The literature recommends that standardized testing be continued through the secondary level, including broad field tests and appropriate separate subject tests. The importance of testing certain basic skills, particularly reading and work-study, is also cited. Surveys of current practices[1/] reported widespread use of achievement testing at the secondary level. Findings of the present study pertaining to the extent of standardized testing in grades 9 to 12, inclusive, are presented in Tables 69 to 82.

The extent to which the basic academic areas are measured by standardized tests is depicted in Tables 69 to 74.

Table 69. Extent to Which Mathematics Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	26.0	51.4	61.5	67.5	57.6	54.5	50.6
Individuals or class groups as need arises	6.0	11.4	12.3	10.0	18.2	31.8	11.9
One-time or sporadic basis	15.0	12.8	9.2	10.0	12.1	9.1	11.6
Not measured	4.0	5.7	3.8	5.0	-	-	3.8
No response	25.0	17.1	11.5	10.0	9.1	13.6	15.7

1/Chapter II, supra.

Table 70. Extent to Which Language Usage Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	30.0	50.0	56.1	55.0	57.6	54.5	48.3
Individuals or class groups as need arises	4.0	11.4	13.1	12.5	24.2	18.2	11.6
One-time or sporadic basis	10.0	10.0	8.5	15.0	6.1	9.1	9.6
Not measured	5.0	5.7	3.8	5.0	-	4.5	4.3
No response	27.0	21.4	16.9	15.0	12.1	22.7	20.0

Table 71. Extent to Which Science Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	24.0	50.0	50.8	60.0	45.4	40.9	43.8
Individuals or class groups as need arises	6.0	10.0	18.5	10.0	30.3	27.3	14.4
One-time or sporadic basis	13.0	11.4	9.2	12.5	6.1	4.5	10.4
Not measured	4.0	5.7	5.4	5.0	-	4.5	4.5
No response	29.0	21.4	14.6	15.0	15.1	27.3	20.2

Table 72. Extent to Which Social Studies Are Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	23.0	48.5	46.1	57.5	39.4	31.8	40.5
Individuals or class groups as need arises	7.0	10.0	15.4	10.0	27.3	31.8	13.7
One-time or sporadic basis	13.0	11.4	7.7	10.0	6.1	9.1	9.9
Not measured	4.0	5.7	6.1	10.0	3.0	4.5	5.6
No response	29.0	22.8	23.1	15.0	21.2	27.3	23.8

Table 73. Extent to Which Literature Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	21.0	37.1	38.5	47.5	30.3	18.2	32.9
Individuals or class groups as need arises	5.0	7.1	8.5	10.0	24.2	22.7	9.6
One-time or sporadic basis	10.0	11.4	9.2	10.0	6.1	-	9.1
Not measured	9.0	7.1	10.0	10.0	3.0	9.1	8.6
No response	31.0	35.7	32.3	22.5	33.3	54.5	32.9

Table 74. Extent to Which Modern Language Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	11.0	22.8	26.9	42.5	21.2	18.2	22.8
Individuals or class groups as need arises	7.0	11.4	20.0	15.0	21.2	36.4	15.7
Not measured	12.0	12.8	14.6	10.0	3.0	9.1	11.9
One-time or sporadic basis	9.0	7.1	6.9	7.5	12.1	4.5	7.8
No response	37.0	42.3	29.2	27.5	39.4	36.4	34.9

The preceding tables reveal that in the academic areas mathematics receives the greatest emphasis in standardized testing, followed in order by language usage, science, social studies, literature, and foreign language. Schools in category IV (3000-5000 enrollment) carry on the greatest amount of systematic testing. Low percentages noted in category I may be due to the fact that some of these systems do not operate high schools. Although the large city system (category VI) does less systematic testing than schools in some of the other categories, it should be noted that they do a considerable amount of testing for individuals or class groups as the need arises.

The importance of reading and work-study skills as a controlling factor in school success is cited by writers. The use of tests for determining proficiency and identification of pupils needing assistance

is advocated in the literature. The extent to which these areas are measured is noted in Tables 75 and 76.

Table 75. Extent to Which Reading Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	31.0	51.4	57.7	65.0	66.7	54.5	51.1
Individuals or class groups as need arises	4.0	11.4	13.8	15.0	12.1	22.7	11.4
One-time or sporadic basis	12.0	10.0	10.0	10.0	6.1	18.2	10.6
Not measured	6.0	5.7	2.3	5.0	-	-	3.8
No response	23.0	20.0	16.1	7.5	15.1	18.2	17.7

Table 76. Extent to Which Work-Study Skills Are Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	VI	V	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	20.0	35.7	34.6	40.0	27.3	22.7	30.4
Not measured	7.0	8.6	14.6	12.5	15.1	13.6	11.4
Individuals or class groups as need arises	8.0	8.6	7.7	15.0	12.1	13.6	10.1
One-time or sporadic basis	11.0	5.7	8.5	7.5	3.0	9.1	8.1
No response	30.0	40.0	32.3	27.5	39.4	31.8	33.2

Reading as a basic tool subject is measured systematically by approximately 50 per cent of the schools. Although work-study skills are recognized as an essential functional aspect of learning, less than one third of the schools systematically test this area. This is particularly noticeable in the small schools.

Standardized tests in such areas as business education, music, art, and related areas are not as well developed as those in the basic academic areas. The literature indicated that achievement in many of these areas consists primarily of products which lend themselves chiefly to rating. The extent to which these areas are measured by standardized measures is noted in Tables 77 to 81.

Table 77. Extent to Which Business Education Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentages of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	10.0	18.6	18.5	32.5	24.2	4.5	17.5
Individuals or class groups as need arises	7.0	11.4	20.8	22.5	21.2	31.8	16.4
One-time or sporadic basis	8.0	5.7	6.9	7.5	6.1	18.2	7.6
Not measured	13.0	14.3	17.7	10.0	9.1	9.1	13.9
No response	38.0	48.5	33.8	30.0	36.4	45.4	38.0

Table 78. Extent to Which Home Economics Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	2.0	4.2	7.7	12.5	-	-	5.1
Individuals or class groups as need arises	7.0	7.1	8.5	2.5	12.1	18.2	8.1
One-time or sporadic basis	5.0	8.6	3.1	7.5	-	-	4.5
Not measured	2.0	11.4	21.5	30.0	21.2	22.7	15.7
No response	74.0	68.7	59.2	47.5	66.7	59.1	66.6

Table 79. Extent to Which Industrial Arts Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	2.0	2.8	5.4	10.0	3.0	-	4.0
Individuals or class groups as need arises	6.0	7.1	7.7	7.5	12.1	18.6	7.8
One-time or sporadic basis	4.0	7.1	1.5	5.0	-	-	3.3
Not measured	3.0	12.8	24.6	30.0	21.2	27.3	17.5
No response	65.0	30.2	60.8	47.5	63.7	54.1	67.4

Table 80. Extent to Which Music Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	4.0	4.2	3.1	7.5	3.0	-	3.8
Individuals or class groups as need arises	4.0	7.1	6.1	5.0	15.1	18.2	7.1
One-time or sporadic basis	2.0	2.8	4.6	12.5	3.0	-	4.3
Not measured	4.0	10.0	24.6	27.5	18.2	27.3	16.5
No response	86.0	75.9	61.6	47.5	60.7	54.5	68.3

Table 81. Extent to Which Art Is Measured by Standardized Tests in Grades 9 Through 12

Extent Measured	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	1.0	4.2	3.8	12.5	-	-	3.5
Individuals or class groups as need arises	4.0	4.2	4.6	-	3.0	22.7	5.1
One-time or sporadic basis	3.0	4.2	2.3	7.5	-	-	3.0
Not measured	4.0	11.4	25.4	32.5	27.3	22.7	18.2
No response	88.0	76.0	63.9	47.5	69.7	54.6	70.2

Other than business education, which is measured by standardized tests on a systematic basis by 17.5 per cent of the schools, an insignificant amount of standardized testing is carried on by the New England public schools in these curriculum areas. It is noted that the large city system (category VI) reports systematic use of standardized tests in only business education. Schools in category IV make the greatest use of standardized measures in these areas, as well as in other areas. This may be due to the fact that a large number of schools in this category are located in the more wealthy urban community. It is also noted that measures in these curriculum areas find some use for individuals or class groups as the need arises, particularly in categories V and VI. Standardized testing in these curriculum areas finds greatest use for this purpose as opposed to systematic use at regular intervals.

Table 82 provides a general overview of the extent of standardized testing in grades 9 through 12, as presented in detail by the thirteen preceding tables.

Table 82. Extent to Which Standardized Tests Are Used in the Basic Skills or Informational Areas in Grades 9 Through 12

Skill or Informational Area	Percentage of Responses According to Extent of Use			
	Systematic Basis at Regular Intervals Throughout System	To Individuals or Class Groups as Need Arises	One-Time or Sporadic Basis	Not Measured
(1)	(2)	(3)	(4)	(5)
Reading	51.5	11.4	10.6	21.5
Mathematics	50.6	11.9	11.6	19.5
Language Usage	48.3	11.6	9.6	24.3
Science	43.8	14.4	10.4	24.7
Social Studies	40.5	13.7	9.9	29.4
Literature	32.9	9.6	9.1	41.5
Study Skills	30.4	11.4	10.1	41.3
Modern Lang.	22.8	15.7	11.9	42.7
Business Ed.	17.5	16.4	7.6	51.9
Home Econ.	5.1	8.1	4.5	82.3
Industrial Arts	4.0	7.8	3.3	84.9
Music	3.8	7.1	4.3	84.8
Art	3.5	5.1	18.2	88.4

Table 82 reveals the extensiveness of standardized testing in various curriculum areas. Reading is the area receiving the greatest emphasis, followed by the general academic subjects. Very little testing occurs in home economics, music, and related fields. Standardized measures are not used as extensively at the senior high school level as in the elementary grades, the basic areas attaining approximately the 50 per cent level in the former, as compared with the 90 per cent level in the elementary grades. It is recognized that achievement testing at the secondary level poses many problems not encountered at the elementary level. Problems of curriculum validity, wide variety

of course offering, and varying course content emphasis dictate that standardized measures be used with caution and full knowledge of their limitations. In spite of these restrictions, findings reveal that the schools fall far short of an adequate program cited in the literature. This is true in both the basic skills areas and academic subject fields.

Attention is called to the significant number of respondents who did not mark items pertaining to this area of investigation. It may be assumed that no reply indicates the school system did not offer the skill or informational area being investigated, no testing was being done, or the testing was of such a minor nature as to preclude inclusion in the inquiry form. "No responses" are reflected in column (4) of Table 82. The investigation does not concern itself with the type of test. Measurement of functional knowledge or general educational development rather than specific subject content is not shown in the data.

Time of year of testing.-- The literature indicates that most testing takes place in the spring or fall. Advantages and disadvantages in each of these periods are noted. Most writers recommend fall testing but point out that scheduling is largely dependent upon the purpose for which the tests are to be used. Current practices surveys reveal that the large majority of schools test largely in the fall and spring, with the beginning of the school year being favored for most programs. Findings pertaining to this aspect are presented in Tables 83 to 85.

Table 83. Time of Year at Which Standardized Achievement Tests Are Generally Administered in the Elementary Schools

Time of Year	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
End of year or term	61.0	60.0	41.5	50.0	54.5	40.9	51.7
Beginning of year or term	26.0	22.8	40.8	25.0	54.5	36.4	33.2
Middle of year or term	9.0	15.7	19.2	30.0	9.1	31.8	17.0
Completion of topic or unit	1.0	2.8	3.8	2.5	-	4.5	2.5
No definite policy	2.0	4.2	1.5	5.0	-	4.5	2.5
Other	4.0	5.7	3.1	7.5	-	9.1	4.3

Table 84. Time of Year at Which Standardized Achievement Tests Are Generally Administered in the Junior High School

Time of Year	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
End of year or term	36.0	44.3	39.2	47.5	57.6	18.2	40.5
Beginning of year or term	16.0	12.8	22.3	30.0	33.3	45.4	22.0
Middle of year or term	3.0	12.8	20.8	22.5	6.1	27.3	14.7
No definite policy	2.0	5.7	3.8	5.0	-	9.1	3.8
Completion of topic or unit	1.0	2.8	1.5	-	3.0	-	1.5
Other	3.0	5.7	2.3	2.5	-	9.1	3.3

Table 85. Time of Year at Which Standardized Achievement Tests Are Generally Administered in the Senior High School

Time of Year	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
End of year or term	23.0	35.7	43.1	40.0	42.4	27.3	35.4
Beginning of year or term	4.0	11.4	20.8	30.0	39.4	45.4	18.7
Middle of year or term	6.0	10.0	13.8	20.0	9.1	13.6	11.4
No definite policy	4.0	11.4	6.9	5.0	-	4.5	6.1
Completion of topic or unit	1.0	2.8	4.6	2.5	3.0	4.5	3.0
Other	2.0	1.4	0.8	5.0	-	4.5	1.8

At all levels in the New England public schools end-of-the-year or term testing is most frequently practiced, followed by beginning of the year and middle of the year in the order named. The frequency of spring testing to fall testing is near the ratio of two to one for the total sample. This ratio is due to the influence of the smaller systems in the total sample, where end-of-the-year testing predominates. It is noted that the larger systems (categories IV, V) do more fall testing to the point where, in the large city system (category VI), it exceeds spring testing at the secondary level. It may be assumed that responses indicated for "other" denotes no definite pattern or plan.

Findings presented in Tables 83 to 85 imply that, with the exception of the large school systems, achievement tests are used more extensively for administrative and survey purposes than for instructional and classroom purposes.

Non-test techniques.-- The literature indicates that the pupil-centered testing and evaluation program, while leaning heavily on testing, must make full use of the other approaches to the appraisal of the individual, particularly in the area of personality evaluation. Non-test techniques are thus required in the evaluation process. The extent to which such techniques find use in the New England public schools is depicted in Tables 86 to 93.

Techniques such as the rating scale, anecdotal record, and sociometry include observation and are presented in Tables 86 to 89.

Table 86. Extent to Which Anecdotal Records Are Used in the School System

Extent Used	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	35.0	40.0	39.2	45.0	39.4	27.3	38.2
Individuals or class groups as need arises	30.0	30.0	38.5	35.0	45.4	53.6	36.4
Not used	8.0	7.1	7.7	5.0	6.1	-	6.6
No response	28.0	21.4	15.4	17.5	12.1	9.1	19.2

Table 87. Extent to Which Behavior or Personality Ratings Are Used in the School System

Extent Used	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	16.0	27.1	33.1	42.5	33.3	31.8	28.6
Individuals or class groups as need arises	23.0	21.4	30.8	27.5	48.5	40.9	28.9
Not used	14.0	15.7	12.3	10.0	3.0	9.1	11.8
No response	48.0	34.3	23.8	22.5	21.2	18.2	31.1

Table 88. Extent to Which Directed Observation Is Used in the School System

Extent Used	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at intervals throughout system	13.0	12.8	13.1	20.0	6.1	-	12.4
Individuals or class groups as need arises	24.0	32.8	43.8	42.5	63.6	59.1	39.2
Not used	14.0	10.0	9.2	7.5	6.1	4.5	10.4
No response	49.0	42.8	33.8	30.0	24.2	27.3	37.7

Table 89. Extent to Which Sociometric Techniques Are Used in the School System

Extent Used	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	1.0	5.7	3.1	-	-	4.5	2.5
Individuals or groups as need arises	10.8	18.6	26.9	50.6	42.4	63.6	26.8
Not Used	24.0	21.4	26.9	17.5	12.1	4.5	22.1
No Response	65.0	52.8	43.8	32.5	45.4	27.3	48.9

Schools make limited use of the techniques depicted in the four preceding tables. Anecdotal records are used more extensively than the other procedures. The large city system makes the greatest use of this technique, with 90 per cent of the systems reporting varying uses. With the noticeable exception of category VI (Table 86), about equal use on a "systematic basis at regular intervals throughout the system" and with "individuals or groups as the need arises" is noted. To attain maximum utility, anecdotes must extend over a period of time and include an adequate number of observations. It would seem that teachers must be trained in reading and interpreting these records, an adequate records system must be established, and teacher time must be provided to adequately administer the program. These factors must receive further investigation before any general conclusions can be drawn regarding the extensiveness and effectiveness of this technique of appraisal.

Behavior or personality ratings (Table 87) find greater use in the large schools than in the small schools, as evidenced by use reported in 81.8 and 72.7 per cent of the systems in categories V and VI, respectively, as compared to 39.0 and 48.5 per cent reported in the systems in categories I and II, respectively. The degree of use on a "systematic basis at regular intervals" or "with individuals or groups as the need arises" varies from category to category with no noticeable trend.

Directed observation as a technique is used primarily with individuals or groups as the need arises. Although all of the techniques depicted in Tables 86 to 89 are dependent upon observation, directed observation as a technique per se is not used as extensively as anecdotal records and behavior ratings. The observation technique has been developed primarily in connection with child study and finds greatest use in studying individual cases. This use is reflected in Table 88, where greatest use is reported under "with individuals or groups as the need arises." It is further noted that the use of this technique is greater in the larger school systems. Training and skill in observation cannot be overemphasized. Schools using this technique must be certain that teachers are properly trained prior to its general adoption.

Sociometry is a relatively new technique which can provide a teacher with information that will assist him in understanding the group and in working more effectively with the group. Its nature lends itself primarily to group work. This is reflected in findings reported in Table 89, where use is primarily "with individuals or groups as the need

arises." Again it is noted that the larger school system makes greater utility of this technique, only 11 per cent of the small systems reporting its use. This may be due to aid and encouragement provided by specialized personnel that is not available in the small school system.

The extent of use of the autobiography, problem checklist, and case study is noted in Tables 90 to 92.

Table 90. Extent to Which Autobiographies Are Used in the School System

Extent Used	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	7.0	7.1	16.9	10.0	3.0	13.6	10.6
Individuals or groups as need arises	15.0	18.6	21.5	25.0	33.3	31.8	21.3
Not used	20.0	20.0	23.8	27.5	18.2	13.6	21.3
No response	58.0	52.8	37.7	40.0	45.4	40.9	46.6

Table 91. Extent to Which Problem Checklists Are Used in the School System

Extent Used	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at regular intervals throughout system	8.0	5.7	7.7	15.0	6.1	9.1	8.1
Individuals or groups as need arises	12.0	18.6	25.4	32.5	21.2	40.9	22.0
Not used	22.0	20.0	24.6	17.5	27.3	18.2	22.3
No response	58.0	54.3	42.3	35.0	45.4	31.8	47.3

Table 92. Extent to Which Case Studies Are Used in the School System

Extent Used	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Systematic basis at intervals throughout system	5.0	8.6	4.6	10.0	15.1	-	6.6
Individuals or groups as need arises	26.0	38.6	60.0	65.0	72.7	90.9	50.9
Not used	13.0	11.4	6.1	12.5	3.0	-	8.9
No response	56.0	40.0	29.2	12.5	9.1	9.1	33.4

The autobiography and problem checklist represent self-report techniques and find primary use in the schools "with individuals or groups as the need arises," as noted in Tables 90 and 91. The autobiography may include spontaneous introduction of personal material

and serve as a basis for an interview between counselor and student. Its use may have greater utility in the larger system where the close personal contact between pupil and teacher may not prevail. This is reflected in Table 90 where it is noted that the use of this technique increases as the school system grows in size.

Problem checklists find greater use in the larger system and there primarily "with individuals or groups as the need arises." Varying use of this technique is reported by only 20 per cent of the small schools, while 50 per cent of the large city systems report its use. Many writers point to serious limitations and inherent dangers in the use of this technique. Writers cite that these measures should be used with extreme caution and sound judgment. Data revealing only limited use by the schools imply that this belief prevails in the majority of the New England public schools.

Case study is taking on increasing importance as a tool of pupil appraisal and is advocated in the literature as a means of synthesizing pertinent facts about a pupil in one framework. Table 92 reveals that the large school system makes extensive use of this technique "with individuals or groups as the need arises," as reported by 90.9 per cent of these schools. The extensiveness of use decreases as the system becomes smaller, with less than one third of the schools in category I reporting the use of the case conference. The case conference requires the "team" approach and varied pupil personnel functionaries, together with teachers, sitting down to consider diagnosis and treatment of the pupil under consideration. This technique implies the availability of

specially trained personnel to aid in this process. The absence of these functionaries in the small system may limit its use.

A general overview of non-test techniques is presented in Table 93.

Table 93. Extent to Which Non-Testing Techniques Are Used in the School System

Technique	Percentage of Responses According to Use		
	Systematic Basis at Regular Intervals Throughout System	With Individuals or Groups as Need Arises	Measures Not Used
(1)	(2)	(3)	(4)
Anecdotal Records	38.2	36.4	25.3
Behavior or Personality Ratings	28.6	28.9	42.9
Directed Observation	12.4	39.2	48.1
Autobiography	10.6	21.3	67.9
Problem Checklist	8.1	22.0	69.6
Case Study	6.6	50.9	42.3
Sociometric Techniques	2.5	26.8	71.0
Others[a]	1.0	-	98.5

a/In order of decreasing frequency: projective techniques, situation-response type test.

Findings presented in Table 93 and the preceding seven tables reveal that non-testing techniques are not used as extensively as objective testing devices. It is noted that observational techniques generally find greater use than self-report techniques. With the exception of anecdotal records, all techniques are used "with individuals or groups as the need arises," rather than on a systematic basis throughout the system. Minor use of projective techniques and situation-

response type test is noted. Until such time as additional specialized personnel is available, schools will not make extensive use of these techniques. The literature reveals that the whole area of personal-social development, involving non-test techniques, is in need of further development and exploration. This situation is true throughout the country and is not unique to the New England area. Data presented in this study point up the need for the schools to make more extensive use of these measures in appraising the "total" pupil. Although findings indicate advances in the larger systems where specialized personnel is employed, need for greater use still remains in this area.

A significant number of "no response" is noted in Tables 86 to 92. Failure to respond to these items implies non-use or such insignificant use that it may be considered non-use. "No responses" are thus included in Table 93 as "Measure Not Used" in column (4).

Locally-constructed achievement tests.-- Locally-constructed achievement tests should represent an integral part of the achievement testing program. Since this device is more closely related to the local instructional objectives, its use represents a definite factor in the improvement of instruction and evaluation. The literature strongly recommends these measures. Current practices surveys reviewed in this study,[1/] however, reveal that very few systems use these measures. Table 94 indicates the use of this type of measure in the New England public schools.

1/Chapter II, supra.

Table 94. Extent to Which Achievement Tests Are Constructed Locally for Use on a System-Wide Basis

Extent of Use	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Never	72.0	78.5	73.8	52.5	51.5	54.5	69.1
Occasionally	15.0	11.4	16.9	22.5	27.3	40.9	18.2
Regularly	8.0	5.7	2.3	17.5	15.1	4.5	7.1

Findings presented in Table 94 indicate that locally-constructed achievement tests find little use on a regular basis in the schools, only 7.1 per cent reporting such use. Although a significant number of large systems, 40.9 per cent, report occasional use, the New England public schools as a whole make far too little use of this type of test, approximately 70 per cent reporting no use. This is an area requiring attention and in need of strengthening.

Areas requiring attention.-- In order to determine opinion regarding adequacy of present measuring instruments and need for additional measures, the inquiry form asked the following question:

> "In what areas, not adequately covered at the present time by available standardized tests, are tests and measuring devices most needed in the school system?"

A variety of replies was received from respondents. The responses received most frequently are listed below in order of decreasing frequency:

1. Tests to measure educational development at high school level
2. Special aptitude tests, particularly in fields of music and art

3. More suitable tests in elementary science
4. More adequate measures in the field of social studies
5. Reliable personality measures
6. Work-study skills tests
7. Additional skills tests in the primary grades
8. More adequate interests measures
9. Diagnostic arithmetic measures
10. Measures in home economics.

It is evident from the responses that educators are concerned with a universe of measurement problems covering the whole gamut of the testing field. Although the variety of responses precludes any conclusion, a few implications may be gathered from the data. These are noted below:

1. Apparently many educators in the field may not be aware of some of the more recent tests that are available, as evidenced by responses 1, 6, 7, and 9 noted above.
2. Because of instability and flux in the fields of elementary science and social studies content, validity becomes a problem with many schools, as implied in responses 3 and 4.
3. The measurement of personality, interest, and special aptitude remains the concern of many educators. Until such time as more valid and reliable measures are devised, this will continue to be a problem area not susceptible to ready measurement (items 2, 5, and 8).

5. Organizing and Reporting Test Results

Circulating intelligence and standardized achievement test results.-- A preliminary step to test interpretation and utilization is organizing the results in understandable form and circulating these data to school functionaries concerned with the test results. Tables 95 and 96 reveal the extent to which these data are made available to school personnel.

Table 95. Extent to Which the Results of Intelligence Tests Are Made Available to Teachers and Staff Members

Extent Made Available	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Results are regularly circulated to each teacher or staff member concerned with class group or subject tested	68.0	82.8	57.7	75.0	67.5	60.6	70.4
Results are occasionally circulated to each teacher or staff member concerned with class group or subject tested	19.0	14.3	30.0	25.0	12.5	12.1	22.0
Results are non-circulating, but are available to teachers or staff members upon request	6.0	1.4	8.5	7.5	-	-	5.3
Results are kept solely for administrative or guidance personnel use	4.0	1.4	5.4	2.5	2.5	-	3.5

Table 96. Extent to Which the Results of Standardized Achievement Test Results Are Made Available to Teachers and Staff Members

Extent Made Available	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Results are regularly circulated to each teacher or staff member concerned with the class group or subject tested	71.0	84.3	72.3	87.5	84.8	86.4	77.5
Results are occasionally circulated to each teacher or staff member concerned with the class group or subject tested	19.0	11.4	16.9	12.5	9.1	13.6	15.2
Results are non-circulating, but are available to teachers or staff members upon request	5.0	-	7.7	7.5	3.0	4.5	5.1
Results are kept solely for administrative or guidance personnel use	2.0	1.4	3.1	2.5	3.0	-	2.3

Approximately 70 per cent of the systems regularly circulate intelligence test results, while 77.5 per cent report the practice of regularly circulating the results of standardized achievement tests. Although schools in category IV follow this practice to a greater extent than other schools, no noticeable relationship between size of school and practice is evident. The literature recommends that maximum utilization of test results dictates the necessity of getting these data into the hands of those individuals concerned with the tests as quickly as possible after testing. This requires a program for regularly circulating the results to all school functionaries. Approximately 75 per cent of the schools report this procedure. This procedure dictates that all functionaries understand and make full utiliza-

tion of the results once they are in their hands.

Person to whom results are made available.-- Test interpretations must reach those who work directly with pupils, the pupils themselves, the parents, and those who make decisions affecting the curriculum. Tables 97 to 102 present data pertaining to practices in this aspect of the testing program.

Intelligence and achievement tests form the core of the group testing and have the greatest utilization in the schools. Data pertaining to these tests are presented in Tables 97 and 98.

Table 97. Person to Whom the Results of Intelligence Tests Are Made Available

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Principal	92.0	90.9	98.4	97.5	100.0	95.4	95.2
Classroom teacher	85.0	91.4	92.3	92.5	97.0	90.9	90.6
Guid. personnel	12.0	31.4	74.6	87.5	90.9	86.4	54.4
Supervisor	22.0	32.8	52.3	45.0	81.8	72.7	44.0
Schl. psychologist	5.0	11.4	23.1	42.5	66.7	81.8	25.3
Parent	9.0	12.8	14.6	22.5	21.2	9.1	13.9
Pupil	3.0	14.3	10.0	17.5	18.2	13.6	10.6

Table 98. Person to Whom the Results of Standardized Achievement Tests Are Made Available

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Principal	91.0	85.7	98.4	95.0	100.0	95.4	93.9
Classroom teacher	89.0	94.2	94.6	92.5	97.0	90.9	92.9
Guid. personnel	12.0	31.4	73.8	82.5	90.9	86.4	53.7
Parent	40.0	47.1	60.8	47.5	51.5	40.9	49.9
Supervisor	22.0	32.8	55.4	40.0	84.8	72.7	44.8
Pupil	28.0	38.6	53.8	45.0	45.4	45.4	42.0
Schl. psychologist	5.0	11.4	25.4	35.0	66.7	77.3	25.1

Results of both intelligence and standardized achievement tests are universally made available to the building principal and the classroom teacher, upwards of 90 per cent of the school systems reporting this practice (Tables 97 and 98). These data are usually made available to the guidance specialists and other school functionaries in most cases. The lower percentage of responses under categories I and II for specialized personnel may be due to the fact that these systems do not employ such personnel. The predominance of the small school system in the New England area may thus unduly weigh the percentage of responses for these school functionaries in the total New England sample noted in column (8).

Only 13.9 per cent of the schools make intelligence test results available to parents. Those who follow this practice report that parental conferences are held for general understanding, results are usually transmitted only in general terms and not exact scores, and

extreme caution is exercised in discussing results. A larger proportion of schools, approximately 50 per cent, make standardized achievement test results available to parents. Misinterpretation of test results by parents poses a grave danger. Schools need to give added attention to this phase of the testing program and embark on a program of parent education aimed at allowing for more extensive sharing of results with parents.

Intelligence and standardized achievement test results are made available to pupils by 10.6 and 49.9 per cent of the school systems, respectively. Pupils need to have a knowledge of their capacities, learning advancement, strengths, and weaknesses. This is not possible unless results are shared with pupils. Findings reveal that schools are woefully weak in this practice. Cautions and limitations regarding test interpretations are cited throughout the literature. These factors should not deter schools from embarking on a program of taking pupils into partnership in utilizing test results as soon and as completely as is feasible. This is particularly true in the area of academic aptitude and achievement.

Special aptitude tests, interest inventories, and personality tests are not used as extensively as standardized measures of academic aptitude or achievement. They are integral parts of the total testing program. Findings pertaining to circulating the results of these measures are presented in Tables 99 to 101.

Table 99. Person to Whom the Results of Special Aptitude Tests Are Made Available

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Principal	43.0	64.3	72.3	82.5	78.8	68.2	64.8
Classroom teacher	39.0	62.8	60.0	75.0	84.8	68.2	59.2
Guid. personnel	6.0	31.4	65.4	85.0	84.8	81.8	48.9
Pupil	21.0	41.4	46.1	42.5	68.6	45.4	40.0
Parent	24.0	37.1	43.8	45.0	60.6	36.4	38.7
Supervisor	8.0	18.6	35.4	35.0	66.7	45.4	28.6
Schl. psychologist	2.0	11.4	16.9	45.0	63.6	63.6	21.5
No response	57.1	28.6	20.8	10.0	9.1	18.2	29.1

Table 100. Person to Whom the Results of Interest Inventories Are Made Available

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Principal	36.0	57.1	66.9	65.0	78.8	59.1	57.7
Classroom teacher	31.0	47.1	56.9	62.5	84.8	59.1	51.6
Guid. personnel	7.0	32.8	67.7	75.0	84.8	77.3	48.9
Pupil	23.0	41.4	53.1	55.0	60.6	40.9	43.5
Parent	21.0	31.4	50.8	47.5	60.6	31.8	39.2
Supervisor	7.0	12.8	30.8	20.0	63.6	40.9	23.8
Schl. psychologist	1.0	11.4	16.1	32.5	60.6	50.0	18.7
No response	62.0	37.1	19.2	22.5	12.1	18.2	32.9

Table 101. Person to Whom the Results of Personality Measures Are Made Available

Person	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Principal	27.0	47.1	52.3	55.0	51.5	45.4	44.8
Guid. personnel	6.0	27.1	53.8	67.5	63.6	54.5	39.2
Classroom teacher	28.0	42.8	40.8	47.5	51.5	40.9	38.2
Supervisor	8.0	11.4	24.6	25.0	45.4	27.3	20.0
Parent	10.0	22.8	19.2	22.5	27.3	-	17.5
Pupil	6.0	21.4	23.8	25.0	15.1	-	17.0
Schl. psychologist	1.0	8.6	14.6	32.5	51.5	45.4	16.7
No response	72.0	45.7	36.9	32.5	36.4	40.9	47.0

Special aptitude tests, interest inventories, and personality measures are used extensively in the guidance process. Tables 99 to 101 reveal that results of these measures are generally made available to school functionaries with the guidance specialists being mentioned most frequently by schools having such services. The low percentage responses pertaining to school psychologist, supervisor, and guidance specialist, noted particularly in the smaller school systems (categories I and II) implies that such functionaries are not available in these systems. Results of special aptitude and interest measures are made available to parents in the majority of systems. The smaller systems (categories I and II) do not follow this practice to the extent of the larger systems. Results of personality measures are made available principally to school functionaries and are given to parents and pupils only to a limited degree. Previous findings reveal that a significant number of schools did not make extensive use of personality,

interest, and special aptitude measures. This is reflected in "no responses" noted in the three preceding tables. A general overview of the findings is presented in Table 102.

Table 102. Person to Whom the Results of Various Standardized Measures Are Made Available

Person	Percentage of Responses According to Type of Measure				
	Intelligence Tests	Standardized Achievement Tests	Special Aptitude	Interest Inventories	Personality Tests
(1)	(2)	(3)	(4)	(5)	(6)
Principal	95.2	93.9	64.8	57.5	44.8
Supervisor	44.0	44.8	28.6	23.8	20.0
Guidance counselor	54.4	53.7	48.9	48.9	39.2
School psychologist	25.3	25.1	21.5	18.7	16.2
Classroom teacher	90.6	92.9	59.2	51.6	38.2
Parent	13.9	49.9	38.7	39.2	17.5
Pupil	10.6	42.0	40.0	43.5	17.0

Table 103 reveals that the results of standardized measures are generally made available to school functionaries. The principal and classroom teacher are mentioned most frequently as receiving results of intelligence and standardized achievement tests, with the guidance specialists, together with these individuals, most frequently mentioned for special aptitudes, interest, and personality measures. The lower percentage of responses pertaining to supervisor, school psychologist, and guidance counselor is due in part to the fact that many of the small schools included in the total sample do not number such specialists among their staff. Attention is called to previous findings which

indicate that measures of special aptitudes, interests, and personality are not used as extensively as intelligence and standardized achievement tests. Percentage of responses in columns (4), (5), and (6) are thus somewhat lower than those in columns (1) and (2), since percentages are figured on the total sample of schools participating in the study. Data presented indicate, however, that test results are generally made available to school functionaries requiring such results.

Findings indicate that schools are not adequately making test results available to parents and pupils. A glaring example of this fact is noted in Table 102, column (2), where results are made available to parents and pupils by 13.9 and 10.6 per cent of the schools, respectively. The literature strongly advocates the presentation of test data to parents and pupils with appropriate explanation and interpretation. To accomplish this task, schools must embark upon a program to instruct and inform parents in bare essentials of test understanding and score interpretation. Pupils must also be given training in test interpretation.

Schools evidently are not carrying out this aspect of the testing program. This problem is not unique to the New England public schools, as evidenced by previous surveys reviewed in this study.[1/] If test data are to function in the lives of pupils and assist them in personal assessment and educational growth, schools must make results available to all those concerned with the educational process.

1/Chapter II, *supra*.

Form in which test results are made available.-- The literature advocates that test results be presented in a meaningful way so that the strengths, weaknesses, patterns, and trends may be easily discernible to all school functionaries. The practice followed in the schools for presenting intelligence and standardized achievement test results is noted in Tables 103 and 104.

Table 103. Form in Which Intelligence Test Results Are Generally Made Available to Teachers and Staff Members

Form	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Class record sheet giving total scores of pupils	61.0	64.3	59.2	62.5	48.5	50.0	59.5
Individual pupil profile chart	48.0	42.8	42.3	45.0	57.6	31.8	44.8
Class record sheet giving part and total scores of pupils	33.0	32.8	26.1	40.0	39.4	36.4	32.1
Class or group profile charts	30.0	32.8	31.5	37.5	30.3	36.4	32.1
Others[a/]	3.0	7.1	10.0	5.0	15.1	27.3	8.6

a/In order of decreasing frequency: individual pupil permanent record card giving total score; individual test booklet; mimeographed summary sheets giving scores for various schools as well as grade, class, and city medians; quintile charts.

Table 104. Form in Which Standardized Achievement Results Are Generally Made Available to Teachers and Staff Members

Form	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individual pupil profile charts	65.0	62.8	66.1	45.0	60.6	45.4	61.6
Class record sheet giving total scores of pupils	54.0	57.1	55.4	60.0	39.4	45.4	53.9
Class or group profile charts	34.0	41.4	41.7	50.0	36.4	45.4	42.3
Class record sheet giving part and total scores of pupils	38.0	34.3	33.8	35.0	48.5	36.4	36.4
Others[a/]	2.0	5.7	10.8	2.5	18.2	27.3	8.3

a/In order of decreasing frequency: individual pupil permanent record card, mimeographed summary for schools throughout the system, cumulative record card giving total scores.

Class record sheets giving total scores of pupils and individual pupil profile charts are mentioned most frequently as the forms used for recording intelligence test results, 59.5 per cent reporting use of the former and 44.8 per cent of the latter. Approximately one third of the schools report the use of class record sheets giving part and total scores of pupils, and class or group profile charts. The use of individual profile charts is strongly advocated by the literature, particularly if the multiscore type is used for studying the individual's performance. This type of measure would almost dictate the use of the class record sheets giving part and total scores of pupils. Responses provided in Table 103 may imply that only approximately one

third of the schools administer multiscore intelligence tests.

It is noted that individual profile charts are most frequently used in recording achievement test scores. These are usually provided by test publishers. It is also significant to note that 42.3 per cent of the schools report the use of class or group profiles. Some schools report the use of several forms.

The aim of the recording system is to place the results in the hands of the staff in a meaningful and comparable basis which is understandable and easy to use. Individual pupil profile charts and class or group profile charts meet these criteria and provide the best means of analyzing individual pupil scores and composite group scores. These means may well serve as an added stimulus for effecting improvements in curriculum emphasis, teaching procedures, and methods of study. Findings imply that the New England public schools need to develop more fully this method of presentation. Too many schools report simply the use of class record sheets giving total scores.

The literature reveals that, in addition to providing group data in understandable form to school personnel, a systematic plan for organizing pertinent test data about the individual pupil on a cumulative basis is essential. The practice followed by the New England public schools in recording test data for the individual pupil is presented in Table 105.

Table 105. Recording the Test Results of Individual Pupils

Practice	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Recorded on pupil cumulative permanent record card	85.0	90.0	90.8	92.5	97.0	95.4	90.1
Recorded on individual pupil profile card	27.0	24.3	32.3	25.0	27.3	31.8	28.3
Recorded on individual pupil test record sheet or card covering all tests administered at given gr. level	29.0	21.4	24.6	40.0	42.4	22.7	28.1
Others[a]	1.0	1.4	1.5	5.0	6.1	13.6	2.8

a/In order of decreasing frequency: test covers and profile sheets, recorded on summary sheet, IBM typed sheets.

Approximately 90 per cent of the schools report that results are recorded on the pupil's permanent cumulative record card. This practice is generally followed in other schools throughout the country[1/] and follows procedures advocated in the literature. Approximately 28 per cent of the schools report the use of an individual pupil profile card and/or an individual pupil test record sheet or card. These are kept in the individual cumulative folder and are advocated by writers as providing a ready means with a high "glance value" for studying the pupil's cumulative test record. There appears to be a recent trend for schools to make more use of these forms. This is particularly apparent in schools in category V.

1/Chapter II, *supra*.

Using derived scores.-- Individual test scores acquire meaning only when they are compared with the scores of well-identified groups. Practices followed by the New England public schools in deriving scores and in developing norms are indicated in Tables 106 and 107.

Table 106. Derived Scores in Which Standardized Achievement Tests Are Generally Recorded in the School Record

Derived Scores	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grade Equivalents	90.0	92.8	88.4	95.0	90.9	95.4	90.9
Percentile Rank	42.0	58.6	58.5	60.0	75.7	54.5	55.7
Age Equivalents	36.0	32.8	18.5	35.0	36.4	27.3	29.1
Standard Scores	27.0	25.7	19.2	27.5	27.3	31.8	24.5
Stanines	3.0	2.8	3.1	7.5	6.1	9.1	4.0
Others[a]	1.0	-	-	-	-	-	0.7

a/In order of decreasing frequency: achievement quotient, numerical standing in class, local quartiles, T scores.

Table 106 reveals that grade equivalents are the most widely used derived scores, 90.9 per cent of the schools reporting their use, followed by percentile rank, age equivalent, and standard scores. Stanines find little use in the New England public schools, only 4 per cent of the schools reporting the use of these scores. This is undoubtedly due to the recent development of this type of score. Responses also indicate that some systems make use of several types of derived scores. Current practices surveys[1/] reveal that grade equivalent, age equivalent, and percentile rank are the most commonly used types, while the

1/Chapter II, supra.

literature generally lists age, grade, percentile, and standard score as the most frequently used type of norms in the schools today. Findings presented in this study reveal that the grade equivalent and percentile rank are used much more extensively in the New England public schools, regardless of size, than other types of derived scores.

Data pertaining to the type of norm being used in the New England public schools are included in Table 107.

Table 107. Norms Generally Used in Interpreting Standardized Test Data

Norm	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Norms furnished by test publisher	93.0	95.7	93.8	97.5	97.0	86.4	94.2
Norms computed on local level for school system	8.0	17.1	19.2	30.0	48.5	50.0	21.3
Norms computed on regional basis	11.0	15.7	27.7	20.0	21.2	4.5	18.8

Table 107 reveals that the majority of school systems make use primarily of national norms furnished by the test publisher, 94.2 per cent of the schools reporting this practice. Data also indicate that some schools make use of more than one norm. This is particularly true of systems in categories III and IV. It is noted that local norms find primary use in the larger systems (categories V and VI). The literature advocates the use of local norms as supplementary to the national norm provided by the test publisher. This applies to the

small as well as the larger school system. Although the larger school district is making limited use of local norms, only 21.3 per cent of the total sample report their use. Regional norms find less use than local norms. Data reveal that the public schools in the New England area can well expand the use of both the local and regional norms.

6. Using the Results of the Testing and Evaluation Program

The literature underlines the fact that the heart of the testing and evaluation program is the use that is made of the results of the measures of that program. Many and varied uses may be made of measures. All members of the staff may utilize the results of the testing program for administrative purposes, instructional purposes, educational and vocational guidance, and research. Since the results of various measures are interrelated and difficult to pinpoint, many purposes and uses may be served by the same data. Findings presented in this study pertain to the main administrative and supervisory uses reported for measures of academic aptitude, achievement, special aptitude, interest, and personal-social adjustment.

Academic aptitude or intelligence tests.-- The main administrative and supervisory uses of the results of intelligence tests are revealed in Table 108.

Table 108. Main Administrative and Supervisory Uses Being Made of the Results of Intelligence Tests

Use	Percentage of Responses		
	Common Practice	Occasional Use	Not Used
(1)	(2)	(3)	(4)
To identify those pupils achieving significantly below their capacity to achieve	68.6	26.1	5.6
To aid in grouping within grades	64.0	23.3	12.9
To aid in placement of pupils in grades	53.2	23.3	23.8
To aid in determining the expectancy level in achievement within the class group and schools	47.8	29.4	23.7
To assist in planning special programs to meet the needs of individual pupils and class groups	45.8	31.9	22.3
To determine how different groups compare in ability levels from school to school and within schools	31.4	26.1	42.5
Other uses	8.6	38.0	87.6

Other varying uses are reported by respondents. As indicated in Table 108, these do not receive a high degree of emphasis or utility, only 8.6 per cent of the schools reporting their use as a common practice (column 3). Those most frequently appearing in responses are listed in order of decreasing frequency:

1. To identify exceptional pupils for special class and honors groups
2. To identify over- and underachievers
3. To aid in educational and vocational guidance
4. To aid in determining promotion and nonpromotion

5. To aid in program evaluation and curriculum planning

6. To aid in determining expectancy level of the school system

7. To aid in making referrals for psychological clinics, etc.

Findings disclose that many uses are being made of intelligence test results by the New England public schools. Major uses reported in this study are pupil identification, grouping and placement, and curriculum planning, with emphasis in the order named. The majority of schools report greater use in aiding the individual's school adjustment and development as contrasted with the broader aspect of over-all program planning, curriculum evaluation and adjustment, and the evaluation and improvement of instruction. The literature advocates full utilization in all these areas. No use is reported for educational research or interpreting the schools to the public. Responses in columns (3) and (4) point clearly to the fact that many school systems are not making full utilization of test results.

Standardized achievement test.-- Practices pertaining to the uses made of standardized achievement test results are reported in Table 109.

Table 109. Main Administrative and Supervisory Uses Being Made of the Results of Standardized Achievement Tests

Use	Percentage Responses		
	Common Practice	Occasional Use	Not Used
(1)	(2)	(3)	(4)
To identify instructional areas needing special emphasis	71.1	21.0	7.8
To aid in counseling of individuals regarding educational and vocational planning	63.3	22.8	13.9
To reveal extent to which instructional objectives are being attained	59.5	22.8	17.7
To aid in grouping within the grade or class group	58.2	25.6	16.2
To aid in placement of pupils in grades	53.4	24.0	22.5
To determine how achievement of local groups compares with that of other systems	46.8	24.5	28.6
To aid in curriculum revision and evaluation	31.6	34.4	33.9
To interpret the work of the school to the public	25.6	39.2	35.2
To compare the achievement of different class groups within the system	18.5	23.5	58.0
To evaluate different teaching methods and materials of instruction	11.4	27.8	60.7
To rate the proficiency of different teachers	2.8	14.7	82.5
Others	4.0	2.8	93.2

In addition to the uses stated in Table 109, respondents indicate many other varied uses of standardized achievement tests. Listed in order of decreasing frequency, they are:

1. To identify under- and overachievers
2. To determine proper electives within the program

3. To inform parents of pupils' progress in certain areas
4. To identify gifted pupils
5. To aid in planning special programs for the slow learner and the gifted
6. To aid in establishing special groups
7. To provide informational material for school board
8. To compare teachers' achievement ratings with those of test ratings
9. To aid in determining promotion and nonpromotion
10. To serve as a teaching device with teachers.

Uses listed in the accompanying table point to the following major areas of use of achievement test results: evaluation and improvement of instruction, grouping and placement of pupils, curriculum evaluation and revision. Wide use is made of results as noted by the variety of responses. The relatively low percentage of responses in column (2) denotes the fact that many schools are not making maximum use of the results. Responses for use, "To evaluate different teaching methods and materials of instruction," is cited as one example. Certainly more than 11.4 per cent of the schools should take advantage of the data provided by the results of these measures to study and improve methodology and teaching materials.

Attention is called to the responses noted for "To rate the proficiency of different teachers." It is gratifying to note that 93.2 per cent of the schools recognize this malpractice. Several administrators expressed concern over this item and other areas of test

utilization, as evidenced by the following comment extracted from one inquiry form:

> "We are extremely careful to defend the testing program from any suggestion that it will be used in any way to evaluate individual teachers and from the danger that the goals of instruction come to be the outcomes measured by standardized instruments."

Although standardized achievement test results are being put to many uses, responses clearly indicate that many schools are not making full utilization of these measures. Percentage responses recorded in columns (3) and (4) reveal that a large proportion of the schools are merely scratching the surface in receiving maximum benefit of the testing program.

Special aptitude tests, interest inventories, and personality measures have only limited use in the New England public schools (Tables 53, 55, 57) and find primary utility for guidance purposes. These measures, together with other data, can be used to great advantage by administrators and supervisors concerned with the total educational program. Major administrative and supervisory uses made of the results of these measures by the New England public schools are indicated in Tables 110 to 112.

Special aptitude tests.-- Table 110 provides data pertaining to uses made of special aptitude test results.

Table 110. Main Administrative and Supervisory Uses Being Made of the Results of Special Aptitude Tests

Use	Percentage Responses		
	Common Practice	Occasional Use	Not Used
(1)	(2)	(3)	(4)
To aid in the placement of pupils in grades and for grouping within the grade or class	23.5	24.5	51.9
To place pupils in remedial or corrective classes	18.2	22.5	59.2
To determine the needs for special school programs	13.9	18.7	67.3
Others[a]	7.8	7.6	84.5

a/In order of decreasing frequency: to aid the counseling procedure; to aid in educational and vocational planning; to aid in course selection at the senior high school; to aid placement of pupils in trade and vocational schools; to aid in determining proper electives with the program of studies; to aid in setting up accelerated class programs; to aid in identifying special talents; to aid in job placement; to aid in establishing special school and out-of-school programs.

The primary use made of the results of special aptitude tests is for pupil placement and grouping, with minor use being reported for curriculum revision and evaluation. Other uses reported may be considered primarily in the general area of individual pupil guidance. It is somewhat disturbing to note that less than one fourth of the schools report that it is common practice to use the results of special aptitude tests for the more commonly accepted administrative and supervisory uses. It is evident that the majority of the schools are making only minor use of these measures for administrative and supervisory purposes.

Interest inventories.-- Administrative and supervisory uses made of the results of interest measures are presented in Table 111.

Table 111. Main Administrative and Supervisory Uses Being Made of the Results of Interest Inventories

Use	Percentage Responses		
	Common Practice	Occasional Use	Not Used
(1)	(2)	(3)	(4)
To provide a partial basis for parent-pupil-teacher and/or counselor conferences on educational and vocational planning	45.8	22.3	32.1
To identify special interest areas coinciding with special or high level general ability	27.6	22.0	50.4
To provide a basis for an occupation unit in various courses	9.1	14.4	76.4
Others[a]	22.8	2.5	95.2

a/In order of decreasing frequency: to aid in educational and vocational planning; to aid in counseling; to aid in English department in assigning outside reading; to aid pupils in subject choices (exploratory study); to aid in selecting avocational activity; to provide information to outside agencies for the terminal student seeking employment, etc.; to aid in school referrals.

Interest inventories find major use in the general area of individual pupil guidance as evidenced by the responses provided in Table 111 (item 1). Responses to items 2 and 3 in the table imply some use for curriculum evaluation and revision. As indicated in Table 55, interest inventories have limited use in the New England public schools. Those schools that do include these measures in the testing program make only minor uses of the results for administrative and supervisory purposes.

Personality measures.-- Major administrative and supervisory uses reported for personality measures are depicted in Table 112.

Table 112. Main Administrative and Supervisory Uses Being Made of the Results of Personality Measures

Use	Percentage Responses		
	Common Practice	Occasional Use	Not Used
(1)	(2)	(3)	(4)
To determine the need for special attention to certain areas of the curriculum	5.1	14.4	80.5
To determine the personal and mental health problems which characterize class groups, schools, or school districts	4.5	19.0	76.4
To determine class groups or schools that need special assistance in the social school situation	4.5	13.2	82.3
Others[a]	7.1	8.6	84.3

a/In order of decreasing frequency: used chiefly in individual counseling situations and case work--individual analysis; to discover basic emotional needs in individual cases; to aid in referring individual problem cases to specialists; to aid in identifying pupils needing special help; to aid in appraising individual maladjustments; to aid in determining personal and mental health problems which characterize individuals; to aid counselors to give direction to interviews with emotionally disturbed pupils.

Personality measures have very limited use in the New England public schools (Table 57). Many respondents indicate that group measures are rarely used, and then only with parental permission. Findings further indicate that these measures are used chiefly by guidance counselors and specialists. It thus becomes evident that results of these measures are used principally for individual pupil guidance and

case work. A review of Table 112 shows this fact to be very apparent, the large majority of schools reporting "not used" (column (4)) for the items enumerated in the table. Limited use of the results of these measures for administrative and supervisory purposes will continue until these measures have greater use in the schools.

Other uses.-- Many and varied uses are being made of the results of the testing program. Findings reported previously reveal that tests serve many purposes, several uses or purposes being satisfied by the same data. Standardized measures are not the exclusive tool of any one branch of the profession. To attain maximum utilization, these measures must be considered in relation to other available data.

As a further means of determining the use made of test results, respondents were asked to cite specifically anything that has been done in the past two years by way of use of the results of the testing program, not noted in responses previously recorded in the inquiry form.

Upwards of 375 separate comments were received. These comments represent specific ways in which test results have been used at the local level in affecting significant improvements and changes in the total school program. All comments were carefully analyzed and tabulated, and may be found in Appendix H. A study of these comments reveals that test results are being used in many and varied ways at the local level and are influencing many aspects of the school program. Findings in this study reveal that comments may be categorized under the following five administrative and supervisory uses:

1. An aid for expanding educational services and effecting major improvements in the instructional program
2. An aid for initiating program expansion and instituting enrichment activities
3. An aid for initiating curriculum revision
4. An aid for instituting changes in methodology and instructional procedures
5. A public relations medium--an aid for informing the public of basic need of the local school system.

CHAPTER VI

SUMMARY, RECOMMENDATIONS, AND LIMITATIONS

Responses received on the 395 inquiry forms analyzed in the present study reveal certain discernible trends and patterns in the pupil-centered testing and evaluation program in the New England public schools. This chapter summarizes the general findings of the study, sets forth certain conclusions and recommendations pertaining to present practices, and makes specific recommendations for further research and study in the area under consideration. The reader is referred to Chapter IV for detailed findings pertaining to various aspects of the appraisal program.

1. Summary of Findings

General program.-- Major findings pertaining to general aspects of the over-all program in the New England public schools are:

1. Wide diversity of practices is noted in the pupil-centered testing and evaluation programs ranging from a comprehensive program to the traditional program hopelessly mired in the past.
2. The expenditure per pupil (exclusive of salaries) for the pupil appraisal program ranges from less than twenty-five cents to slightly more than $2.00, with a median of forty cents, the 25th percentile at twenty-five cents, and the 75th percentile at sixty-two cents.

3. The larger school systems provide a more comprehensive appraisal program than the small school systems, both in terms of areas of evaluation and specialized personnel to aid in the program.

Organizing and administering the program.-- Salient features of the organizational and administrative aspect of the program are noted in the findings. These are summarized below.

1. The general school administrators are responsible for directing the pupil appraisal program in the majority of the schools. This responsibility shifts to pupil personnel functionaries in the larger school systems, with the guidance director most frequently assuming this responsibility. Many varying titles are reported for the functionary directing the appraisal program.
2. Less than 5 per cent of the school systems report a full-time director of the testing and evaluation program.
3. In approximately one third of the school systems the person responsible for directing the testing and evaluation program holds an advanced degree with specialization in the field of tests and measurements.
4. It is not common practice for school systems to have a school-wide testing committee to aid in program planning and development, only 5 per cent of the systems reporting such a committee.
5. Scheduling and test selection are done mainly by the general school administrators. This responsibility shifts to pupil personnel functionaries in the larger school systems.

6. The classroom teacher most frequently administers group measures in the standardized testing program.
7. The majority of schools hand-score standardized group tests at the local level. This task is usually done by the classroom teacher.
8. Very limited use is made of outside consultative services. The majority of school systems indicate that they would use such services if they could be made available. The small school system evidences the greatest desire in this area.
9. The New England public schools do not participate to any great extent in state-wide or regional testing programs, only 13 per cent reporting regular participation.
10. Approximately one half of the school systems have an in-service program in measurements for teachers. The larger school systems are more active in this area than the small systems.
11. During the past five years the pupil-centered testing and evaluation program has steadily expanded in approximately 70 per cent of the school systems. Fifty-eight per cent of the systems reported that this was brought about chiefly because of recognition of new needs.
12. The major problem or weakness reported by the schools centers around test utilization and staff limitations.

Measurement devices used in the program.-- Certain trends and practices pertaining to the general areas of pupil behavior being

evaluated by standardized measures are summarized below.

1. The major emphasis in the pupil-centered testing and evaluation program is upon tests.
2. Intelligence tests and standardized achievement tests form the hard core of the group testing program, both measures being administered on a systematic basis at regular intervals throughout the school systems.
3. Group tests of intelligence are usually administered at grades 1, 3, and 6 or 7. The majority of schools use individual intelligence tests to retest and check questionable cases, the larger school systems leading in this practice.
4. Tests of special aptitude, interest, and personality are available for special needs or purposes but are not used extensively on a systematic basis in the majority of schools. Their use is largely limited to the secondary level.
5. Standardized measures in reading, arithmetic, language usage, and spelling are used regularly on a systematic basis in the schools at the elementary level. Social studies, science, and work-study skills are measured regularly by the majority of schools, but not as extensively as the first four named areas.
6. A wide variety of patterns pertaining to grade level of usage of standardized achievement measures is noted. The greatest level of concentration is at grades 6, 8, 4, and 3, in the order named.

7. Standardized achievement testing is not carried on as extensively at the secondary level as at the elementary level. Reading receives the greatest emphasis, followed by mathematics, language usage, science, and social studies. Limited use is made of standardized measures in the fields of literature, study-skills, modern language, and business education. Standardized measures are rarely used in home economics, industrial arts, music, and art.
8. The end of the year or spring testing predominates in the schools, with the smaller school systems leading in this practice.
9. Non-test techniques find limited use in the schools. Anecdotal records and behavior or personality ratings have the greatest frequency of use. Directed observation, autobiography, problem checklist, case study, and sociometric techniques have very little usage.
10. Very few school systems make use of locally constructed achievement tests of a system-wide basis, only 8 per cent of the schools reporting regular use of these measures.
11. The areas where educators indicate the greatest need for standardized measuring devices are: general educational development at the senior high school, aptitude in music and art, elementary science, social studies, and personality.

Organizing and reporting test results.-- Provisions made for organizing and recording test data and pertinent information about the

individual pupil are integral phases of the testing and evaluation program. Circulating these data to various individuals is considered to be essential to the success of the program. General findings pertaining to this aspect are noted below.

1. The majority of schools regularly circulate the results of intelligence and standardized achievement tests to each teacher or staff member concerned with the class group or subject tested.
2. Although the majority of schools make test results available to school functionaries, only a limited number of schools make such results available to pupils and parents.
3. The form used most frequently for recording test results is the class record sheet giving total scores and the individual pupil profile card.
4. Test data for individual pupils are recorded on a cumulative basis in the individual pupil's permanent cumulative record.
5. The most widely used derived scores are: grade equivalent, percentile rank, age equivalent, and standard score. The first named score has by far the widest use.
6. National norms furnished by the test publishers are primarily used by the schools. Local and regional norms are used on a limited basis, chiefly by the larger school systems.

Using test results.-- Major findings pertaining to utilization of test results are:

1. Many and varied uses are made of test results.
2. Major uses of intelligence and standardized achievement tests are: pupil identification, grouping and placement, curriculum planning and evaluation, and evaluation and improvement of instruction.
3. Results of special aptitude tests, interest inventories, and personality measures find primary usage for individual pupil guidance purposes.
4. Test results are being used, to a limited extent, for the following administrative and supervisory purposes: (1) an aid for expanding educational services and affecting major improvements in the instructional program; (2) an aid for initiating program expansion and instituting enrichment activities; (3) an aid for initiating curriculum revision; (4) an aid for instituting changes in methodology and instructional procedures; and (5) a public relations media--an aid for informing the public of basic needs of the local school system.

2. Recommendations

Several recommendations suggest themselves to the writer:

1. The expenditure for pupil appraisal programs by the New England public schools appears to be exceedingly low when reviewed in terms of the total per pupil educational costs. Schools must exert a greater effort in providing funds for this program if it is to meet the needs of a modern education.

2. Expanded budgetary provisions are needed for research, testing materials, specialized personnel (including psychological services) essential in a comprehensive evaluation program.
3. Adequate time and personnel should be provided to give the necessary direction to the program. In the large system a specially trained person should be charged with the responsibility of the total program. In the small system a teacher should be allowed some time to function as the program specialist. The schools must strive to attain these aims if the program is to attain maximum effectiveness.
4. Considerable upgrading in training of the professional staff is required if the appraisal program is to attain its rightful professional status.
5. Added consultative services and increased personnel are needed to operate an expanded testing and evaluation program required in today's schools. The small schools may secure this by banding together and sharing specialists on a regional basis. The State Department of Education could well consider increasing its consultative or advisory services to schools to meet this need.
6. Professional personnel still are required to score tests in the majority of schools. In the interest of efficiency and economy greater use should be made of clerks, outside agencies, and machine scoring.

7. Since the classroom teacher plays an important part in the testing program, schools should direct added emphasis to the education of the teacher in test administration, interpretation, and utilization, as well as in newer evaluation techniques. This implies an expanded in-service education for teachers and other staff members.
8. The New England public schools as a group seem to be meeting minimal standards in the measurement of basic skills and informational areas in the elementary grades, with reading and arithmetic receiving the greatest coverage. They appear to fall far short of a minimal program in the secondary level, both in the broad field areas and in separate subject measurement. Increased test coverage is evidently required at the secondary level.
9. Although many schools are providing excellent testing coverage at critical points in the pupil's school career, the New England public schools as a group could increase the effectiveness of their testing program by concentrating more testing at transitional periods in the pupil's career. This implies grades 3 or 4, grades 6 or 7, grades 9 or 10, and grade 12.
10. The traditional test approach dominates the New England scene, with emphasis upon the three R's and subject content areas. Newer techniques for the most part are neglected. Additional emphasis on measurement of the functional aspects of learning is strongly advocated.

11. The greatest need for improvement in the present programs might be found in the area of personal-social development. Very little is being done to evaluate social attitude, emotional adjustment, social adjustment, and related needs of the learner. Increased use of informal non-test devices, creative attempts to secure new devices, and research on improved methods of measurement remain a crying need of the schools. More attention and effort should be directed toward these needs.
12. Data imply that achievement tests are used more extensively for administrative and survey purposes than for instructional and classroom purposes. Programs might attain greater effectiveness if the reverse were true.
13. The New England public schools, as a whole, make little use of the locally constructed achievement test on a system-wide basis. This is an area requiring attention and added emphasis.
14. The schools, as a whole, do an adequate job in circulating and making test results available to school functionaries, but fall far short in making results available to parents and pupils. It is evident that schools need to present these data to pupils and parents with proper explanation and interpretation. This dictates a pupil and parent education program in test interpretation. This would seem to be a definite need requiring attention.
15. Additional use of stanines as derived scores and more extensive use of the local and regional norm as a means of making scores

more meaningful and comparable are advocated.

16. The primary use of test results is more for purposes of guidance and aiding the maximum school adjustment of the pupil. Greater utilization would be achieved if results were used to a greater extent to improve the more broad aspects of the educational program. Greater use for the following purposes is thus advocated: curriculum evaluation and improvement, evaluating and improving instruction, educational research, and interpreting the schools to the public.

3. Possibilities for Further Research

It is apparent that a real need exists for further investigation of many aspects of the pupil-centered testing and evaluation program. A few suggested studies for profitable research follow.

1. A follow-up study of the present investigation could be made to determine the impact of the National Defense Education Act of 1958 on the testing and evaluation programs in New England public schools.
2. The present study could be expanded to include a sampling of schools throughout the country in order to determine the present testing and evaluation practices in the public (or private) schools in the United States or in a selected geographical area.
3. A more detailed analysis could be made of many aspects of the testing and evaluation program investigated in this study.

4. A follow-up study of measurement practices of classroom teachers in New England public schools would prove to be a profitable supplement to the present study.
5. A study to determine the criteria used to select tests could be made. This study would involve an analysis of the purposes behind the use of a particular test. It could seek to determine if there is any relationship between the school's educational objectives and the purposes underlying test selection.
6. A detailed analysis could be made to determine the specific aspects of achievement being emphasized in the school's standardized achievement testing program. This would involve an analysis of the specific measures used in the school-wide program.

4. Limitations of the Study

This study is limited to those major items included in the inquiry form. It does not include every possible testing and evaluation practice. The study shares the limitations of all such surveys using an inquiry form in that it depends for its validity on the truthfulness of the person reporting. Except for the six systems which were checked in person by the author, it was impossible to judge the accuracy of the replies. The reliability of the replies is limited by variation in individual interpretation.

APPENDIX A

PRELIMINARY INQUIRY FORM

TESTING AND EVALUATION PRACTICES

IN

NEW ENGLAND PUBLIC SCHOOLS

Tentative Inquiry Form (Working Copy)

Winston B. Keck

General Information

Note: The information requested in this Inquiry Form applies to the school year 1955 - 1956.

1. Name of town (city)_____ Date_____

2. Population of town (city)_____

3. Total number of pupils enrolled in school system as of October 1st_____

4. Under what plan is the school system in the town (city) organized? Check (v) one.

a( ) 8 - 4
b( ) 6 - 3 - 3
c( ) 6 - 2 - 4
d( ) 6 - 6
e( ) Other (Specify) _____

5. If there is no high school in the town, where do the majority of students attend high school?

_____ _____
name of school town (city)

6. Is a definite provision made in the school budget for the pupil appraisal program? Check (v) one.

Note: The pupil appraisal program is intended to include the school program for measuring pupil intelligence, achievement, special aptitudes, interest, and personality traits.

a( ) Yes
b( ) No

If "Yes" in item 6 is checked, answer the next item.

7. How is this amount allocated in the school budget? Check (v)

a( ) Allocated separately to elementary and secondary schools
b( ) Allocated separately to individual school buildings in the system
c( ) Included in the budget in one lump sum
d( ) Other (Specify) _____

8. Approximately what amount is spent each year for the pupil appraisal program?

Note: This amount should include the purchase of tests, materials, scoring services, and other items exclusive of salaries of personnel.

$_____

9. Name of person filling out this Inquiry Form_____

Official position_____

Part I. HOW IS THE PROGRAM ORGANIZED AND ADMINISTERED AT THE LOCAL LEVEL?

A. Responsibility for Administering the Program

1. Who is charged with the responsibility of planning, organizing and supervising the system-wide pupil appraisal program in the school system? Check (v) one or more.

   a( ) Superintendent of Schools
   b( ) Assistant Superintendent of Schools
   c( ) Committee on testing and evaluation
   d( ) School Principal
   e( ) Director of Testing
   f( ) Director of Guidance
   g( ) School Psychologist
   h( ) Other (Specify) _____

2. If a committee on testing and evaluation exists in the school system, who are members of this committee? Check (v)

   a( ) Superintendent of Schools
   b( ) Assistant Superintendent of Schools
   c( ) Principal
   d( ) Supervisor
   e( ) Director of Testing
   f( ) Director of Guidance
   g( ) School Psychologist
   h( ) Guidance Counselor
   i( ) Department Head
   j( ) Classroom Teacher
   k( ) Parent
   l( ) Pupil
   m( ) Others (Specify) _____

3. What is the title of the person in charge of the system-wide testing program? (Specify) _____

   a. Has this person special training in tests and measurement which qualify him for this position? Check (v) one.

      a( ) Yes
      b( ) No

   b. Approximately what proportion of this person's time is spent in testing and evaluation activities?

      (Specify) _____%

4. Who is charged with the responsibility of scheduling the system-wide testing program? Check (v).

   a( ) Classroom teacher
   b( ) Supervisor
   c( ) Principal
   d( ) Director of Testing
   e( ) Committee on testing and evaluation
   f( ) Assistant Superintendent of Schools
   g( ) Superintendent of Schools
   h( ) Other (Specify) _____

Indicate by checking (v) in the appropriate columns who is responsible for choosing the various types of measures used in the testing and evaluation program.

Person	Type of Measures					
	I Group Intell. Tests	II Individ. Intell. Tests	III Stand. Achieve. Tests	IV Special Aptit. Tests	V Person- ality Tests	VI Interest Inventor- ies
assroom teacher						
partment head						
pervisor						
rector of Guidance						
idance Counselor						
sting Specialist						
mmittee on Testing						
acher Committee						
incipal						
ist. Supt. of Sch						
pt. of Schools						
her (Specify)						

Indicate by checking (v) in the appropriate column who usually administers the various types of measures used in the testing and evaluation program.

School Personnel	Type of Measures					
	I Group Intell. Tests	II Individ. Intell. Tests	III Stand. Achieve. Tests	IV Special Aptit. Tests	V Person- ality Tests	VI Interest Inventor- ies
assroom teacher						
partment head						
pervisor						
idance specialist						
hool psychologist						
sting specialist						
incipal						
her (Specify)						

Indicate by checking (v) in the appropriate columns the procedure that is usually followed in scoring tests in the elementary and secondary schools.

Type of Measure	Scoring Procedure					
	Elementary Schools			Secondary Schools		
	I Local Hand Scoring	II Local Machine Scoring	III Outside Agency or Person	IV Local Hand Scoring	V Local Machine Scoring	VI Outside Agency of Person
oup Intelligence ests						
andardized chievement Tests						
ecial Aptitude ests						
rsonality ests						
terest nventories						

8. If intelligence and standardized achievement tests are scored locally by hand, indicate by checking (v) in the appropriate columns who generally scores the tests.

School Personnel	I Intelligence Tests	II Standardized Achievement Tests
a. Classroom teacher		
b. Committee of teachers		
c. Guidance specialist		
d. Principal		
e. Clerk		
f. Other (Specify)		

B. Using Outside Services

1. To what extent does the school system use the services of outside testing specialists on a consultant basis? Check (v)

   a( ) Regularly
   b( ) Occasionally
   c( ) Never

2. Would the school system use the services of a consultant if one could be made available? Check (v) one.

   a( ) Yes
   b( ) No
   c( ) Comment_____

3. Does the school system make use of the services of an outside testing agency or service?

   a( ) Yes
   b( ) No

   If "Yes" in Item 3 is checked, answer the following two items

   a. Who directs or sponsors this service? Check (v)

      1( ) College or University
      2( ) State Department of Education
      3( ) Commercial Service Agency
      4( ) Other (Specify)_____

   b. What testing services are provided by the agency or servic Check (v).

      1( ) Complete "packaged" testing program
      2( ) Tests provided on rental basis
      3( ) Test scoring
      4( ) Consultant and field service
      5( ) Others (Specify)_____

      _____

4. Does the school system participate in a state-wide or regional testing program? Check (v) one.

   a( ) Yes
   b( ) No

C. Personnel Orientation

1. Is there an organized in-service training program to assist teachers in interpreting and using tests and other evaluation devices in carrying out their regular instructional functions? Check (v) one.

a( ) Yes
b( ) No

If "Yes" in item 1 is checked, answer the following three items.

a. Who is responsible for directing and coordinating the program? Check (v)

1( ) Superintendent of Schools
2( ) Assistant Superintendent of Schools
3( ) Principal
4( ) Supervisor
5( ) Director of Testing
6( ) Director of Guidance
7( ) Other (Specify)_____

b. What means are used in the in-service training program? Check (v)

1( ) Extension courses offered by a University or College
2( ) Group conference and discussions
3( ) Grade level or subject meetings
4( ) Workshop in measurements
5( ) General teachers' meeting devoted to measurements and testing problems
6( ) Others (Specify)_____

c. Are teachers given released time for this in-service work? Check (v).

1( ) Yes
2( ) No

2. Are mimeographed teacher directions, supplementing those of the test publisher, prepared for classroom teachers administering standardized tests? Check (v) one.

a( ) Regularly
b( ) Occasionally
c( ) Never

D. Status of Program

1. Check (v) the statement below which best describes the status of your testing and evaluation program during the past five years.

a( ) Program has steadily expanded
b( ) Program has remained essentially the same
c( ) Program has substantially diminished in scope

2. Has the budgetary allotment for testing and evaluation increased any during the past five years in proportion to the total budgetary allotment for instructional services? Check (v) one.

a( ) No appreciable increase
b( ) Slight increase
c( ) Significant increase

3. If the program has undergone change or revision during the pas five years, indicate the chief reason for the change. Check (

a( ) The change came about because of new needs
b( ) The change came about because of dissatisfaction with t previous program
c( ) Other (Specify)_____

4. What do you consider to be the major problem or weakness in yo present testing and evaluation program? Check (v)

a( ) Lack of trained personnel in testing and measurements
b( ) Teachers not properly trained in use of test results
c( ) Only limited use being made of test results
d( ) Not proper follow-up of testing program
e( ) Program too limited in scope
f( ) Other (Specify)_____

Part II WHAT TESTS AND OTHER EVALUATION DEVICES ARE USED IN THE SYSTEM

A. Standardized Testing Program

1. Measuring Intelligence, Special Aptitude, Interest, and Personality

a. Indicate by checking (v) in the appropriate columns the extent to which various measures are used in the school system.

Measures	I Used on systematic basis at regular intervals thruout school system	II Used on one-time or spasmodic basis	III Used only with individuals or class groups as need arises	IV Not us
1. Group Intell. Tests				
2. Individual Intell. Tests				
3. Special Aptitude Tests				
4. Interest Inventories				
5. Personality Tests				

b. Indicate the grade levels at which the following measures are generally administered by encircling the appropriate numbers to the right.

. Group Intelligence Tests.....K 1 2 3 4 5 6 7 8 9 10 11 12

. Individual Intell. Tests.....K 1 2 3 4 5 6 7 8 9 10 11 12

. Special Aptitude Tests.....K 1 2 3 4 5 6 7 8 9 10 11 12

. Interest Inventories.....K 1 2 3 4 5 6 7 8 9 10 11 12

. Personality Tests.....K 1 2 3 4 5 6 7 8 9 10 11 12

2. Measuring Achievement

a. Indicate by checking (v) in the appropriate columns the extent to which different skills and informational areas are measured by standardized achievement tests in grades K to VIII inclusive.

	Practice			
	I	II	III	IV
Skill or ormational Area	Systematic basis at regular intervals thru-out system	One time or spasmodic basis	To individuals or class groups as need arises	Not measured
Arithmetic				
Reading				
Lang. usage				
Spelling				
Science				
Social Stud.				
Work-study skills				

b. Indicate the grade levels at which the following skills or informational areas are generally measured by encircling the appropriate numbers to the right.

1. Arithmetic.....K 1 2 3 4 5 6 7 8

2. Reading... .....K 1 2 3 4 5 6 7 8

3. Language Usage.....K 1 2 3 4 5 6 7 8

4. Spelling.....K 1 2 3 4 5 6 7 8

5. Science.....K 1 2 3 4 5 6 7 8

6. Social Studies.....K 1 2 3 4 5 6 7 8

7. Work-study Skills.....K 1 2 3 4 5 6 7 8

c. Indicate by checking (v) in the appropriate columns the extent to which different skills and subject areas are measured by standardized achievement tests in grades IX to XII inclusive.

Skill or Informational Area	Practice I Systematic basis at regular intervals thruout system	II One time or spasmodic basis	III To Individuals or class groups as need arises	IV Not meas
1. Reading				
2. Lang. Usage				
3. Literature				
4. Mathematics				
5. Science				
6. Soc. Studies				
7. Study Skills				
8. Bus. Educat.				
9. Modern Lang.				
10. Home Economics				
11. Indust. Arts				
12. Music				
13. Art				

d. If standardized achievement tests are administered in the school system, indicate by checking (v) in the appropriat columns the time of year at which they are generally administered at the various school levels.

Time of Year	I Elementary	II Jr. H. S.	III Sr. H. S
1. Beginning of the year or term			
2. Middle of the year or term			
3. End of the year or term			
4. At the completion of a topic or unit			
5. No definite policy			
6. Other (Specify)			

B. Non-Test Techniques

Indicate by (v) checking in the appropriate columns the extent to which non-test techniques are used in the school system.

	Practice I Systematic basis at regular intervals thruout the system	II With individuals or groups as need arises	III Measures used
1. Anecdotal Records			
2. Autobiography			
3. Behavior or personality ratings			
4. Case studies			
5. Directed Observation			
6. Sociometric Techniques			
7. Others (Specify)			

C. Locally-Made Tests

To what extent are achievement tests constructed locally for use throughout the system? Check (v) one.

1( ) Regularly
2( ) Occasionally
3( ) Never

D. In what areas, not adequately covered at the present time by available standardized tests, are tests and measuring devices most needed in the school system?

Specify)_____

_____

t III HOW ARE THE RESULTS OF THE PROGRAM ORGANIZED AND REPORTED?

A. Recording and Circularizing Test Results for Staff and Teacher Use.

1. Indicate by checking (v) in the appropriate columns the statement which most nearly describes the extent to which the results of intelligence and standardized achievement tests are made available to teachers and staff members.

Practice	I Intelligence Tests	II Standardized Achievement Tests
ults are regularly circulated to each teacher staff member concerned with the class group or ject tested		
ults are occasionally circulated to each teacher staff member concerned with the class group subject tested		
ults are non-circulating, but are available teachers or staff members upon request		
ults are kept solely for administrative or dance personnel use		

2. Indicate by checking (v) in the appropriate columns the person to whom the results of the various types of tests are made available.

Person	Type of Test				
	I Intell. Tests	II Standard Achieve. Tests	III Special Aptit. Tests	IV Personality Tests	V Interest Inventories
ncipal					
ervisor					
dance Counselor					
ool Psychologist					
ssroom Teacher					
ent					
il					

3. Indicate the form in which intelligence and standardized achievement test results are made available to teachers and staff members. Check (v)

a( ) Class testing record sheet giving total scores of pupil
b( ) Class analysis sheet giving part or subject as well as total scores of pupils
c( ) Class or group profile charts
d( ) Individual pupil profile chart
e( ) Other (Specify)_____

4. In terms of what derived scores are the standardized achieveme tests recorded in the school records? Check (v)

a( ) Standard scores
b( ) Grade equivalents
c( ) Age equivalents
d( ) Percentile rank
e( ) Stanines
f( ) Other (Specify)_____

5. What norms are generally used in interpreting standardized achievement test data? Check (v)

a( ) Norms furnished by the test publisher
b( ) Norms computed on a regional basis
c( ) Norms computed on the local level for the school system
d( ) Other (Specify)_____

6. How are the test results of individual pupils recorded? Check

a( ) Recorded on an individual pupil test record sheet or card covering all the tests administered at a given grade level.
b( ) Recorded on an individual pupil profile card
c( ) Recorded on the permanent individual pupil cumulative record card.
d( ) Others (Specify)_____

B. Pupil Personnel Record

1. Is an individual pupil cumulative record card kept for each pupil in the school system? Check (v) one.

a( ) Yes
b( ) No

If "Yes" in Item 1 is checked answer the following three items.

2. Check (v) the statement which most nearly describes the pupil cumulative record used in the school system.

a( ) The pupil cumulative record is started when the pupil first enters the school system.
b( ) The pupil cumulative record is continuous from grade to grade.
c( ) The pupil cumulative record follows the pupil from grade to grade and from school to school within the system.
d( ) Other (Specify)_____

3. Indicate by checking (v) in the appropriate columns where the pupil cumulative record is filed at the various school levels.

Place Filed	I Elementary School	II Junior H. S.	III Senior H. S.
Central administrative office of the school system			
Individual school office			
Guidance Office			
Homeroom			
Other (Specify)			

4. If test results are recorded on the individual pupil cumulative record card, indicate by checking (v) in the appropriate columns the person who is responsible for recording and keeping these results up-to-date at the various school levels.

Person	I Elementary School	II Junior H. S.	III Senior H. S.
Classroom teacher			
Homeroom teacher			
Guidance Counselor			
Principal			
Clerk			
Other (Specify)			

rt IV WHAT USE IS MADE OF THE RESULTS OF THE TESTING AND EVALUATION PROGRAM?

Directions: Indicate the main administrative and supervisory uses that are being made of the results of the measures administered in the school system. Check (v) <u>once</u> if use is only occasional or spasmodic. Check (v) (v) <u>twice</u> if use is regular or common practice.

A. Intelligence Tests

1. ( ) ( ) To determine how different groups compare in ability levels from school to school and within schools
2. ( ) ( ) To aid in the placement of pupils in grades
3. ( ) ( ) To aid in grouping within grades
4. ( ) ( ) To aid in determining the expectancy level in achievement within the class group and schools
5. ( ) ( ) To assist in planning special programs to meet the needs of individual pupils and class groups
6. ( ) ( ) To identify those pupils achieving significantly below their capacity to achieve
7. ( ) ( ) Others (Specify other major uses of intelligence tests in the system)_____

_____

_____

B. Standardized Achievement Tests

1. ( ) ( ) To reveal the extent to which instructional objectives are being attained
2. ( ) ( ) To identify instructional areas needing special emphas and attack
3. ( ) ( ) To evaluate different teaching methods and materials c instruction
4. ( ) ( ) To rate the proficiency of different teachers
5. ( ) ( ) To determine how the achievement of local school group compare with those of comparable groups in other syste
6. ( ) ( ) To compare the achievement of different class groups within the system
7. ( ) ( ) To aid in the placement of pupils in grades
8. ( ) ( ) To aid in grouping within the grade or class group
9. ( ) ( ) To aid in curriculum revision and evaluation
10. ( ) ( ) To interpret the work of the school to the public
11. ( ) ( ) To aid in counseling of individuals with respect to ed tional and vocational planning
12. ( ) ( ) Others (Specify other major uses that are made of standardized achievement tests)_____

_____

_____

_____

C. Special Aptitude Tests

1. ( ) ( ) To determine the need for curriculum revision and special school programs
2. ( ) ( ) To aid in the placement of pupils in grades and for grouping within the grade or class
3. ( ) ( ) Others (Specify other major uses that are made of special aptitude tests)_____

_____

_____

_____

D. Personality Measures

1. ( ) ( ) To determine the personal and mental health problems which characterize class groups, schools, or school districts
2. ( ) ( ) To determine the need for special attention to certain areas of the curriculum
3. ( ) ( ) To determine class groups or schools that need special assistance in the social school situation
4. ( ) ( ) Others (Specify other major uses that are made of personality measures)_____

_____

_____

_____

E. Interest Tests or Inventories

1. ( ) ( ) To assist in curriculum planning in meeting the needs of class and school groups
2. ( ) ( ) To determine if the curriculum offerings are meeting the needs and interests of class and school groups
3. ( ) ( ) To aid in the placement of pupils in grades and for grouping within the grade or class
4. ( ) ( ) To aid in educational and vocational planning
5. ( ) ( ) Others (Specify other major uses made are made of Interest Tests or Inventories)_____

_____

_____

_____

F. Comments: It would be appreciated if you could cite specifically anything that has been done in the last two years by way of use of the results of the testing program. For example, "We reivsed our 10th grade mathematics program last year as the result of a 10th grade survey testing program in mathematics."

_____

_____

_____

_____

_____

_____

APPENDIX B

REVISED INQUIRY FORM

TESTING AND EVALUATION PRACTICES

IN

NEW ENGLAND PUBLIC SCHOOLS

By

Winston B. Keck

Curriculum Coordinator
New Britain Public Schools

-1956-

GENERAL INFORMATION

Note: The information requested in this Inquiry Form applies to the school year 1955-1956

A. Name of town(city)_____ State_____

B. Name of Superintendent of Schools_____

C. Name of person completing this form_____

D. Title of person completing this form_____

E. Number of pupils enrolled in the school system_____

F. Under what plan is the school system in the town(city) organized? Check(v) one.

( ) 1. 8-4
( ) 2. 6-3-3
( ) 3. 6-2-4
( ) 4. 6-6
( ) 5. Other(Specify)_____

_____

G. Is a definite provision made in the school budget for the pupil appraisal program? Encircle: Yes No

Note: The "pupil appraisal program" is intended to include the school program for measuring pupil intelligence, achievement, special aptitudes, interests, and personality traits.

If "Yes" in item G. is encircled, answer the next item.

H. How is this amount allocated in the school budget? Check(v) one.

( ) 1. Allocated separately to elementary and secondary schools
( ) 2. Allocated separately to individual school buildings in the system
( ) 3. Included in the budget in one lump sum
( ) 4. Other(Specify)_____

_____

I. Approximately what amount is spent each year for the pupil appraisal program?

Note: This amount should include the purchase of tests, materials, scoring services, and other items exclusive of salaries of personnel.

$_____

HOW IS THE PROGRAM ORGANIZED AND ADMINISTERED AT THE LOCAL LEVEL?

A. Responsibility for Administering the Program

1. Who is given the responsibility of planning, organizing, and supervising the system-wide pupil appraisal program in the school system? Check(v) one or more.

( ) a. Superintendent of Schools
( ) b. Assistant Superintendent of Schools
( ) c. Committee on testing and evaluation
( ) d. School Principal
( ) e. Director of Testing
( ) f. Director of Guidance
( ) g. School Psychologist
( ) h. Other(Specify)_____

_____

2. What is the title of the person who is charged with coordinating and administering the system-wide testing program?

(Specify)_____

a. Does this person hold an advanced degree with specialization in the field of tests and measurements? Encircle: Yes No

b. Approximately what proportion of this person's time is spent in testing and evaluation activities?

(Specify)_____per cent

3. If a committee on testing and evaluation exists in the school system, who are members of this committee? Check(v) one or more.

( ) a. Superintendent of Schools
( ) b. Assistant Superintendent of Schools
( ) c. School Principal
( ) d. Director of Testing
( ) e. Director of Guidance
( ) f. School Psychologist
( ) g. Guidance Counselor
( ) h. Department Head
( ) i. Classroom Teacher
( ) j. Parent
( ) k. Pupil
( ) l. Other(Specify)_____

_____

4. Who is charged with the responsibility of scheduling the system-wide testing program? Check(v) one or more.

( ) a. Superintendent of Schools
( ) b. Assistant Superintendent of Schools
( ) c. School Principal
( ) d. Committee on Testing and Evaluation
( ) e. Director of Testing
( ) f. Director of Guidance
( ) g. Classroom Teacher
( ) h. Other(Specify)_____

4. Indicate who is given the responsibility for choosing the various types of measures used in the testing program by checking(v) the appropriate columns.

Person	Type of Measures					
	I Group Intell-- Tests	II Individ-- Intell-- Tests	III Standard- Achieve-- Tests	IV Special Aptitude Tests	V Personal-- ity Tests	VI Interest Inventor-- ies
a. Superintendent of Schools						
b. Assistant Superintendent of Schools						
c. Testing Committee						
d. School Principal						
e. Elementary Supervisor						
f. Director of Testing						
g. Guidance Director						
h. Guidance Counselor						
i. Department Head						
j. Teacher Committee						
k. Classroom Teacher						
l. Other(Specify)						

5. Indicate who generally administers the various types of measures used in the testing program by checking(v) the appropriate columns.

School Personnel	Type of Measures					
	I Group Intell-- Tests	II Individ-- Intell-- Tests	III Standard- Achieve-- Tests	IV Special Aptitude Tests	V Person- ality Tests	VI Interest Inventories
a. School Principal						
b. Elementary Supervisor						
c. Director of Testing						
d. School Psychologist						
e. Guidance Specialist						
f. Department Head						
g. Classroom Teacher						
h. Other(Specify)						

. Indicate the procedure that is usually followed in scoring tests in the elementary and secondary schools by checking(v) the appropriate columns.

Type of Measure	Scoring procedure					
	Elementary Schools			Secondary Schools		
	I Local Hand Scoring	II Local Machine Scoring	III Outside Agency or Person	IV Local Hand Scoring	V Local Machine Scoring	VI Outside Agency or Person
a. Group Intelligence Tests						
b. Standardized Achievement Tests						
c. Special Aptitude Tests						
d. Personality Tests						
e. Interest Inventories						

. If intelligence and standardized achievement tests are scored locally by hand, indicate who generally scores the tests by checking(v) the appropriate columns.

	I Intelligence Tests	II Standardized Achievement Tests
a. Classroom Teacher		
b. Committee of Teachers		
c. Guidance Specialist		
d. Principal		
e. Clerk		
f. Other(Specify)		

Outside Services

. To what extent does the school system use the services of outside testing specialists on a consultant basis? Check(v) one.

( ) a. Regularly
( ) b. Occasionally
( ) c. Never

. Would the school system use the services of a consultant if one could be made available? Encircle: Yes No

Comment: _____

_____

Does the school system make use of the services of an outside testing agency or service? Encircle: Yes No

If "Yes" in item 3 is encircled, kindly answer the following two items.

a. Who directs or sponsors this service? Check(v) one or more.

( ) 1. College or University
( ) 2. State Department of Education
( ) 3. Commercial Service Agency
( ) 4. Other(Specify)_____

_____

b. What testing services are provided by the agency or service? Check(v) one or more.

( ) 1. Complete "packaged" testing program
( ) 2. Tests provided on rental basis
( ) 3. Test scoring
( ) 4. Consultant and field service
( ) 5. Other(Specify)_____

_____

. To what extent does the school system participate in a state-wide or regional testing program? Check(v) one.

( ) a. Regularly
( ) b. Occasionally
( ) c. Never

Personnel Orientation

. Is there an organized in-service training program to assist teachers in interpreting and using tests and other evaluation devices in carrying out their regular instructional functions? Encircle: Yes No

If "Yes" is encircled in item 1, kindly answer the following three items.

a. Who is given the/responsibility for directing and coordinating the program? Check(v) one.

( ) 1. Superintendent of Schools
( ) 2. Assistant Superintendent of Schools
( ) 3. School Principal
( ) 4. Elementary Supervisor
( ) 5. Director of Testing
( ) 6. Director of Guidance
( ) 7. Other(Specify)_____

_____

b. What means are used in the in-service training program? Check(v) one or more.

( ) 1. Extension courses offered by a University or College
( ) 2. Group conferences and discussions
( ) 3. Grade level or subject meetings
( ) 4. Workshop in measurements
( ) 5. General teachers' meeting devoted to measurements and testing problems
( ) 6. Others(Specify)_____

_____

c. Are teachers given released time for this in-service work? Check(v) one.

( ) 1. Regularly
( ) 2. Occasionally
( ) 3. Never

. Are mimeographed teacher directions, supplementing those of the test publisher, prepared for classroom teachers administering standardized tests? Check(v) one.

( ) a. Prepared regularly for all standardized testing
( ) b. Prepared only for certain types of standardized tests
( ) c. Not prepared for teachers use

Status of the Program

. Check(v) the statement below which best describes the status of your testing and evaluation program during the past five years.

( ) a. The program has steadily expanded
( ) b. The program has remained essentially the same
( ) c. The program has substantially diminished in scope

. Has the budgetary expenditure for testing and evaluation increased any during the past five years in proportion to the total budget expenditure for instructional services? Check(v) one.

( ) a. No appreciable increase
( ) b. Slight increase
( ) c. Significant increase

. If the program has undergone change or revision during the past five years, kindly indicate the chief reason for the change. Check(v) one or more.

( ) a. The change came about because of the recognition of new needs
( ) b. The change came about because of dissatisfaction with the previous program
( ) c. Other(Specify)_____

_____

. What do you consider to be the major problem or weakness in your present testing and evaluation program? Check(v) one.

( ) a. Lack of trained personnel in testing and measurements
( ) b. Teachers not properly trained in the use of test results
( ) c. Only limited use being made of test results
( ) d. Not proper follow-up of testing program
( ) e. Program too limited in scope
( ) f. Insufficient time and resources for full utilization
( ) g. Other(Specify)_____

. WHAT TESTS AND OTHER EVALUATION DEVICES ARE USED IN OUR SYSTEM

. Standardized Testing Program

1. Measuring Intelligence, Special Aptitudes, Interests, and Personality

a. Indicate the extent to which various measures are used in the school system by checking (✓) the appropriate columns

Measured	I Used on systematic basis at regular intervals throughout the school system	II Used on one-time or sporadic basis	III Used only with individuals or class groups as need arises	IV Not used
1. Group Intelligence Tests				
2. Individual Intelligence Tests				
3. Special Aptitude Tests				
4. Interest Inventories				
5. Personality Tests				

b. Indicate the grade levels at which the following measures are generally administered by encircling the appropriate numbers to the right.

1. Group Intelligence Tests.....K 1 2 3 4 5 6 7 8 9 10 11 12

2. Individual Intelligence Tests.....K 1 2 3 4 5 6 7 8 9 10 11 12

3. Special Aptitude Tests.....K 1 2 3 4 5 6 7 8 9 10 11 12

4. Interest Inventories.....K 1 2 3 4 5 6 7 8 9 10 11 12

5. Personality Tests.....K 1 2 3 4 5 6 7 8 9 10 11 12

. Measuring Achievement

a. Indicate the extent to which different skills and informational areas are measured by standardized achievement tests in grades K to VIII inclusive by checking (✓) the appropriate columns.

Skill or Informational Area	Practice			
	I Systematic basis at regular intervals throughout the system	II One time or sporadic basis	III To individuals or class groups as the need arises	IV Not measured
1. Arithmetic				
2. Reading				
3. Language usage				
4. Spelling				
5. Science				
6. Social Studies				
7. Work-study skills				

b. Indicate the grade levels at which the following skills or informational areas are generally measured by encircling the appropriate numbers to the right.

1. Arithmetic.....K 1 2 3 4 5 6 7 8

2. Reading.....K 1 2 3 4 5 6 7 8

3. Language Usage.....K 1 2 3 4 5 6 7 8

4. Spelling.....K 1 2 3 4 5 6 7 8

5. Science.....K 1 2 3 4 5 6 7 8

6. Social Studies.....K 1 2 3 4 5 6 7 8

7. Work-study Skills.....K 1 2 3 4 5 6 7 8

c. Indicate the extent to which different skills and subject areas are measured by standardized achievement tests in grades IX to XII inclusive by checking (✓) the appropriate columns.

Skill or Informational Area	Practice			
	I Systematic basis at regular intervals throughout the system	II One time or sporadic basis	III To individuals or class groups as the need arises	IV Not measured
1. Reading				
2. Language Usage				
3. Literature				
4. Mathematics				
5. Science				
6. Social Studies				
7. Study Skills				
8. Business Education				
9. Modern Language				
10. Home Economics				
11. Industrial Arts				
12. Music				
13. Art				

If standardized achievement tests are administered in the school system, indicate the time of year at which they are generally administered at the various school levels by checking (✓) the appropriate columns.

Time of Year	I Elementary	II Junior High School	III Senior High School
1. Beginning of the year or term			
2. Middle of the year or term			
3. End of the year or term			
4. At the completion of a topic or unit			
5. No definite policy			
6. Other (Specify)			

n-Test Techniques

dicate the extent to which non-test techniques are used in the school system by ecking (✓) the appropriate columns.

Technique	Practices		
	I Systematic basis at regular intervals throughout the system	II With individuals or groups as the need arises	III Measures not used
1. Anecdotal Records			
2. Autobiography			
3. Behavior or Personality Ratings			
4. Case Studies			
5. Directed Observation			
6. Problem Check List			
7. Sociometric Techniques			
8. Other (Specify)			

Locally-Made Tests

To what extent are achievement tests constructed locally for use on a system-wide basis? Check (✓) one.

( ) 1. Regularly
( ) 2. Occasionally
( ) 3. Never

In what areas, not adequately covered at the present time by available standardized tests, are tests and measuring devices most needed in the school system?

(Specify) _____

HOW ARE THE RESULTS OF THE PROGRAM ORGANIZED AND REPORTED?

Recording and Circularizing Test Results for Staff and Teacher Use

. Indicate the statement which most nearly describes the extent to which the results of intelligence and standardized achievement tests are made available to teachers and staff members by checking (✓) the appropriate columns.

Practice	I Intelligence Tests	II Standardized Achievement Tests
a. Results are regularly circulated to each teacher or staff member concerned with the class group or subject tested.		
b. Results are occasionally circulated to each teacher or staff member concerned with the class group or subject tested.		
c. Results are non-circulating, but are available to teachers or staff members upon request.		
d. Results are kept solely for administrative or guidance personnel use.		

. Indicate the person to whom the results of the various types of tests are made available by checking (✓) the appropriate columns.

	Type of Test				
	I Intelligence Tests	II Standardized Achievement Tests	III Special Aptitude	IV Personality Tests	V Interest Inventories
a. Principal					
b. Supervisor					
c. Guidance Counselor					
d. School Psychologist					
e. Classroom Teacher					
f. Parent					
g. Pupil					

. In what form are intelligence and standardized achievement test results usually made available to teachers and staff members? Check (✓) one or more.

( ) a. Class testing record sheet giving total scores of pupils
( ) b. Class analysis sheet giving part or subject as well as total scores of pupils
( ) c. Class or group profile charts
( ) d. Individual pupil profile chart
( ) e. Other (Specify) _____

4. In terms of what derived scores are the standardized achievement tests generally recorded in the school record? Check (√) one or more.

( ) a. Standard scores
( ) b. Grade equivalents
( ) c. Age equivalents
( ) d. Percentile rank
( ) e. Stanines
( ) f. Other (Specify) _____

_____

5. What norms are generally used in interpreting standardized achievement test data? Check (√) one or more.

( ) a. Norms furnished by the test publisher
( ) b. Norms computed on a regional basis
( ) c. Norms computed on the local level for the school system
( ) d. Other (Specify) _____

_____

6. How are the test results of individual pupils recorded? Check (√) one or more.

( ) a. Recorded on an individual pupil test record sheet or card covering all the tests administered at a given grade level.
( ) b. Recorded on an individual pupil profile card
( ) c. Recorded on the pupil cumulative permanent record card
( ) d. Other (Specify)_____

_____

Pupil Personnel Record

1. Is an individual pupil cumulative record card kept for each pupil in the school system? Encircle: Yes No

If "Yes" is encircled in item 1, answer the following item.

2. Check (√) the statements which more nearly describe the pupil cumulative record used in the school system. Check (√) one or more.

( ) a. The pupil cumulative record is started when the pupil first enters the school system.
( ) b. The pupil cumulative record is continuous from grade to grade.
( ) c. The pupil cumulative record follows the pupil from grade to grade within the system.
( ) d. The pupil cumulative record follows the pupil from one school level to the other within the system.
( ) e. Other (Specify)_____

_____

V. WHAT USE IS MADE OF THE RESULTS OF THE TESTING AND EVALUATION PROGRAM?

Directions: Indicate the main administrative and supervisory uses that are being made of the results of the measures administered in the school system. Check(v) once if use is only occasional or sporadic. Check (v) (v) twice if use is regular or common practice. Leave spaces blank if not used.

A. Intelligence Tests

( ) ( ) 1. To determine how different groups compare in ability levels from school to school and within schools
( ) ( ) 2. To aid in the placement of pupils in grades
( ) ( ) 3. To aid in grouping within grades
( ) ( ) 4. To aid in determing the expectancy level in achievement within the class group and schools
( ) ( ) 5. To assist in planning special programs to meet the needs of individual pupils and class groups
( ) ( ) 6. To identify those pupils achieving significantly below their capacity to achieve
( ) ( ) 7. Other(Specify other major uses of intelligence tests in the system)

_____

_____

_____

B. Standardized Achievement Tests

( ) ( ) 1. To reveal the extent to which instructional objectives are being attained
( ) ( ) 2. To identify instructional areas needing special emphasis and attack
( ) ( ) 3. To evaluate different teaching methods and materials of instruction
( ) ( ) 4. To rate the proficiency of different teachers
( ) ( ) 5. To determine how the achievement of local school groups compare with those of comparable groups in other systems
( ) ( ) 6. To compare the achievement of different class groups within the system
( ) ( ) 7. To aid in the placement of pupils in grades
( ) ( ) 8. To aid in grouping within the grade or class group
( ) ( ) 9. To aid in curriculum revision and evaluation
( ) ( ) 10. To interpret the work of the school to the public
( ) ( ) 11. To aid in counseling of individuals with respect to educational and vocational planning
( ) ( ) 12. Other(Specify other major uses that are made of standardized achievement tests) _____

_____

_____

_____

C. Special Aptitude Tests

( ) ( ) 1. To determine the need for curriculum revision and special school programs
( ) ( ) 2. To aid in the placement of pupils in grades and for grouping within the grade or class

( ) ( ) 3. Other(Specify other major uses that are made of special aptitude tests) _____

_____

_____

_____

D. Personality Measures

( ) ( ) 1. To determine the personal and mental health problems which characterize class groups, schools, or school districts
( ) ( ) 2. To determine the need for special attention to certain areas of the curriculum
( ) ( ) 3. To determine class groups or schools that need special assistance in the social school situation
( ) ( ) 4. Other(Specify other major uses that are made of personality measures)

_____

_____

_____

E. Interest Tests or Inventories

( ) ( ) 1. To assist in curriculum planning in meeting the needs of class and school groups
( ) ( ) 2. To determine if the curriculum offerings are meeting the needs and interests of class and school groups
( ) ( ) 3. To aid in the placement of pupils in grades and for grouping within the grade or class
( ) ( ) 4. To aid in educational and vocational planning
( ) ( ) 5. Others(Specify other major uses that are made of Interest Tests or Inventories) _____

_____

_____

_____

F. Comments: It would be appreciated if you could cite specifically anything that has been done in the last two years by way of use of the results of the testing program. For example, "We revised our 10th grade mathematics program last year as the result of a 10th grade survey testing program in mathematics."

_____

_____

_____

_____

_____

_____

_____

_____

_____

APPENDIX C

FINAL INQUIRY FORM

TESTING AND EVALUATION PRACTICES IN NEW ENGLAND PUBLIC SCHOOLS

By
WINSTON B. KECK
Curriculum Coordinator
New Britain Public Schools

— 1958 —

I. HOW IS THE PROGRAM ORGANIZED AND ADMINISTERED?

A. Responsibility for Administering the Program

1. Who is given the responsibility of planning, organizing, and supervising the **system-wide** pupil appraisal program in the school system? Check (√) one or more.

( ) a. Superintendent of Schools
( ) b. Assistant Superintendent of Schools
( ) c. Committee on testing and evaluation
( ) d. School Principal
( ) e. Director of Testing
( ) f. Director of Guidance
( ) g. School Psychologist
( ) h. Director of Pupil Personnel Services
( ) i. Other (Specify)_____

_____

2. What is the title of the person who is charged with coordinating and administering the **system-wide** testing program?

(Specify)_____

a. Does this person hold an advanced degree with specialization in the field of tests and measurements? Encircle: **Yes No**

b. Approximately what proportion of this person's time is spent in testing and evaluation activities?

(Specify)_____per cent

3. If a committee on testing and evaluation exists in the school system, who are members of this committee? Check (√) one or more.

( ) a. Superintendent of Schools
( ) b. Assistant Superintendent of Schools
( ) c. School Principal
( ) d. Director of Testing
( ) e. Director of Guidance
( ) f. School Psychologist
( ) g. Guidance Counselor
( ) h. Department Head
( ) i. Classroom Teacher(s)
( ) j. Parent(s)
( ) k. Pupil(s)
( ) l. Other (Specify)_____

_____

4. Who is charged with the responsibility of scheduling the **system-wide** testing program? Check (√) one or more.

( ) a. Superintendent of Schools
( ) b. Assistant Superintendent of Schools
( ) c. School Principal
( ) d. Committee on Testing and Evaluation
( ) e. Director of Testing
( ) f. Director of Guidance
( ) g. Classroom Teacher
( ) h. Other (Specify)_____

7. Indicate the procedure that is usually followed in **scoring** tests in the elementary and secondary schools by checking (√) the appropriate columns.

	SCORING PROCEDURE					
	ELEMENTARY SCHOOLS			SECONDARY SCHOOLS		
	I LOCAL HAND SCORING	II LOCAL MACHINE SCORING	III OUTSIDE AGENCY OR PERSON	IV LOCAL HAND SCORING	V LOCAL MACHINE SCORING	VI OUTSIDE AGENCY OR PERSON
A. GROUP INTELLIGENCE TESTS						
B. STANDARDIZED ACHIEVEMENT TESTS						
C. SPECIAL APTITUDE TESTS						
D. PERSONALITY TESTS						
E. INTEREST INVENTORIES						

8. If intelligence and standardized achievement tests are scored locally by hand, indicate who generally **scores** the tests by checking (√) the appropriate columns.

	I INTELLIGENCE TESTS	II STANDARDIZED ACHIEVEMENT TESTS
A. CLASSROOM TEACHER		
B. COMMITTEE OF TEACHERS		
C. GUIDANCE SPECIALIST		
D. PRINCIPAL		
E. CLERK		
F. OTHER (SPECIFY)		

B. Outside Services

1. To what extent does the school system use the services of outside testing specialists on a consultant basis? Check (√) one.

   ( ) a. Regularly
   ( ) b. Occasionally
   ( ) c. Never

2. Would the school system use the services of a consultant if one could be made available? Encircle: **Yes** **No**

   Comment: _____

   _____

b. What means are used in the in-service training program? (Check (√) one or more.

( ) 1. Extension courses offered by a University or College
( ) 2. Group conferences and discussions
( ) 3. Grade level or subject meetings
( ) 4. Workshop in measurements
( ) 5. General teachers' meeting devoted to measurements and testing problems
( ) 6. Others (Specify) _____

_____

c. Indicate the extent to which teachers are given released time for this in-service work? Check (√) one.

( ) 1. Regularly
( ) 2. Occasionally
( ) 3. Never

2. Are mimeographed teacher directions, supplementing those of the test publisher, prepared for classroom teachers administering standardized tests? Check (√) one.

( ) a. Prepared regularly for all standardized testing
( ) b. Prepared only for certain types of standardized tests
( ) c. Not prepared for teachers use

D. Status of Program

1. Check (√) the statement below which best describes the status of your testing and evaluation program during the past **five years.**

( ) a. The program has steadily expanded
( ) b. The program has remained esentially the same
( ) c. The program has substantially diminished in scope

2. Has the budgetary expenditure for testing and evaluation increased any during the past five years in proportion to the total budget expenditure for instructional services? Check (√) one.

( ) a. No appreciable increase
( ) b. Slight increase
( ) c. Significant increase

3. If the program has undergone change or revision during the past **five** years, kindly indicate the chief reason for the change. Check (√) one or more.

( ) a. The change came about because of the recognition of new needs
( ) b. The change came about because of dissatisfaction with the previous program
( ) c. Other Specify) _____

_____

4. What do you consider to be the **major** problem or weakness in your present testing and evaluation program? Check (√) one.

( ) a. Lack of trained personnel in testing and measurements
( ) b. Teachers not properly trained in the use of test results
( ) c. Only limited use being made of test results
( ) d. Not proper follow-up of testing program
( ) e. Program too limited in scope
( ) f. Insufficient time and resources for full utilization
( ) g. Other (Specify) _____

_____

b. Indicate the **grade levels** at which the following skills or informational areas are generally measured by standardized achievement tests by encircling the appropriate numbers to the right.

1. Arithmetic ---------- K 1 2 3 4 5 6 7 8
2. Reading ---------- K 1 2 3 4 5 6 7 8
3. Language Usage ---------- K 1 2 3 4 5 6 7 8
4. Spelling ---------- K 1 2 3 4 5 6 7 8
5. Science ---------- K 1 2 3 4 5 6 7 8
6. Social Studies ---------- K 1 2 3 4 5 6 7 8
7. Work-study Skills ---------- K 1 2 3 4 5 6 7 8

c. Indicate the **extent** to which different skills or subject areas are measured by standardized achievement tests in **grades IX to XII** inclusive by checking (V) the appropriate columns.

	PRACTICE			
SKILL OR INFORMATIONAL AREA	I SYSTEMATIC BASIS AT REGULAR INTERVALS THROUGHOUT THE SYSTEM	II ONE TIME OR SPORADIC BASIS	III TO INDIVIDUALS OR CLASS GROUPS AS THE NEED ARISES	IV NOT MEASURED
1. READING				
2. LANGUAGE USAGE				
3. LITERATURE				
4. MATHEMATICS				
5. SCIENCE				
6. SOCIAL STUDIES				
7. STUDY SKILLS				
8. BUSINESS EDUCATION				
9. MODERN LANGUAGE				
10. HOME ECONOMICS				
11. INDUSTRIAL ARTS				
12. MUSIC				
13. ART				

III. HOW ARE THE RESULTS OF THE PROGRAM ORGANIZED AND REPORTED?

A. Indicate the statement which most nearly describes the **extent** to which the results of intelligence and standardized achievement tests are made available to teachers and staff members by checking (√) the appropriate columns.

PRACTICE	I INTELLIGENCE TESTS	II STANDARDIZED ACHIEVEMENT TESTS
A. RESULTS ARE **REGULARLY** CIRCULATED TO EACH TEACHER OR STAFF MEMBER CONCERNED WITH THE CLASS GROUP OR SUBJECT TESTED.		
B. RESULTS ARE **OCCASIONALLY** CIRCULATED TO EACH TEACHER OR STAFF MEMBER CONCERNED WITH THE CLASS GROUP OR SUBJECT TESTED.		
C. RESULTS ARE **NON-CIRCULATING,** BUT ARE AVAILABLE TO TEACHERS OR STAFF MEMBERS UPON REQUEST.		
D. RESULTS ARE KEPT SOLELY FOR ADMINISTRATIVE OR GUIDANCE PERSONNEL USE.		

B. Indicate the person to whom the results of the various types of tests are made available by checking (√) the appropriate columns.

	TYPE OF TEST				
	I INTELLIGENCE TESTS	II STANDARDIZED ACHIEVEMENT TESTS	III SPECIAL APTITUDE	IV PERSONALITY TESTS	V INTEREST INVENTORIES
A. PRINCIPAL					
B. SUPERVISOR					
C. GUIDANCE COUNSELOR					
D. SCHOOL PSYCHOLOGIST					
E. CLASSROOM TEACHER					
F. PARENT					
G. PUPIL					

C. Indicate the form in which intelligence and standardized achievement test results are usually made available to teachers and staff members by checking (√) the appropriate columns.

FORM	I INTELLIGENCE TESTS	II STANDARDIZED ACHIEVEMENT TESTS
A. CLASS RECORD SHEET GIVING TOTAL SCORES OF PUPILS		
B. CLASS RECORD SHEET GIVING PART AND TOTAL SCORES OF PUPILS		
C. CLASS OR GROUP PROFILE CHARTS		
D. INDIVIDUAL PUPIL PROFILE CHART		
E. OTHER (SPECIFY)		

B. Standardized Achievement Tests

( ) ( ) 1. To reveal the extent to which instructional objectives are being attained
( ) ( ) 2. To identify instructional areas needing special emphasis
( ) ( ) 3. To evaluate different teaching methods and materials of instruction
( ) ( ) 4. To rate the proficiency of different teachers
( ) ( ) 5. To determine how the achievement of local school groups compare with those of comparable groups in other systems
( ) ( ) 6. To compare the achievement of different class groups within the system
( ) ( ) 7. To aid in the placement of pupils in grades
( ) ( ) 8. To aid in grouping within the grade or class group
( ) ( ) 9. To aid in curriculum revision and evaluation
( ) ( )10. To interpret the work of the school to the public
( ) ( )11. To aid in the counseling of individuals with respect to educational and vocational planning
( ) ( )12. Other (Specify other major uses that are made of standardized achievement tests) _____

_____

_____

_____

C. Special Aptitude Tests

( ) ( ) 1. To determine the need for special school programs
( ) ( ) 2. To aid in the placement of pupils in grades and for grouping within the grade or class
( ) ( ) 3. To place pupils in remedial or corrective classes
( ) ( ) 4. Other (Specify other major uses that are made of special aptitude tests)

_____

_____

_____

D. Personality Measures (Including problem check lists)

( ) ( ) 1. To determine the personal and mental health problems which characterize class groups, schools, or school districts
( ) ( ) 2. To determine the need for special attention to certain areas of the curriculum
( ) ( ) 3. To determine class groups or schools that need special assistance in the social school situation
( ) ( ) 4. Other (Specify other major uses that are made of personality measures)

_____

_____

_____

APPENDIX D

LETTERS PERTAINING TO INQUIRY FORM DISTRIBUTION

PUBLIC SCHOOLS OF NEW BRITAIN
ADMINISTRATION BUILDING
NEW BRITAIN, CONN.

A comprehensive testing and evaluating program is an integral part of the teaching procedure and must assume a position of major administrative importance. In this year of increased attention to student development in basic instructional areas and current charges of academic shortcomings in public education, additional emphasis must be placed upon the pupil appraisal program in order to secure valid data regarding student development in all areas of learning.

School systems are concerned with maintaining a sound pupil appraisal program based upon accepted practices. Although much has been written regarding these practices, relatively little is known concerning the actual status of the programs. This study is being made to obtain this data for the New England Public Schools. It is being conducted as a doctorate study at Boston University as part of a nation-wide study being sponsored by the American Educational Research Association under the auspice of the Test Usage Committee. It also has the full endorsement of the Executive Committee of the New England School Development Council.

Will you kindly aid in this project by supplying the data requested on the enclosed data form? Although the form looks formidable, only a reasonable amount of time is required to complete it since most of the responses need only check marks.

It is hoped that answers will reflect as nearly as possible the actual situation in the pupil appraisal program in the system. A compilation of this data will contribute materially to a better understanding of what the schools are doing and allow you to compare your program with practices in other schools in the New England area.

The information you provide will be held in strict confidence and will be used only by the writer for analysis and study. A summary of the results will be made available to you when the study has been completed.

It is hoped that all inquiry forms will be returned within two weeks. I am enclosing a self-addressed, stamped envelope for your convenience in replying.

Your cooperation will be deeply appreciated.

Very truly yours,

Winston Keck

Winston B. Keck
Curriculum Coordinator

PUBLIC SCHOOLS OF NEW BRITAIN
ADMINISTRATION BUILDING
NEW BRITAIN, CONN.

A comprehensive testing and evaluating program is an integral part of the teaching procedure and must assume a position of major administrative importance. In this year of increased attention to student development in basic instructional areas and current charges of academic shortcomings in public education, additional emphasis must be placed upon the pupil appraisal program in order to secure valid data regarding student development in all areas of learning.

School systems are concerned with maintaining a sound pupil appraisal program based upon accepted practices. Although much has been written regarding these practices, relatively little is known concerning the actual status of the programs. This study is being made to obtain this data for the New England Public Schools. It is being conducted as a doctorate study at Boston University as part of a nation-wide study being sponsored by the American Educational Research Association under the auspice of the Test Usage Committee. It also has the full endorsement of the Executive Committee of the New England School Development Council.

This study is concerned with the small community as well as the large city. A community has been chosen at random from each Superintendency Union in order to obtain a representative sampling of the smaller towns in New England. The community chosen from your district is noted on the Inquiry Form.

Will you kindly aid in this project by supplying the data requested on the enclosed Inquiry Form? Although the form looks formidable, only a reasonable amount of time is required to complete it, since most of the responses need only check marks.

It is hoped that answers will reflect as nearly as possible the actual situation in the pupil appraisal program in the system. A compilation of this data will contribute materially to a better understanding of what the schools are doing and allow you to compare your program with practices in other schools in the New England area.

The information you provide will be held in strict confidence and will be used only by the writer for analysis and study. A summary of the results will be made available to you when the study has been completed.

It is hoped that all Inquiry Forms will be returned within two weeks. I am enclosing a self-addressed, stamped envelope for your convenience in replying.

Your cooperation will be deeply appreciated.

Very truly yours,

Winston Keck

Winston B. Keck
Curriculum Coordinator

u

PUBLIC SCHOOLS OF NEW BRITAIN
ADMINISTRATION BUILDING
NEW BRITAIN, CONN.

Recently you received an inquiry form on Testing and Evaluation Practices in the New England Public Schools. This form has been sent to communities throughout New England in order to obtain a rather detailed picture of pupil appraisal programs in the area.

This study, endorsed by the Executive Committee of N.E.S.D.E.C. and the Test Usage Committee of A.E.R.A. should contribute materially to a better understanding of what schools are actually doing in their pupil appraisal programs.

Inquiry forms have been received from a great many communities. To date I have not received the form sent to you. Would it be possible for you to take time from your busy schedule to have the form completed? I can assure you that the information will be held in strict confidence by the writer and will be used only in summary form without identifying any specific school system. Administrators throughout the New England area could profit from your contribution to this study.

The return of the inquiry form at your earliest convenience will be greatly appreciated.

Very truly yours,

Winston B. Keck

Winston B. Keck
Curriculum Coordinator

PUBLIC SCHOOLS OF NEW BRITAIN
ADMINISTRATION BUILDING
NEW BRITAIN, CONN.

Several weeks ago you received an inquiry form on Testing and Evaluation Practices in the New England Public Schools. This form, part of a study endorsed by the Executive Committee of N.E.S.D.E.C. and the Test Usage Committee of the A.E.R.A., has been distributed throughout New England. Returns have been received from the majority of school systems.

A check on the inquiry forms returned to date reveals that no reply has been received from you. The study is nearing completion and plans have been made to start tabulating the results on May 29. I hope that your system will be represented in the study. However, it will be impossible to include it unless the form is received by May 29. I hope that a reply will be received from you by this date.

The information that you can provide will be most helpful in obtaining a complete picture of the pupil appraisal programs in the New England Public Schools. All information will be treated as confidential, and no individual school system will be identified in the final report.

If you have misplaced the inquiry form, please let me know, and I will be happy to send you another.

Very truly yours,

Winston B. Keck

Winston B. Keck
Curriculum Coordinator

APPENDIX E

LIST OF EDUCATORS PARTICIPATING IN INQUIRY FORM REVISION AND PILOT STUDY

EDUCATORS PARTICIPATING IN REVISION

AND PILOT STUDY OF INQUIRY FORM

*** 1. Walter N. Durost, Director, Test Service and Advisement Center, Dunbarton, New Hampshire

* 2. Joseph Bedard, Guidance Director, New Britain Public Schools, New Britain, Connecticut

*** 3. Roger T. Lennon, Director, Division of Test Research and Service, World Book Company, Tarrytown-on-Hudson, New York

* 4. George Prescott, Director, Pupil Personnel Services, Norwalk Public Schools, Norwalk, Connecticut

*** 5. T. Joseph McCook, Superintendent of Schools, Springfield, Massachusetts

*** 6. William Edgar, Superintendent of Schools, Fairfield, Connecticut

* 7. Margaret E. Allen, Supervisor of Testing, Portland Public Schools, Portland, Maine

* 8. Mary McGauvran, Dean, State Teachers College, Lowell, Massachusetts

* 9. Harold A. Mahoney, Chief, Bureau of Pupil Personnel and Special Educational Services, Connecticut State Department of Education, Hartford, Connecticut

*** 10. Maurice J. Ross, Chief, Bureau of Research and Statistics, Connecticut State Department of Education, Hartford, Connecticut

* 11. Ruth Kimball, Director, Testing and Research, New Britain Public Schools, New Britain, Connecticut

* 12. Anthony G. Brackett, Superintendent of Schools, Weston, Massachusetts

* 13. Thomas Warren, Superintendent of Schools, Randolph, Massachusetts

* 14. George C. Roy, Superintendent of Schools, Millis, Massachusetts

* 15. Merle A. Sturtevant, Superintendent of Schools (retired), Shrewsbury, Massachusetts

*** 16. Robert A. Stoughton, Guidance Specialist, Connecticut State Department of Education, Hartford, Connecticut

* 17. James Hayden, Assistant Superintendent of Schools, New Bedford, Massachusetts

** 18. John Davis, Executive Secretary, New England School Development Council, Cambridge, Massachusetts

** 19-26. Members of the Executive Committee, New England School Development Council, Cambridge, Massachusetts

** 27. Harold Seashore, Director, Test Division, The Psychological Corporation, New York, New York

** 28. George Forlano, Research Associate, Bureau of Educational Research, Board of Education of the City of New York, New York, New York

** 29. John W. Polley, Executive Secretary, Associated Public School Systems, Teachers College, Columbia University, New York

** 30. Edward J. Rogean, Superintendent of Schools, Bloomfield, Connecticut

** 31. John W. Wallace, Superintendent of Schools, Newington, Connecticut

** 32. E. Perley Eaton, Superintendent of Schools, Berlin, Connecticut

** 33. E. Davis Woodbury, Superintendent of Schools, Milton, Massachusetts

** 34. Leo T. Doherty, Superintendent of Schools, Worcester, Massachusetts

Key:

* Indicates participation in revision study.
** Indicates participation in pilot study.
*** Indicates participation in both revision and pilot studies.

APPENDIX F

SELECTED BIBLIOGRAPHIES USED IN THE STUDY

LIST OF TEXTBOOKS TREATING ASPECTS OF THE TESTING AND EVALUATION PROGRAM

1. Ahman, J. Stanley, and Marvin D. Glock. Evaluating Pupil Growth. Allyn and Bacon, Boston, 1958, pp. 136-182, 468-496.

2. Bradfield, James M., and H. Stewart Moredock. Measurement and Evaluation in Education. The Macmillan Company, New York, 1957, Chap. XVI.

3. Dailey, John G. Testing and Counseling in the High School Guidance Program. Science Research Associates, Chicago, 1943, Chaps. IV and V.

4. Greene, Harry A., Albert N. Jorgensen, and J. Raymond Gerberich. Measurement and Evaluation in the Elementary School (Second Edition). Longmans, Green and Company, New York, 1953, pp. 119-135.

5. Jordan, A. M. Measurement in Education. McGraw-Hill Book Company, Inc., New York, 1953, pp. 67-95.

6. Noll, Victor H. Introduction to Educational Measurement. Houghton Mifflin Company, Boston, 1957, pp. 318-343.

7. Remmers, H. H., H. L. Gage, and J. Francis Rummel. A Practical Introduction to Measurement and Evaluation. Harper and Brothers, New York, 1960, Chaps. IV, V, and VI.

8. Ross, C. C., and Julian C. Stanley. Measurement in Today's Schools (Third Edition). Prentice-Hall, Inc., New York, Chap. VIII.

9. Schwartz, Alfred L., and Stuart C. Tiedman. Evaluating Student Progress in the Secondary School. Longmans, Green and Company, New York, 1957, pp. 17-30.

10. Smith, Eugene R., Ralph W. Tyler, and others. Appraising and Recording Student Progress. Harper and Brothers, New York, 1942, Chap. I, pp. 3-34.

11. Thorndike, Robert L., and Elizabeth Hagen. Measurement and Evaluation in Psychology and Education. John Wiley and Sons, Inc., New York, 1955, Chap. XVI.

12. Togerson, Theodore L., and Georgia S. Adams. Measurement and Evaluation for the Secondary School Teacher. Dryden Press, New York, 1956, Chaps. XXIV, XXV, and XXVI.

13. Traxler, Arthur E. Techniques of Guidance. Harper and Brothers, New York, 1957, Chaps. IX and X.

14. Traxler, Arthur E., Robert Jacobs, Margaret Selover, and Agatha Townsend. Introduction to Testing and the Use of Test Results in Public Schools. Harper and Brothers, New York, 1953, pp. 1-113.

15. Wrightstone, J. Wayne, Joseph Justman, and Irving Robbins. Evaluation in Modern Education. American Book Company, New York, 1956, Chaps. II and IV.

SELECTED BIBLIOGRAPHY

TESTING PATTERNS AND PROGRAMS

1. Ahman, J. Stanley, and Marvin D. Glock. Evaluating Pupil Growth. Allyn and Bacon, Boston, 1958, pp. 483-484.

2. California Test Bureau. Basic Testing Program for Educational Diagnosis and Pupil Guidance, Educational Bulletin No. 6. California Test Bureau, Los Angeles.

3. Clark, Willis W. Articulated and Integrated Measuring Instruments for Practical Evaluation Programs, Educational Bulletin No. 20. California Test Bureau, Los Angeles, 1954.

4. Durost, Walter N. What Constitutes a Testing Program for Elementary and Junior High School, Test Service Notebook No. 1. World Book Company, Tarrytown-on-Hudson, New York, 1956.

5. Fisher, Marie R. Using Test Results to Improve the Education of the Individual Child, Test Service Bulletin No. 76. World Book Company, Tarrytown-on-Hudson, New York, p. 3.

6. Kent Area Guidance Council. A Proposed 12-Year Testing Program, Ohio Scholarship Tests. State Department of Education, Columbus, Ohio, 1959, pp. 14-18.

7. Noll, Victor H. Introduction to Educational Measurement. Houghton Mifflin Company, Boston, 1957, pp. 343-352.

8. Remmers, H. H., N. L. Gage, and J. Francis Rummel. A Practical Introduction to Measurement and Evaluation. Harper and Brothers, New York, 1960, pp. 101-103.

9. Ross, C. C., and Julian C. Stanley. Measurement in Today's Schools. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1954, p. 210.

10. Schwartz, Alfred, and Stuart C. Tiedman. Evaluating Student Progress in the Secondary School. Longmans, Green and Company, New York, 1957, pp. 273-274.

11. Thorndike, Robert L., and Elizabeth Hagen. Measurement and Evaluation in Psychology and Education. John Wiley and Sons, Inc., New York, 1955, pp. 433-438.

12. Traxler, Arthur E., Robert Jacobs, Margaret Selover, and Agatha Townsend. Introduction to Testing and the Use of Test Results in Public Schools. Harper and Brothers, New York, 1953, pp. 20-27.

13. World Book Company. Planning the Secondary School Testing Program. World Book Company, Tarrytown-on-Hudson, New York.

APPENDIX G

GRADE LEVEL AND FREQUENCY OF TESTING

BY STANDARDIZED TESTS IN KINDERGARTEN THROUGH GRADE 8

Table 113. Grade Level at Which Reading Is Generally Measured by Standardized Tests in Kindergarten Through Grade 8

Grade Level	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	8.0	10.0	11.5	17.5	12.1	4.5	10.6
I	65.0	62.8	48.4	52.5	57.6	36.4	55.7
II	74.0	74.3	59.2	67.5	57.6	45.4	65.6
III	88.0	77.1	80.0	85.0	81.8	68.2	81.5
IV	85.0	77.1	76.9	90.0	78.8	54.5	79.2
V	90.0	78.5	75.4	80.0	69.7	59.1	78.7
VI	91.0	78.5	82.3	90.0	90.9	86.4	85.6
VII	86.0	81.4	73.1	70.0	75.7	40.9	76.2
VIII	91.0	84.3	69.2	87.5	78.8	72.7	80.2

Table 114. Grade Level at Which Arithmetic Is Generally Measured by Standardized Tests in Kindergarten Through Grade 8

Grade Level	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	6.0	4.2	4.6	10.0	3.0	4.5	5.3
I	48.0	47.1	28.5	35.0	21.2	18.2	36.2
II	62.0	62.8	50.0	52.5	33.3	18.2	52.4
III	90.0	72.8	73.8	77.5	60.6	54.5	73.4
IV	86.0	80.0	73.1	87.5	72.7	45.4	77.5
V	87.0	87.1	75.4	82.5	75.7	50.0	76.7
VI	94.0	78.5	80.8	90.0	87.9	63.6	84.3
VII	88.0	80.0	66.9	62.5	63.6	40.9	72.4
VIII	87.0	84.3	78.5	85.0	84.8	54.5	81.5

Table 115. Grade Level at Which Language Usage is Generally Measured by Standardized Tests in Kindergarten Through Grade 8

Grade Level	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	4.0	2.8	3.1	5.0	3.0	4.5	3.3
I	43.0	34.3	23.1	17.5	15.1	13.6	28.3
II	58.0	50.0	40.0	27.5	27.3	13.6	42.5
III	81.0	75.6	72.3	55.0	60.6	68.2	72.1
IV	86.0	72.8	73.8	75.0	69.7	40.9	74.7
V	83.0	74.3	72.3	70.0	66.7	45.4	73.2
VI	87.0	77.1	78.5	82.5	81.8	77.3	81.0
VII	83.0	77.1	70.0	62.5	63.6	40.9	71.6
VIII	88.0	85.7	77.7	77.5	75.7	72.7	81.3

Table 116. Grade Level at Which Spelling Is Generally Measured by Standardized Tests in Kindergarten Through Grade 8

Grade Level	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	5.0	17.1	3.1	5.0	3.0	4.5	6.3
I	48.0	42.8	26.9	20.0	12.1	13.6	32.4
II	63.0	57.1	44.6	40.0	27.3	13.6	47.8
III	81.0	71.4	70.7	60.0	54.5	59.1	70.3
IV	84.0	71.4	70.7	80.0	66.7	40.9	73.2
V	85.0	72.8	72.3	72.5	57.6	54.5	73.4
VI	84.0	75.6	68.4	85.0	81.8	86.4	80.0
VII	83.0	75.6	66.9	60.0	54.5	40.9	69.3
VIII	87.0	55.7	74.6	82.5	69.7	72.7	74.7

Table 117. Grade Level at Which Social Studies Is Generally Measured by Standardized Tests in Kindergarten Through Grade 8

Grade Level	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	3.0	2.8	2.3	2.5	3.0	4.5	2.8
I	30.0	22.8	16.9	12.5	12.1	4.5	19.7
II	38.0	31.4	25.4	20.0	12.1	4.5	26.8
III	57.0	47.1	40.8	25.0	18.2	27.3	41.8
IV	65.0	60.0	50.8	45.0	27.3	13.6	51.4
V	72.0	71.4	56.9	55.0	33.3	13.6	58.7
VI	75.0	68.6	59.2	67.5	42.4	40.9	63.3
VII	74.0	72.8	56.1	50.0	33.3	13.6	58.7
VIII	78.0	78.5	61.5	72.5	42.4	40.9	67.1

Table 118. Grade Level at Which Science Is Generally Measured by Standardized Tests in Kindergarten Through Grade 8

Grade Level	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	3.0	2.8	1.5	2.5	3.0	-	2.3
I	30.0	21.4	13.1	12.5	12.1	-	18.0
II	39.0	27.1	21.5	17.5	12.1	4.5	24.8
III	55.0	40.0	35.4	22.5	15.1	18.2	37.2
IV	62.0	52.8	45.4	45.0	24.2	9.1	47.1
V	66.0	62.8	50.0	52.5	30.3	9.1	52.6
VI	71.0	65.6	56.1	62.5	42.4	31.8	59.7
VII	70.0	64.3	53.8	52.5	33.3	9.1	55.4
VIII	75.0	71.4	53.8	70.0	39.4	40.9	62.0

Table 119. Grade Level at Which Work-Study Skills Are Generally Measured by Standardized Tests in Kindergarten Through Grade 8

Grade Level	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K	5.0	1.4	2.3	5.0	3.0	4.5	3.3
I	28.0	20.0	10.8	15.0	6.1	4.5	16.4
II	37.0	21.4	16.9	22.5	6.1	4.5	21.8
III	55.0	38.6	35.4	32.5	21.2	27.3	39.0
IV	60.0	41.4	46.9	60.0	24.2	18.2	47.1
V	64.0	58.6	55.4	57.5	33.3	22.7	54.7
VI	69.0	58.6	60.0	67.5	42.4	36.4	60.0
VII	68.0	64.3	57.7	45.0	36.4	27.3	56.7
VIII	70.0	67.1	56.9	67.5	42.4	36.4	60.7

Table 120. Number of Times Reading Is Measured by Standardized Tests During Pupil's School Career in Kindergarten Through Grade 8

Number of Times	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Once	2.0	-	1.5	2.5	-	9.1	1.8
Twice	4.0	7.1	4.6	5.0	6.1	4.5	5.1
Three times	4.0	5.7	8.5	2.5	15.1	13.6	7.1
More than three times	88.0	79.2	81.8	90.0	75.1	72.7	82.9
Annually	55.0	55.1	40.8	37.5	30.2	13.6	44.5

Table 121. Number of Times Arithmetic Is Measured by Standardized Tests During Pupil's School Career in Kindergarten Through Grade 8

Number of Times	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Once	2.0	-	1.5	2.5	6.1	9.1	2.3
Twice	6.0	10.0	6.9	10.0	9.1	9.1	7.6
Three times	5.0	8.6	8.5	10.0	24.2	40.9	10.6
More than three times	85.1	74.1	77.6	80.0	54.5	40.9	75.2
Annually	45.0	43.6	29.2	27.5	15.1	9.1	33.7

Table 122. Number of Times Language Usage Is Measured by Standardized Tests During Pupil's School Career in Kindergarten Through Grade 8

Number of Times	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Once	2.0	-	3.1	2.5	3.0	4.5	5.3
Twice	6.0	10.0	6.1	15.0	12.1	9.1	8.3
Three times	7.0	5.7	10.0	10.0	21.2	40.9	11.1
More than three times	38.0	32.8	22.3	15.0	9.1	9.1	72.8
Annually	80.0	75.5	76.1	72.5	34.5	40.9	25.5

Table 123. Number of Times Spelling Is Measured by Standardized Tests During Pupil's School Career in Kindergarten Through Grade 8

Number of Times	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Once	3.0	-	2.3	-	3.0	4.5	2.0
Twice	3.0	10.0	4.6	12.5	6.1	9.1	6.3
Three times	8.0	4.2	9.2	10.0	21.2	45.4	11.1
More than three times	78.0	75.5	74.5	75.0	54.5	36.3	71.9
Annually	41.0	35.6	24.6	20.0	9.1	9.1	28.1

Table 124. Number of Times Social Studies Is Measured by Standardized Tests During Pupil's School Career in Kindergarten Through Grade 8

Number of Times	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Once	2.0	-	3.8	5.0	12.1	18.2	4.3
Twice	7.0	10.0	10.0	25.0	6.1	13.6	10.6
Three times	6.0	8.6	4.6	12.5	6.1	18.2	7.3
More than three times	69.0	64.2	55.3	45.0	27.2	9.0	54.5
Annually	26.0	22.8	13.0	10.0	3.0	4.5	16.5

Table 125. Number of Times Science Is Measured by Standardized Tests During Pupil's School Career in Kindergarten Through Grade 8

Number of Times	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Once	3.0	-	3.8	7.5	12.1	22.7	5.1
Twice	10.0	10.0	9.2	25.0	9.1	9.1	11.3
Three times	6.0	10.0	8.5	10.0	6.1	9.1	8.1
More than three times	63.0	56.9	49.1	42.5	24.2	9.0	49.1
Annually	25.0	21.3	12.3	10.0	3.0	4.5	15.7

Table 126. Number of Times Work-Study Skills Are Measured by Standardized Tests During Pupil's School Career in Kindergarten Through Grade 8

Number of Times	Percentage of Responses According to Enrollment Categories						
	I	II	III	IV	V	VI	VII
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Once	3.0	1.4	3.1	10.0	12.1	9.1	4.5
Twice	10.0	8.6	8.5	15.0	3.0	4.5	8.9
Three times	5.0	7.1	5.4	2.5	12.1	18.2	6.6
More than three times	63.0	54.0	54.5	52.5	30.2	18.0	52.4
Annually	23.0	18.5	10.7	12.5	3.0	4.5	14.4

APPENDIX H

COMMENTS PERTAINING TO TEST UTILIZATION

PROGRAM EXPANSION AND ENRICHMENT

1. Extended the developmental reading program to include grades seven and eight.

2. Established a broadened reading program at the senior high school level including a reading laboratory, etc.

3. Established review math classes in grade twelve after discovering via a testing program, a lower percentile rating in math than seemed desirable.

4. Established a ninth grade accelerated algebra group.

5. Initiated a class for "gifted" pupils in grades III-VI.

6. Added specific work in English Composition in the junior and senior years.

7. Introduced an advanced math seminar in grade XII.

8. Introduced an advanced placement program as the results of Iowa Developmental Tests.

9. Introduced a program for students of superior ability at the secondary level.

10. Initiated an enriched program in library and language arts for certain 5th and 6th grade students

11. Revised graduation requirements to include more work in math and science.

12. Established special reading groups at the secondary level.

13. Established a program for the "gifted" pupil in the secondary schools.

14. Established a special class at secondary level for low ability level pupils.

15. Expanded guidance services throughout the schools.

16. Developed an expanded program for the non-academic student.

17. Established accelerated classes in math and science.

EXPANDING SERVICES AND IMPROVING INSTRUCTIONAL PROGRAM

1. Aided in adding a remedial reading supervisor.
2. Used to refine groups and effecting homogeneous grouping, particularly at the junior high school level, in basic academic areas.
3. Used to determine the effectiveness of grouping at the junior high school.
4. Used to inaugurate classes for educable and trainable students.
5. Used to set up special classes in the elementary school for seriously retarded pupils.
6. Initiated tutorial groups at the junior high schools for underachievers.
7. Used as the basis for establishing two in-service programs for teachers for improving the teaching in language arts and mathematics.
8. Test summaries served as the basis for in-service program and workshop for teachers.
9. Used as a basis for regrouping graded pupils.
10. Aided in adding a remedial teacher for the junior high schools.
11. Used as the basis for textbook selection by textbook selection committee.
12. Used to aid in selection of students for trade school, as a result we have the lowest drop-out rate in the history of the school.
13. Used to establish a new system of reporting, using stanines as the basis of this reporting system.
14. Used to set up a new "track" program at the senior high school.
15. Used to eliminate grade barriers in grades I-III.

CURRICULUM REVISION

1. Revised mathematics curriculum at secondary level.
2. Stressed fundamentals of arithmetic at secondary level as well as in elementary program.
3. Revised arithmetic program and selected new texts as result of weakness shown in testing.
4. Revised entire development reading program.
5. Revised spelling program.
6. Revised English program.
7. Revising social studies and science program as result of weakness evidenced in test results.
8. Changes science sequence at the senior high school level.
9. Revised entire junior high school curriculum due to study made with Iowa Test of Educational Development.
10. Revised group guidance content in grades VII-XII.
11. Revised the elementary time allotment schedule.
12. Revised the English program in grades I-VII.
13. Revised the foreign language offerings at the junior and senior high school.

INSTRUCTIONAL PROCEDURES AND METHODOLOGY

1. Increased emphasis upon problem solving in entire mathematics program.

2. System-wide study being made of the teaching of spelling and arithmetic because of weaknesses noted in test results.

3. Concerted system-wide effort made to improve study skills.

4. Greater stress placed in teaching arithmetic fundamentals in grades IV-VI.

5. Studying method in teaching vocabulary as result of language tests.

6. Placing greater emphasis upon written composition at the secondary level.

7. Using scattergram noted considerable underachievement in arithmetic. Total program is now being studied and will be overhauled to improve teaching effectiveness.

8. Developed study guide for teachers in order to strengthen the teaching of work and study skills and habits.

9. As result of item analysis of arithmetic results, are placing more emphasis upon drill in the teaching of arithmetic at all levels.

10. Directed more attention to map-reading techniques and knowledge in grade 6 when results of Iowa Every-pupil Tests of Basic Study Skills showed pupils to be below par in this area.

11. Strengthened foreign language program, particularly the oral and aural approach.

12. Stressed understanding in teaching methods as contrasted with purely teaching facts, this because of a definite weakness noted in this area in tests recently administered.

13. Revised system of teaching spelling.

14. Formed an arithmetic vocabulary at each grade level to strengthen a weakness in arithmetic reasoning shown in testing.

PUBLIC RELATIONS MEDIA TO IMPROVE OVER-ALL PROGRAM

1. A discussion of test results has helped to get the Board of Education to see the need for adequate instructional supplies.

2. Test results are used to educate the Board of Education to the fact that wide range of abilities exist at each grade level, thus the need for differentiated instruction.

3. Test results were used to convince the Board of Education of the need for a remedial reading specialist at the senior high school level.

4. Mathematics achievement tests of high school pupils were administered and results were presented to the Board of Education to prove that many changes in teachers are not good for pupils. They must realize that they are responsible to have an attractive salary schedule to retain teachers and thus have normal achievement.

5. Using test results we are trying to educate voters on remedial work and to recognize the need for trained personnel in the guidance field. We have also used these results to aid in the elimination of "mongrel" course selection by students and revising the high school schedule.

6. Used results to initiate a program of early entrance to school for youngsters whose C.A. does not permit entrance, but whose M.A. indicates they are ready for school.

7. Arguments for an upward revision of the minimum school entrange age were strengthened by reference to findings on readiness tests.

8. To aid in selling the school board and lay people the advantage of forming a larger administrative unit.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Afra, Sister M., S.C.L., Standardized Tests: Their Uses and Abuses, Test Service Bulletin No. 58. The World Book Company, Tarrytown-on-Hudson, New York, 1945.

2. Ahman, J. Stanley, and Marvin D. Glock, Evaluating Pupil Growth. Allyn and Bacon, Boston, 1958.

3. Ahrens, Edward H. J., "The Validity of the Questionnaire," Science Education (February, 1950), 34:41-42.

4. Allen, Margaret E., Status of Measurement Courses for Undergraduates in Teacher-Training Institutions, Thirteenth Yearbook, National Council on Measurements Used in Education. The Council (Secretary, Robert D. North, 21 Audubon Avenue, New York), 1956, pp. 69-73.

5. Allen, Richard D., Tests for Instruction and Guidance, Test Service Bulletin No. 46. World Book Company, Tarrytown-on-Hudson, New York.

6. Alpern, Hyman, "Evaluation in the Academic High Schools of New York City," High Points (March, 1946), 28:13-26.

7. American Association of School Administrators, American School Curriculum, Thirty-first Yearbook. Washington, D. C., 1953.

8. American Educational Research Association, "Educational and Psychological Testing," Review of Educational Research (February, 1947), 17:1.

9. American Educational Research Association, "Educational and Psychological Testing," Review of Educational Research (February, 1950), 20:1.

10. American Educational Research Association, "Educational and Psychological Testing," Review of Educational Research (February, 1953), 23:1.

11. American Educational Research Association, "Educational and Psychological Testing," Review of Educational Research (February, 1956), 26:1.

12. American Educational Research Association, "Educational and Psychological Testing," Review of Educational Research (February, 1959), 29:1.

13. American Psychological Association, "Technical Recommendations for Psychological Tests and Diagnostic Techniques," Supplement to the Psychological Bulletin (March, 1954), 51:2:2. American Psychological Association, Washington, D. C.

14. Bacon, Francis L., "Using Tests in the Secondary Schools," The Bulletin of the National Association of Secondary School Principals (April, 1943), 33:41-50.

15. Beer, Carl R., A County Program of Standardized Testing, Test Service Bulletin No. 60. World Book Company, Tarrytown-on-Hudson, New York.

16. Belanger, Laurence L., "Testing in California Secondary Schools," California Journal of Secondary Education (February, 1947), 22: 108-111.

17. Bingham, Walter Van Dyke, Aptitude and Aptitude Testing. Harper and Brothers, New York, 1937.

18. Bloom, Benjamin (Editor), Taxonomy of Educational Objectives. Longmans, Green and Company, New York, 1956.

19. Boag, Audrey K., "Standardized Tests: How, When, Why," Instructor (October, 1935), 65:24.

20. Boyer, Philip A., "The Administration of Learning Groups in Elementary Schools," The Grouping of Pupils, Part I, The Thirty-fifth Yearbook, National Society for the Study of Education. Public School Publishing Company, Bloomington, Illinois, 1936.

21. Bradfield, James M., and H. Stewart Moredock, Measurement and Evaluation in Education. The Macmillan Company, New York, 1957.

22. Bridges, Claude F., Some Basic Considerations in Determining the Significance of Achievement Test Results, Test Service Bulletin No. 66. World Book Company, Tarrytown-on-Hudson, New York.

23. Brown, Stanley B., Mary Louise Lyda, and Carter V. Good, Research Studies in Education, A Subject Index. Phi Delta Kappa, Bloomington, Indiana, 1940-1959.

24. Brown, Woodrow A., Testing in Pennsylvania's Public Kindergartens, Test Service Bulletin No. 67. World Book Company, Tarryton-on-Hudson, New York.

25. Brownell, William A., "Some Neglected Criteria for Evaluating Classroom Tests," National Elementary Principal (July, 1937), 16:485-492.

26. Bureau of Examinations and Testing, New York State Education Department, The School Testing Program, A Guide to the Selection and Use of Standardized Tests, No. 1454. University of the State of New York, Albany, July, 1958.

27. Buros, Oscar K. (Editor), The Fourth Mental Measurements Yearbook. Gryphon Press, Highland Park, New Jersey, 1953.

28. Buros, Oscar K. (Editor), The Fifth Mental Measurements Yearbook. Gryphon Press, Highland Park, New Jersey, 1959.

29. Burton, William H., "Implications for Organization of Instruction and Instructional Organization," Learning and Instruction, Part I, The Forty-ninth Yearbook of the National Society for the Study of Education. University of Chicago Press, Chicago, 1950.

30. California Test Bureau, Administrative Uses of Test Results, Educational Bulletin No. 4. California Test Bureau, Los Angeles, 1951.

31. California Test Bureau, Conducting High School Guidance Programs, Educational Bulletin No. 7. California Test Bureau, Los Angeles, 1945.

32. California Test Bureau, How Tests Can Improve Your Schools, Educational Bulletin No. 1. California Test Bureau, Los Angeles, 1945.

33. California Test Bureau, How to Select Tests, Educational Bulletin No. 2. California Test Bureau, Los Angeles, 1945.

34. Clark, Willis W., Articulated and Integrated Measuring Instruments for Practical Evaluation Programs, Educational Bulletin No. 20. California Test Bureau, Los Angeles, 1954.

35. Clark, Willis W., Identifying Difficulties in Learning Arithmetic, Educational Bulletin No. 9. California Test Bureau, Los Angeles, 1951.

36. Cole, Robert D., and Fred von Borgersrode, "A Scale for Rating Standardized Tests," School of Education Record of the University of North Dakota (October, 1928), 14:11-15.

37. Coleman, William, "Assisting Teachers in Using Test Results," Personnel and Guidance Journal (September, 1950), 36:38-40.

38. Coleman, William, and Dorothy Manley Collett, "Development and Applications of Structured Tests of Personality," Review of Educational Research (February, 1959), 29:57-72.

39. Committee on Test Standards of the American Educational Research Association and the National Council of Measurements Used in Education, Technical Recommendations for Achievement Tests. American Educational Research Association, Washington, D. C., 1955.

40. Connecticut State Department of Education, Testing Programs in the Public Elementary and Secondary Schools of Connecticut, Bureau of Research and Statistics, Connecticut State Department of Education, Hartford, 1955.

41. Cook, Walter W., "The Functions of Measurement in the Facilitation of Learning," Chapter I in Educational Measurement, E. F. Lindquist, Editor. American Council on Education, Washington, D. C., 1951.

42. Cook, Walter W., "Use of Tests in the Supervisory Program," in Appraising the Elementary School Program, The Sixteenth Yearbook, Bulletin of the Department of Elementary School Principals, National Elementary Principals. National Education Association, Washington, D. C., July, 1937.

43. Cook, Walter W., "What Teachers Should Know About Measurement," The Fifteenth Yearbook of the National Council on Measurements Used in Education. Educational Testing Service, Princeton, New Jersey, 1958.

44. Cromwell, Floyd R., A Basic Program of Guidance, Maryland School Bulletin No. 1. State Department of Education, Baltimore, 1943.

45. Crook, Frances E., "The Classroom Teacher and Standardized Tests," Teachers College Record (December, 1956), 58:159-168.

46. Crook, Frances E., "Elementary School Testing Programs, Problems, and Practices," Teachers College Record (November, 1959), 61: 76-85.

47. Dan Golenpaul Associates, Information Please Almanac, 1959. The Macmillan Company, New York, 1959.

48. Darley, John G., Testing and Counseling in the High School Guidance Program. Science Research Associates, Chicago, 1943.

49. Davis, Frederick B., "Testing and the Use of Test Results," Review of Educational Research (February, 1953), 23:5-10.

50. Dobbins, John E., "How Can the Results of a Testing Program Be Used Most Effectively," The Bulletin of the National Association of Secondary School Principals (April, 1958), 42:66-68.

51. Dobbins, John E., "What Parents Need to Know About Tests and Testing," National Elementary Principal (September, 1954), 34: 152-160.

52. Doby, John T. (Editor), An Introduction to Social Research. The Stackpole Co., Harrisburg, Pennsylvania, 1954.

53. Douglass, Harl R., and Herbert F. Spitzer, "The Importance of Teaching for Understanding," Chapter II in The Measurement of Understanding, Part I, The Forty-fifth Yearbook, National Society for the Study of Education. University of Chicago Press, Chicago, 1946.

54. Dunlap, Jack W., To Test or Not to Test, Test Service Bulletin No. 37. World Book Company, Tarrytown-on-Hudson, New York.

55. Durost, Walter N., The Characteristics, Use, and Computation of Stanines, Test Service Notebook No. 23. World Book Company, Tarrytown-on-Hudson, New York, 1959.

56. Durost, Walter N., Tests and the Junior High School Guidance Counselor, Test Service Notebook No. 2. World Book Company, Tarrytown-on-Hudson, New York.

57. Durost, Walter N., What Constitutes a Minimal Testing Program for Elementary and Junior High School, Test Service Notebook No. 1, Revised. World Book Company, Tarrytown-on-Hudson, New York, 1956.

58. Educational Index. H. W. Wilson Co., New York, 1950-1960.

59. Educational Policies Commission of the National Education Association, Education of the Gifted. National Education Association, Washington, D. C., 1950.

60. Educational Testing Service, Essential Characteristics of a Testing Program, Education and Advisory Service Series No. 2. Educational Testing Service, Princeton, New Jersey, 1956.

61. Eells, Walter Crosby, Surveys of Higher American Education. The Cargenie Foundation for the Advancement of Teaching, New York, 1937.

62. Ellis, Albert, "Personality Questionnaire," Review of Educational Research (February, 1947), 17:53-64.

63. Engelhart, Max D., "Making Testing Meaningful to Teachers Through Local Test Construction and Analysis of Test Data," 1953 Invitational Conference on Testing Problems. Educational Testing Service, Princeton, New Jersey, 1954.

64. Engelhart, Max D., "Testing and Use of Test Results," Review of Educational Research (February, 1956), 26:5-13.

65. Erickson, Elmer J., "What Kind of Testing Program for Today's Schools?" The Bulletin of the National Association of Secondary School Principals (April, 1952), 36:160-166.

66. Eurich, Alvin C., and C. Gilbert Wrenn, "Appraisal of Student Characteristics and Needs," Guidance in Educational Institutions, The Thirty-seventh Yearbook of the National Society for the Study of Education, Part I. Public School Publishing Company, Bloomington, Illinois, 1938.

67. Farley, Belmont, What to Tell the People About the Schools. Bureau of Publications, Teachers College, Columbia University, New York, 1929.

68. Fisher, Marie R., Using Test Results to Improve the Education of the Individual Child, Test Service Bulletin No. 76. World Book Company, Tarrytown-on-Hudson, New York.

69. Fox, Benjamin, "Evaluation in the Vocational High Schools," High Points (March, 1946), 28:26-31.

70. French, Will, and associates, Behavioral Goals of General Education in High School. Russell Sage Foundation, New York, 1957.

71. Froehlich, Clifford P., and Arthur S. Benson, Guidance Testing. Science Research Associates, Chicago, 1948.

72. Fullmer, Daniel W., "The Testing Program: What Constitutes Minimum Essentials," The Bulletin of the National Association of Secondary School Principals (May, 1958), 42:80-87.

73. Good, Carter V., Introduction to Educational Research. Appleton-Century-Crofts, Inc., New York, 1959.

74. Green, Harry A., Albert N. Jorgensen, and Raymond J. Gerberich, Measurement and Evaluation in the Elementary School (Second Edition). Longmans, Green and Company, New York, 1953.

75. Grimes, H. F., "A Testing Program's Excuse for Being," Clearing House (September, 1954), 1:33-34.

76. Hahn, Milton E., and Malcolm S. MacLean, General Clinical Counseling in Educational Institutions. McGraw-Hill Book Company, Inc., New York, 1950.

77. Harvey, Philip R., The Use of the Graduate Record Examinations by Colleges and Universities. Educational Testing Service, Princeton, New Jersey, 1957. (Mimeographed.)

78. Heimann, Robert A., and John W. M. Rothney, "Development and Applications of Projective Techniques," Review of Educational Research (February, 1959), 29:73-83.

79. Hildreth, Gertrude, Using Readiness Test Results, Test Service Notebook No. 10. World Book Company, Tarrytown-on-Hudson, New York, 1950.

80. Holmes, John L., Guidance Testing and the Identification of Pupil Characteristics, Educational Bulletin No. 21. California Test Bureau, Los Angeles, 1958.

81. Huddleston, Edith M. (Editor), The Fourteenth Yearbook of the National Council on Measurements Used in Education. Educational Testing Service, Princeton, New Jersey, 1957.

82. Jennings, Helen Hall, Sociometry in Group Relations. American Council on Education, Washington, D. C., 1948.

83. Jordan, A. M., Measurement in Education. McGraw-Hill Book Company, Inc., New York, 1953.

84. Kearney, Nolan C., Elementary School Objectives, Parts II and III. Russell Sage Foundation, New York, 1953.

85. Kent Area Study Council, A Proposed Twelve-Year Testing Program, Ohio Scholarship Tests. State Department of Education, Columbus, Ohio, 1959.

86. King, Joseph E., "Using Tests in the Modern Secondary School," The Bulletin of the National Association of Secondary School Principals (December, 1948), 32:122 pages.

87. Kirby, B. C., "Minimum Testing Programs," Journal of Education (January, 1948), 131:1-24.

88. Koos, Leonard V., The Questionnaire in Education, A Critique and Manual. The Macmillan Company, New York, 1928.

89. Krugman, Judith I., and J. Wayne Wrightstone, A Guide to the Use of Anecdotal Records, Educational Research Bulletin No. 11. Board of Education, New York, 1949.

90. Krugman, Morris, "Changing Methods of Appraising Personality," 1956 Invitational Conference on Testing Problems. Educational Testing Service, Princeton, New Jersey, 1957.

91. Kvaraceus, W. C., In-Service Training in Educational Measurement, Test Service Bulletin No. 64. World Book Company, Tarrytown-on-Hudson, New York,

92. Larson, Carl A., and William H. McCreary, "Testing Programs and Practices in California Public Secondary Schools," California Journal of Secondary Education (November, 1956), 31:389-402.

93. Lass, A. N., and J. Wayne Wrightstone, "Evaluation in the Secondary Schools of New York City," High Points (March, 1946), 28:11-13.

94. Lazar, Mary, and Miriam Aronow, "Research Uses of Testing Program," Education (March, 1957), 77:395-398.

95. Leake, James D., Evaluation in the Denver Public Schools, Test Service Bulletin No. 83. World Book Company, Tarrytown-on-Hudson, New York, 1958.

96. Lee, J. Murray, "Essential Elements of an Adequate Testing Program," in Appraising the Elementary School Program, The Sixteenth Yearbook, National Elementary School Principals. National Education Association, Washington, D. C., July, 1937.

97. Lee, J. Murray, Testing Programs for Secondary Schools, Teachers College Contributions to Education. Teachers College, Columbia University, New York, 1934.

98. Lee, J. Murray, and Doris May Lee, The Child and His Curriculum. Appleton-Century-Crofts, Inc., New York, 1950.

99. Lee, J. Murray, and David Segel, Testing Practices of High School Teachers, United States Office of Education Bulletin No. 9. United States Government Printing Office, Washington, D. C., 1936.

100. Lennon, Roger T., "Discussion of the School Administrators' Problems," in Invitational Conference on Testing Problems, 1957 Proceedings. Educational Testing Service, Princeton, New Jersey, 1958.

101. Lennon, Roger T., Planning a Testing Program, Test Service Bulletin No. 55. World Book Company, Tarrytown-on-Hudson, New York.

102. Lennon, Roger T., Testing in the Secondary School, Test Service Notebook No. 20. World Book Company, Tarrytown-on-Hudson, New York, 1957.

103. Lennon, Roger T., and Harold F. Bligh, Finding Mathematics and Science Talent in the Junior High School, Test Service Bulletin No. 91. World Book Company, Tarrytown-on-Hudson, New York.

104. Leonard, Eugenie A., and Anthony C. Tucker, The Individual Inventory in Guidance Programs in Secondary Schools, United States Office of Education Vocational Division, Bulletin No. 215, Occupational Information and Guidance Series No. 7. United States Government Printing Office, Washington, D. C., 1941.

105. Lewerenz, Alfred Speir, "New Developments in Evaluating Achievement in the Public Schools of Los Angeles," Education (December, 1950), 71:237-244.

106. Lindquist, E. F. (Editor), Educational Measurement. American Council on Education, Washington, D. C., 1951.

107. Lindquist, E. F., "Standardized Achievement Tests and Their Relation to Curriculum Content," in Appraising the Elementary School Program, The Sixteenth Yearbook, Department of Elementary School Principals. National Education Association, Washington, D. C., July, 1937.

108. MacMinn, Paul, "A Few Basic Facts About Tests," in Understanding Tests, United States Department of Health, Education, and Welfare, Office of Education. United States Government Printing Office, Washington, D. C., 1960.

109. Marks, Eli S., "Some Sampling Problems in Educational Research," Journal of Educational Psychology (February, 1951), 42:85-96.

110. Martin, C. W., "The Elementary School Testing Program," The National Elementary Principal (February, 1946), 25:6-9.

111. Mayo, Samuel T., "Testing and the Use of Test Results," Review of Educational Research (February, 1959), 29:5-14.

112. Michaelis, John W., "Current Practices in Evaluation in City School Systems," Educational and Psychological Measurement (Spring, 1950), 9:15-22.

113. Michaelis, John W., and Charles Howard, "Current Practices in City School Systems in California," Journal of Educational Research (December, 1949), 43:250-260.

114. Miller, Carroll H., "Guidance and Programs of Testing," in Understanding Testing, United States Department of Health, Education, and Welfare, Office of Education. United States Government Printing Office, Washington, D. C., 1960.

115. Monroe, Walter S., "Educational Measurement in 1920 and in 1945," Journal of Educational Research (January, 1945), 38:334-340.

116. Monroe, Walter S. (Editor), Encyclopedia of Educational Research. The Macmillan Company, New York, 1952.

117. Mooney, R. L., and M. A. Price, Manual to accompany Ross L. Mooney's Problem Check List, College Form, Bureau of Educational Research. Ohio State University, Columbus, Ohio, 1948.

118. Morrison, Harriet Barthelmess, The Fall Testing Program, Test Service Bulletin No. 62. The World Book Company, Tarrytown-on-Hudson, New York, 1948.

119. National Association of Guidance Supervisors and Counselor Trainers, A National Study of Existing and Recommended Practices for Assisting Youth Adjustment in Selected Elementary Schools of the United States. Ann Arbor Publishers, Ann Arbor, Michigan, 1953.

120. New England Association of School Superintendents, 1957-58 Directory of the New England Association of School Superintendents.

121. New York State Education Department, The School Testing Program, University of the State of New York Bulletin No. 1415. State Education Department, Albany, 1953.

122. Noll, Victor H., Introduction to Educational Measurement. Houghton Mifflin Company, Boston, 1957.

123. Noll, Victor H., "Requirements in Educational Measurement for Prospective Teachers," School and Society (September, 1955), 82:88-90.

124. Noll, Victor H., and Walter N. Durost, Measurement Practices and Preferences of High School Teachers, Test Service Notebook No. 8. World Book Company, Tarrytown-on-Hudson, New York.

125. North, Robert D., "Testing Programs of Public School Members of the Educational Records Bureau--Report of a Questionnaire Survey," 1956 Achievement Testing Program in Independent Schools and Supplementary Studies, Bulletin No. 68. Educational Records Bureau, New York, 1956.

126. North, Robert D., "The Use of Multi-Factor Aptitude Tests in School Counseling," Proceedings of the 1955 Invitational Conference on Testing Problems. Educational Testing Service, Princeton, New Jersey, 1956.

127. Orata, Pedro T., "Evaluating Evaluation," Journal of Educational Research (May, 1940), 33:641-661.

128. Otis, Arthur S., Otis Scale for Rating Standardized Tests. World Book Company, Tarrytown-on-Hudson, New York,

129. Payne, Stanley L., The Art of Asking Questions. Princeton University Press, Princeton, New Jersey, 1951.

130. Psychological Corporation, Human Resources and the Aptitude Inventory, Test Service Bulletin No. 4. The Psychological Corporation, New York, May, 1951.

131. Psychological Corporation, Methods of Expressing Test Scores, Test Service Bulletin No. 48, January, 1955. The Psychological Corporation, New York, 1956.

132. Psychological Corporation, Norms Must Be Relevant, Test Service Bulletin No. 39, May, 1950. The Psychological Corporation, New York, 1950.

133. Remmers, H. H., and N. L. Gage, Educational Measurement and Evaluation. Harper and Brothers, New York, 1943.

134. Remmers, H. H., N. L. Gage, and J. Francis Rummel, A Practical Introduction to Measurement and Evaluation. Harper and Brothers, New York, 1960.

135. Rinsland, H. D., "A Form for Briefing and Evaluating Standardized Tests," Journal of Educational Research (January, 1949), 42:371-375.

136. Rivlin, Harry N., The Teacher's Role in Achievement Testing, Test Service Notebook No. 9. World Book Company, Tarrytown-on-Hudson, New York, 1950.

137. Roeber, Edward C., "A Meaningful Record of Tests," Educational and Psychological Measurement (Autumn, 1948), 8:397-400.

138. Rosinski, Edwin F., "A Status Survey of Existing Testing Programs in Member Schools of the Western New York School Study Council," in The Fourteenth Yearbook of the National Council on Measurements Used in Education, Edith M. Huddleston (Editor). Educational Testing Service, Princeton, New Jersey, 1957.

139. Ross, C. C., and Julian C. Stanley, Measurement in Today's Schools. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1954.

140. Rummel, J. Francis, An Introduction to Research Procedures in Education. Harper and Brothers, New York, 1958.

141. Scates, Douglas E., and Alice V. Yeomans, The Effect of Questionnaire Form on Course Requests of Employed Adults. American Council on Education, Washington, D. C., 1950.

142. Schwartz, Alfred, and Stuart C. Tiedman, Evaluating Student Progress in the Secondary School. Longmans, Green and Company, New York, 1957.

143. Seashore, Harold G., "How May Tests Be Used to Obtain Better Articulation of the Total Educational Program," The Bulletin of the National Association of Secondary School Principals (April, 1954), 38:283-285.

144. Segel, David, Nature and Use of the Cumulative Record, Bulletin No. 3, 1938, United States Office of Education. United States Government Printing Office, Washington, D. C., 1938.

145. Segel, David, "Survey and Trend Studies," Review of Educational Research (December, 1942), 22:492-501.

146. Segel, David, and Raymond J. Gerberich, "Overview of Educational and Psychological Testing 1946 to 1949," Review of Educational Research (February, 1950), 20:5-16.

147. Shane, Harold G., "Recent Developments in Elementary School Evaluation," Journal of Educational Research (March, 1951), 44: 491-500.

148. Smallenburg, Harry, "Developing a Program for Evaluation of Pupil Growth in the Burbank Schools," Educational Administration and Supervision (October, 1942), 28:550-553.

149. Smith, Eugene R., Ralph W. Tyler, and others, Appraising and Recording Student Progress. Harper and Brothers, New York, 1942.

150. Spence, Ralph B., "A Comprehensive Testing Program for Elementary Schools," Teachers College Record (January, 1933), 34:279-284.

151. Stanley, Julian C., Jr., "The Use of Multi-Factor Aptitude Tests in School Counseling," Review of Educational Research (February, 1953), 23:11-32.

152. Stenquist, John L., "The Administration of a Program of Diagnosis and Remedial Instruction," Educational Diagnosis, The Thirty-fourth Yearbook of the National Society for the Study of Education. Public School Publishing Company, Bloomington, Illinois, 1935.

153. Stenquist, John L., "Devices for Testing," The Nation's Schools (September, 1937), 20:30-33.

154. Stoddard, George D., The Meaning of Intelligence. The Macmillan Company, New York, 1943.

155. Stodola, Quetin, How One School System Records and Interprets Test Scores: A Do-It-Yourself Kit for Teachers, Test Service Bulletin No. 89. World Book Company, Tarrytown-on-Hudson, New York, 1958.

156. Strang, Ruth, "Relationships Between Certain Aspects of Intelligence and Certain Aspects of Reading," Educational and Psychological Measurement (Winter, 1943), 3:355-360.

157. Stuit, Dewey B., Gerald C. Helmstadter, and Norman Frederiksen, Survey of College Evaluation Methods and Needs, A Report to the Carnegie Corporation, December, 1956. Educational Testing Service, Princeton, New Jersey, 1957.

158. Super, Donald E., Appraising Vocational Fitness. Harper and Brothers, New York, 1949.

159. Super, Donald E., The Place of Aptitude Testing in the Public Schools, Test Service Bulletin No. 49. World Book Company, Tarrytown-on-Hudson, New York, 1943.

160. Symonds, P. M., Diagnosing Personality and Conduct. D. Appleton-Century Company, New York, 1931.

161. Taba, Hilda, "The Function of Evaluation," Childhood Education (February, 1939), 15:245-251.

162. Taylor, Edward A., "Improving Questionnaires," California Journal of Educational Research (March, 1951), 2:73-76.

163. Thorndike, Edward L., "Tests and Their Uses," Teachers College Record (October, 1924), 26:93-94.

164. Thorndike, Robert L., and Elizabeth Hagen, Measurement and Evaluation in Psychology and Education. John Wiley and Sons, Inc., New York, 1955.

165. Thorpe, Louis P., Appraising Personality and Social Adjustment, Educational Bulletin No. 11. California Test Bureau, Los Angeles, 1953.

166. Thorpe, Louis P., Guiding Child and Adolescent Development in the Modern School, Educational Bulletin No. 16. California Test Bureau, Los Angeles, 1951.

167. Tiegs, Ernest W., Diagnosis in the Reading Program, Educational Bulletin No. 10. California Test Bureau, Los Angeles, 1951.

168. Tiegs, Ernest W., The Proper Use of Intelligence Tests, Educational Bulletin No. 4. California Test Bureau, Los Angeles, 1951.

169. Torbet, David P., "The Attitude of a Select Group of Colorado Secondary School Teachers Toward Informal Teacher-made Tests as Measured by a Projective Interview," Journal of Educational Research (May, 1957), 50:691-700.

170. Torgerson, Theodore L., and Georgia S. Adams, Measurement and Evaluation for the Secondary School Teacher. Dryden Press, New York, 1956.

171. Traxler, Arthur E., "Administering and Scoring the Objective Test," Chapter X in Educational Measurement, E. F. Lindquist (Editor). American Council on Education, Washington, D. C., 1951.

172. Traxler, Arthur E., "Evaluation of Aptitude and Achievement in a Guidance Program," Educational and Psychological Measurements (Spring, 1946), 6:3-16.

173. Traxler, Arthur E., The Nature and Use of Anecdotal Records, Educational Records Bureau Supplementary Bulletin D, Revised. Educational Records Bureau, New York, 1949.

174. Traxler, Arthur E., "Planning and Administering a Testing Program," School Review (April, 1940), 48:253-267.

175. Traxler, Arthur E., "Standardized Tests," National Educational Association Journal (November, 1959), 48:18-20.

176. Traxler, Arthur E., "The Status of Measurement and Appraisal Programs of Large City School Systems," 1953 Achievement Testing Program in Independent Schools and Supplementary Studies, Educational Records Bureau, Bulletin No. 61. Educational Records Bureau, New York, 1953.

177. Traxler, Arthur E., Techniques of Guidance. Harper and Brothers, New York, 1957.

178. Traxler, Arthur E., "Ten Essential Steps in a Testing Program," Education (February, 1959), 79:357-362.

179. Traxler, Arthur E., The Use of Tests and Rating Devices in the Appraisal of Personality, Educational Records Bulletin No. 23. Educational Records Bureau, New York, 1938.

180. Traxler, Arthur E., Robert Jacobs, Margaret Selover, and Agatha Townsend, Introduction to Testing and the Use of Test Results in Public Schools. Harper and Brothers, New York, 1953.

181. Wellman, Frank E., "Administration of the Testing Program," in Understanding Testing, United States Department of Health, Education, and Welfare, Office of Education. United States Government Printing Office, Washington, D. C., 1960.

182. Wertz, Henry, "Minimum Essentials for a Testing Program," American School Board Journal (September, 1957), 135:41-43.

183. Wilkinson, D. H., "Measurement and Evaluation: A Significant Area in Planning Programs of In-Service Education," The Bulletin of the National Association of Secondary School Principals (February, 1957), 41:89-94.

184. Womer, Frank B., "Initiating a Testing Program," Elementary School Journal (January, 1957), 57:193-197.

185. World Book Company, Planning the Testing Program-Testing Program Organizational Chart. World Book Company, Tarrytown-on-Hudson, New York.

186. World Book Company, Types of School Tests, Test Service Bulletin No. 19. World Book Company, Tarrytown-on-Hudson, New York.

187. Wrightstone, J. Wayne, "Design for Evaluation for High Schools," High Points (February, 1944), 26:27-32.

188. Wrightstone, J. Wayne, "Educational Measurements," in Review of Educational Research (June, 1956), 26:268-292.

189. Wrightstone, J. Wayne, "Evaluation," in Encyclopedia of Educational Research. The Macmillan Company, New York, 1950.

190. Wrightstone, J. Wayne, "Techniques of Appraisal," in Readings in General Education, Will S. Gray (Editor). American Council on Education, Washington, D. C., 1940.

191. Wrightstone, J. Wayne, Joseph Justman, and Irving Robbins, Evaluation in Modern Education. American Book Company, New York, 1956.

192. Zehrer, Frederick A., A Basic Testing Program for the High School, Test Service Bulletin No. 47. World Book Company, Tarrytown-on-Hudson, New York.