AN INSERVICE PROGRAM FOR TEACHERS IN THE EVALUATION OF MEASURES OF READING AND READING RELATED PERFORMANCE

THOMAS JOSEPH GRACE

copy 2

AN INSERVICE PROGRAM FOR TEACHERS IN THE EVALUATION OF MEASURES OF READING AND READING RELATED PERFORMANCE

An Internship Report

Presented to

the Faculty of Education

Memorial University of Newfoundland

by

Thomas Grace

In Partial Fulfillment

of the Requirements for the Degree

Master of Education

September 1975

ABSTRACT

Standardized reading tests have played a prominent role in the decision-making processes of our educational systems. Each year millions of students are assigned to various types of instructional programs on the basis of such test results and vast sums of money are expended to rectify or remediate reading problems identified through the use of standardized reading tests. Yet, despite their widespread use, it appears that the strengths and limitations of such tests are little understood by the educators who use them.

This inservice program was designed to provide teachers with a better understanding and appreciation of the concepts related to reading evaluation so that they would be better able to make wise use of standardized reading tests. As a basis for establishing this understanding, the program first examined the nature and functions of visual and auditory ability, phonics, structural analysis, contextual analysis, vocabulary, comprehension and oral reading as they relate to the act of reading. The program then undertook an examination of several appropriate techniques that might be used to evaluate these reading skills. Having analyzed the reading skills, the program proceeded to a consideration of the statistical concepts of validity, reliability, standard error of measurement, norms and test scores and the effects they have upon the strength of a particular test result.

With these aspects of measurement and evaluation serving as a

frame of reference, the program required teachers to analyze the strengths and weaknesses of a number of commonly used standardized reading tests. In this way the program helped teachers develop the ability to evaluate the worth of a standardized reading test in terms of the type and quality of the educational decision that was to be made. This analysis also enabled them to construct better informal reading tests.

A questionnaire designed to evaluate the effectiveness of the program was administered at its conclusion. The results of the questionnaire revealed that teachers did feel that they had gained a greater insight into the concepts of reading evaluation and that they were now much more capable of selecting a test that would enable them to make the type of decision required of them.

The inservice program was conducted through a series of twelve one-hour weekly sessions with seventeen Special Education teachers from schools under the jurisdiction of the St. John's Roman Catholic School Board.

ACKNOWLEDGEMENTS

The intern wishes to express his gratitude to the Director of Curriculum of the St. John's Roman Catholic School Board and to its Supervisor of Special Education for the assistance they gave him in the initial stages of the inservice program.

The intern also wishes to express his deep appreciation to the members of his supervising committee, Dr. Ethel Janes and Dr. Larry Miller, for the time and guidance which they extended to the intern throughout the course of the internship.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

One of the traditional characteristics of the educated individual has been his ability to cope effectively with the printed medium. Modern man has come to realize that much of his knowledge is conveyed to him within the confines of books and if he is to attain this knowledge, he must develop a facility for decoding and comprehending printed materials. The importance of being able to read has been further emphasized by the knowledge explosion of the twentieth century and the tremendous increase in the volume of printed publications.

Whereas in earlier times it was accepted that some people would not have to learn to read, today's society uses every form of persuasion to ensure that each child is taught how to read (Masters, 1973, 59-63). It may be said that of all the educational concerns of our society, the ability to read is viewed as the most essential.

In his discussions of Title I projects, Durost (1968, 291-304, 367) alludes to the enormous expenditures of money and effort that have been directed toward improving the level of literacy within the United States. He points out, however, that while this expenditure is both commendable and necessary, it brings with it the public demand for educational accountability. Educators are being required to defend their teaching practices and to prove that their methods are educationally sound by producing more students who possess a strong reading

ability.

The most frequently used technique for demonstrating that students are progressing along the continuum of reading proficiency is the standardized reading test. The test is administered at the beginning of the learning situation and again at the termination of instruction. The resulting comparison of scores is designed to illustrate that reading gains have been achieved.

Although the technique of utilizing standardized test results as evidence that students are progressing is widely employed, it is a method that is subject to serious criticism. Cox and Starrett (1970, 227-28) have stated that because of the comprehensive nature of most standardized tests, the test results may have little, if any, direct relevance to the instructional program of an individual student. Durost (1968, 291-304) feels that the gains displayed by students on standardized tests may well be within the range of chance occurrence and may represent little more than the variability inherent in the test itself. Other writers have also discussed the problems associated with the use of standardized tests as a means of assessing instructional gains. Sutton (1972, 528-32) maintains that, while accountability has helped to focus greater public attention on the issue of reading instruction, the task of devising valid accountability procedures is a formidable problem that educators have yet to overcome.

The most common purpose for which standardized reading tests are administered is the assessment of individual student reading

needs. Many teachers rely heavily upon such tests to provide them with the diagnostic information required to design appropriate instructional programs for their students. Since the decisions arising from the administration of these tests can have far-reaching effects upon the student's life, it is imperative that they accurately portray the student's reading strengths and weaknesses. Maginnes (1972, 750-53), however, refers to several evaluation factors that can severely limit the diagnostic utility of standardized reading tests and which, when not given sufficient consideration by a teacher, can lead to very erroneous interpretations of a student's reading ability. Others, such as Farr (1970, 52), state that, since there is almost no evidence to support the contention that such tests are valid measures of distinct and separate reading skills, there is very little justification for the use of standardized reading test results as the foundation for the development of an individualized diagnostic teaching program.

Although there are restrictions on ways in which standardized reading tests may be used, they, nevertheless, can provide the teacher with valuable reading information about the students. Rost (1973, 319-20) asserts that, with the aid of test publishers and curriculum specialists, teachers can obtain useful diagnostic information from a standardized test. While acknowledging the limitations of such tests, Wilson (1972, 87-88) illustrates the advantages of using a standardized reading test as a preliminary step in an in-depth evaluation of a reading program. These tests can also fulfill a valuable

function when the teacher is interested in the comparative performances of students.

Standardized reading tests have traditionally been employed by many teachers as a means of determining the grade placement of students. The grade equivalent score has been accepted as indicative of the student's instructional level. Sipay (1964, 265-68) and others have provided research that questions the use of the grade equivalent score as a basis for assigning students to reading levels and programs. Yet, the practice still appears to retain considerable favor among teachers.

It is evident from the preceding discussions that the standardized reading tests have both notable strengths and serious limitations. Consequently, their proper utilization is dependent upon the user's cognizance of these factors. Since these tests appear to be a permanent element of our modern educational system, benefits to be derived from their use can only be realized if teachers have a clear understanding of the purposes for which they can properly be administered.

BACKGROUND OF THE INTERNSHIP

The impetus for this internship project stemmed from the writer's experiences as a classroom teacher and a primary-elementary school principal. As part of his teaching responsibilities, the writer was required to administer standardized reading tests to

his students. At that time, the writer knew very little about the tests' validity, the nature of the scores they reported, the population of items from which they were constructed, the characteristics of their norming groups or their correlation with the specific reading program of the school. The tests were scored, the results were recorded, but very little practical use was ever made of the information obtained. Interpretations that were made were generally of a simplistic nature and sometimes very erroneous.

This lack of knowledge about standardized testing resulted from a teacher training that only incidentally dealt with the proper use and interpretation of tests and from a lack of adequate inservice training on the part of those curriculum specialists who advocated their use. The writer has no reason to believe that this was not the common experience of a majority of his teaching colleagues and that a very similar condition exists among teachers today. Most of the injustices that were, and still are, inflicted on students could have been prevented if the writer and his fellow teachers had been properly instructed regarding the advantages and disadvantages of standardized tests.

In his capacity as a school principal, the writer witnessed the same kind of confusion among the members of his staff. The School Board required teachers to administer a number of standardized reading tests to their students. Although the tests were administered and scored, the teachers held grave reservations about the nature of

the tests and the validity of the scores obtained. In practical terms, the testing usually constituted an interruption of the classroom schedule that yielded very few worthwhile results and only added to the teachers' clerical duties. In general, very few teachers had had sufficient training to understand the real nature of the tests, and fewer still were capable of interpreting the scores meaningfully and accurately.

Since he felt that the School Board would continue the practice of using standardized reading tests, it was apparent to the writer that efforts needed to be made to familiarize teachers more thoroughly with the values and limitations of such tests. Only with such knowledge can teachers extract from standardized tests the value which these tests can afford them in the making of decisions concerning the most effective type of program for their students.

In the initial stages of the internship, the writer contacted the Curriculum Coordinator of the St. John's Roman Catholic School Board and outlined a proposal for an inservice program that would deal with the evaluation of formal measures of reading performance. He further proposed that the inservice program be conducted with Special Education teachers since it appeared that these teachers were more frequently required to administer and use standardized reading tests than were the regular classroom teachers. The Curriculum Coordinator, feeling that the proposed program would be beneficial to this group of teachers, arranged another meeting which the Supervisor

of Special Education attended.

At this latter meeting, there was general agreement that the proposed inservice program should be conducted with the Special Education teachers. The writer was then invited to meet with the Special Education program steering committee in order to obtain some teacher reaction to the proposal. The program steering committee endorsed the concept of the inservice program and recommended that Special Education teachers participate in it.

Following the meeting with the Special Education program steering committee, a letter explaining the purpose and format of the proposed inservice program was sent to all the Special Education teachers within schools under the jurisdiction of the St. John's Roman Catholic School Board. (Appendix A contains a copy of the letter.) The teachers were invited to attend a preliminary meeting at which a more detailed description of the program would be given and at which they would be given an opportunity to ask any questions they had concerning it.

Of the forty-eight teachers to whom letters were sent, twenty-six attended the preliminary session. From this group, seventeen teachers signified their intention to participate in the internship program.

OVERVIEW OF THE REPORT

In Chapter I, the writer has presented a brief introduction to

the field of reading evaluation, along with an explanation of the background of the internship. Chapter II contains a rationale for conducting the inservice program and a review of the related literature. A statement of the goals established for the program, a description of the teachers, instruments used and the methodology employed in the inservice program are found in Chapter III. An evaluation by the teachers of the entire inservice program is provided in Chapter IV. Conclusions and recommendations arising from the program are presented in Chapter V.

CHAPTER II

REVIEW OF RELATED LITERATURE

PROFESSIONAL NEED FOR INSERVICE PROGRAMS IN THE USE AND EVALUATION OF STANDARDIZED TESTS

Standardized reading tests are a very significant element of the modern educational system and exert enormous influence upon the educational futures of students. Davies (1969, 545) estimates that approximately thirty-seven million achievement tests are administered each year in the United States. Livingston (1972, 402) writes that the fates of millions of children are decided on the basis of standardized tests and that in many instances the results of these tests become self-fulfilling prophecies because of the almost total trust and confidence that teachers place in them. One has only to look through Buros' Reading: Tests and Reviews to obtain some idea of the place of standardized reading tests in the present educational system. According to Carver (1972, 301), it will be a long time before such tests will be replaced by more instructionally effective tests of reading.

Because of the proliferation of standardized reading tests, one might be led to believe that they receive unanimous approval from reading specialists and teachers. While these tests do provide some useful services, their popularity seems to be founded more on their availability than on their inherent worth as evaluative instruments for

assessing the reading capacity of individual students. Rost (1973, 319) declares that among the criticisms levied against the standardized test by teachers, its lack of immediate relevance to the instructional needs of the student is paramount. Cox and Starrett (1970, 227) express a similar view when they state that many standardized achievement tests are too comprehensive in their coverage of the subject matter to be capable of furnishing the type of diagnostic information which the classroom teacher requires for the planning of instruction. Although Cox and Starrett made their comments with reference to content area tests, it is reasonable to assume that the global scores reported by many reading tests suffer from the same limitation. Other writers have pointed out that the items of standardized reading tests that differentiate best are those that are confounded by a high degree of abstract reasoning, a primary component of intelligence (Traxler, 1958, 111; Carver, 1972, 300). Ramsey (1972, 210) maintains that the scores obtained on standardized reading tests can be invalidated if the student is permitted or encouraged to guess answers. Sister Julitta (1958, 122) explains that such tests often produce results that reflect the student's frustrational reading level rather than the instructional level.

Although many writers have drawn attention to some of the more serious problems associated with the use of standardized reading tests, these tests are not without their advantages. The most frequently cited advantage is the ability of the standardized reading

test to provide a common basis upon which the reading performance of groups of students, of schools or of school systems may be compared (Mitchell, 1968, 164). Ebel (1961, 29-32) advocates the use of standardized reading tests as a method of identifying those students who may have a serious reading disability and who require a more in-depth diagnosis. This feature of the standardized test is reiterated by Harris (1970, 96).

As the literature demonstrates, the standardized reading test has both strengths and weaknesses and thereby the potential for effective or ineffective decision-making. The degree to which either of these potentialities is realized is dependent upon the testing expertise possessed by the teacher who uses the standardized reading test. Unfortunately, there is considerable scepticism among writers with respect to teachers' awareness of these strengths and weaknesses.

Barry et al. (1974, 590) have stated that "it appears that reading teachers know less about reading tests and their use than almost any other 'tool of the trade'". In the Preface to the Sixth Mental Measurement Yearbook, Buros (1965, xxiii) says that teachers do not appear to be particularly concerned by the fact that so many of the standardized tests are severely criticized as having little or no validity for the task they purport to do. He goes on to state that "teachers have the utmost faith in their own particular choice and use of tests regardless of the absence of supporting research or even of the presence of negating research". In discussing the state of reading

evaluation in North America, Hill (1974, 13) affirms that no substantial improvement in reading evaluation can be anticipated until teachers and reading specialists come to understand better the concepts of measurement and evaluation, both in theory and in practice.

There appears to be a very serious need for teachers to become involved in inservice training that will enable them to make more profitable use of standardized reading tests. Such training might prevent them from making decisions that are based on very questionable premises.

Ruddell (1974, vii) has written that the era of the 1970's has become the "Accountability Era". More and more the general public, and the local and federal governments in particular, are becoming more cognizant of the cost of education. As a consequence of this increased awareness, greater pressure is placed on the teacher to demonstrate the effectiveness of his instruction.

According to Fitzgibbons (1974, 29), "test results are the currency of accountability". This reality makes it imperative that teachers fully understand the characteristics of standardized tests so that they can use them wisely and provide the various interest groups with valid measures of how effective their teaching methods and programs are. Unless teachers become knowledgeable with respect to reading assessment, Farr and Brown (1971, 346) feel that teachers could, as a result of the current emphasis upon accountability, be pressured into adopting evaluative techniques that would be

inappropriate for demonstrating that students are making substantial progress in reading ability.

Associated with and growing out of educational accountability is the phenomenon of performance contracting. Commercial firms sign contracts with a school or school system to raise the reading achievement of individual students by a specified number of score units, usually in terms of grade equivalent score points (Hogan, 1974, 63). The yardstick for determining the degree of reading gain is most frequently the standardized reading test (Stake and Wardrop, 1971, 324). Stake and Wardrop maintain that by pure chance alone a substantial proportion of the students could show reading gains as a result of the variability of the test itself. They further state that reading experts have agreed that standardized reading tests do not have sufficient content validity to be useful in assessing individual student reading achievement. In commenting upon performance contracting, Goodman (1971, 365) states that much of what transpires under the title of performance contracting is little more than the duping of unsophisticated educators. In order for teachers to properly assess the claims of performance contractors, they must be informed regarding the use and interpretation of standardized tests (Durost, 1971, 304).

Teachers have also discovered that within the profession itself there is considerable importance attached to reading evaluation and the use of standardized reading tests. Burnett (1963, 229-34) has made the statement that, of the many ideas that flow in and out of

education, the concept of the teacher as one skilled in classroom diagnosis is one that has become a constant. The gravity of this responsibility is illustrated by Farr and Brown (1971, 341) when they state that it is within the classroom that the most important decisions regarding instruction are made. In a similar vein, Hieronymous (1972, 267) writes that the most significant facet of teacher accountability is the responsibility the teacher has to his student to ensure that the student receives the type of teaching that is most suited to his individual needs. If Durost's statement (1971, 290) that "the quality of instruction is clearly related to the quality of the evaluative effort made" is correct, then the teacher has a very grave obligation to become fully aware of the possibilities and limitations of standardized reading tests. The teacher must, therefore, have the facility to determine whether or not a test will be valuable in assisting him to identify important edumetric needs (Carver, 1972, 302).

AREAS OF PRIMARY CONCERN FOR THE EVALUATION AND USE OF STANDARDIZED READING TESTS

Knowledge of Basic Reading Skills

The evaluation of standardized reading tests must be founded upon an adequate conceptualization of the skills involved in the process of reading. Durkin (1974, 495) confirms this by declaring that the diagnostic proficiency of the teacher rests upon a clear understanding of the skills that constitute the reading act. In his book on methods of

diagnostic teaching, Wallen (1967) pursues a format which first exposes the teacher to the functions and operations of a skill before requiring him to construct an informal test of that skill. It is essential that the teacher comprehend the basic characteristics of the various reading skills before he attempts to pass judgment regarding the value of a test that proposes to measure these skills.

In addition to understanding the nature of the reading skills, the teacher must realize the relevance of each skill to the total development of a mature reading capacity. While there is not total agreement among reading specialists as to the importance of each skill, it is possible to single out a number of skills that have been given serious attention by a majority of reading authorities.

Spache and Spache (1973, 8-14) maintain that visual acuity and discrimination are prerequisite skills for the development of reading ability. Bond and Tinker (1973, 113-15) quote several studies that substantiate the importance of visual acuity and discrimination to the reading act.

Auditory discrimination has been shown by Durkin (1974, 312) and by Heilman (1972, 121) to be essential to the acquisition of a good phonics ability. Both writers state that the student must be capable of auditorily differentiating one sound from another if they are to be able to associate one sound with a particular set of graphic symbols.

The significance of word recognition skills to the development of reading proficiency has been explained by Wiener and Cromer

(1967, 620-42). Wilson (1972, 203) has divided the word recognition skills into phonics, structural analysis and context abilities. A similar division of word recognition skills has been proposed by Spache and Spache (1973, 448-501). With reference to word recognition skills, Bond and Tinker (1973, 304) state that the importance of sight vocabulary to reading proficiency cannot be too strongly emphasized because of its role in the identification of new words and in comprehension.

The topic of comprehension has evoked more discussion among reading experts than has any other reading skill. Its prominence within the list of reading skills has been well documented (Bond and Tinker, 1973, 357; Spache and Spache, 1973, 543-69; Durkin, 1973, 393). The position of meaning vocabularies--speaking, listening and reading--in the growth of reading ability has been affirmed by such writers as Spache and Spache (1973, 511-35), Durkin (1973, 353-91) and Heilman (1972, 447-99).

While oral reading is generally not viewed as a skill in the process of learning to read, it is stressed as a very effective means of diagnosing a student's reading strengths and weaknesses (Letton, 1958, 77). Powell (1971, 89) states that the real value of oral reading is that it affords the teacher the opportunity of evaluating the student's reading behaviour in depth. Durkin (1973, 499) explains how oral reading may also be used to obtain an estimate of the student's independent, instructional and frustrational reading levels.

Considerations Pertaining to the Nature of Standardized Reading Tests

Among the many evaluative decisions which an educator may have to make with respect to the use of a specific test, the most crucial is that of relating the type of test to be used to the type and quality of the educational decision to be resolved (Davies, 1969, 545-47). Heironymous (1972, 264-67) writes that the measurement value of a test is proportional to the degree to which it provides results that facilitate worthwhile decisions concerning the instructional program of the student. While writers agree on the necessity of defining specific purposes for the use of the test and of coordinating these with the stated aims of a test, Anastasi (1966, 3) declares that it is this very consideration that is lacking in many of the testing programs presently operating in schools.

The evaluation of a standardized reading test should incorporate an examination of the subtests to determine exactly what skills are being assessed by the test. Livingston (1972, 402-10) argues that some tests may be measuring aspects of reading other than those which it purports to measure. In this respect, Anastasi (1966, 9) advises teachers against relying upon the test title as evidence of what skill or skills are being assessed.

Educators must also concern themselves with the technique by which a test attempts to measure a skill (Hayward, 1968, 554). A search of the literature revealed very little discussion of the

methodology that should be used to evaluate specific reading skills. Most of the literature related to the methodology of evaluation dealt with the various methods that have been advocated for the measurement of comprehension (Bormuth, 1974, 241-56; King, 1968, 206-13; Kingston, 1970, 230-36; Rankin, Jr., 1970, 237-53). Ramsey (1972, 210) recommends that wherever possible the evaluation of a skill should take place in a manner similar to the way in which the skill is used while the student reads.

While the teacher should be familiar with the purposes, content and methodology of a test, it is also vital that he display a facility for evaluating the statistical qualities of a standardized test (Mour, 1968, 167-71). A firm understanding of these aspects of the test enables the teacher to interpret the results so that only those inferences and conclusions appropriate to the test are made. Among the statistical components of standardized tests which should be studied, Mitchell (1968, 164) considers validity, reliability and test scores the more important.

In arriving at a decision regarding the use or evaluation of a test, Mehrens and Lehmann (1973, 124) and Harnersma (1972, 57) have identified the test's validity as the starting point for the analysis of the statistical aspects of any standardized test. Unless the test possesses an adequate degree of validity, the results obtained from its administration can have little relevance to the purpose for which the test is given.

As a method of determining the content validity of a test, Farr (1970, 50) advises the teacher to compare the items of the test with the objectives of the school's reading program. Cronbach (1969, 23-24) notes that content validity can fluctuate because test items are linked to social conditions and events. With the passage of time, the items of a test may no longer be entirely representative of the universe of behaviours which they are designed to measure. The implication of Cronbach's statement is that the publication date of a test should be checked since tests that were devised many years previously may not be suitable for some current situations.

The content validity of a test may also be affected by the size of the item sample from which the final items were selected and by the manner in which these items were chosen. Mitchell (1968, 165-66) agrues against the validity of meaning vocabulary tests that attempt to estimate the student's depth of vocabulary on the basis of less than thirty words or with items that are not suited to the experiential background of the student.

In order to assess the content validity of a test, the educator must consider all of these factors and weigh his findings in light of the type of decision that will be made as a consequence of administering the test. An adequate critique of a test's validity requires much more than a cursory reading of the test subheadings.

The utility of a test depends as much upon the reliability of the test as it does upon its validity (Ebel, 1961, 30). In order to use

the results of a test, the examiner must have an assurance that the scores obtained will not vary dramatically from one testing occasion to another. The extent to which a test will yield a reliable measurement of a skill is a function of several factors.

Since most tests are but a sample of all the possible behaviours associated with a trait, the number of items comprising a test will have a significant bearing upon the stability of the score (Mour, 1968, 170). Anastasi (1966, 6-7) has asserted that the reliability of a test can be strengthened by increasing the number of items which measure a specific skill. It is the responsibility of the test user to ascertain, in the light of the total item population, whether or not the number of items in a test represents a limitation to the intended use of the scores.

Wesman (1968, 197) states that the range of ability in the group on whom the reliability coefficient is based is a significant consideration in establishing the level of reliability for a test. He further explains that, while a very heterogeneous ability group can furnish a high coefficient, the reliability of the test when used with a very homogeneous ability group can be substantially lower than that which is reported in the test manual.

The statistical technique used to calculate the reliability coefficient will influence the level of reliability that will be obtained for a test. A comprehensive discussion of the strengths and limitations of the various reliability formulae is presented by Thorndike and

Hagen (1969, 183-86). The decision as to which form of reliability should be sought by the teacher will be dictated by the use that will be made of the test results. Mehrens and Lehmann (1973, 116) state that the time restrictions of a test will also affect the consistency of the test and they argue that the use of internal measures of reliability with speeded tests will produce spuriously high coefficients. For such tests an estimate of stability would be a more logical method of determining the reliability.

The teacher is also faced with the task of selecting an appropriate level of reliability to demand of a test. Hamersma (1972, 58) has suggested that tests with reliability levels above .80 should be sought. Because of the various conditions that can affect the derivation of a reliability coefficient, Hamersma's recommendation may not prove to be necessary or feasible under a given set of circumstances. A more reasonable formula is offered by Wesman (1968, 195) when he writes that the level of reliability to be desired should be determined by the nature of the decisions to be made.

While the reliability coefficient is an important consideration for the teacher, he must also recognize that even with a highly reliable test an individual student's score will contain some degree of chance error (Hamersma, 1972, 119). Because of this, the teacher must possess an understanding of the use and interpretation of the standard error of measurement statistic (Mehrens and Lehmann, 1969, 34-35).

It has been previously stated that the need to obtain an indication of a student's ability in a discipline is the prime purpose for which a test is administered. In most test situations, however, the raw score that is registered does not convey any instructional meaning in and of itself (Hagan, 1961, 11). To overcome this drawback of testing, test publishers have developed sets of norms that permit the teacher to translate the student's raw score into a normed score. The normed score enables the teacher to make comparisons between the performance of the individual student and that of the standardization group (Massad, 1972, 287).

In order for the norms of a test to form a proper basis for comparing the performance of the student to that of the norm population, Mehrens and Lehmann (1969, 51) contend that the characteristics of the norm population and those of the student must be very similar. Any significant variation in the nature of either group would invalidate any comparisons that would be made. In another place, Mehrens and Lehmann (1973, 142) explain that the test norms should be recent. Norms that are not current could, depending upon the discipline the test measures, no longer be representative of the present population of people with whom the test might be used. Massad (1972, 292) maintains that norms are only useful insofar as they legitimately contribute to the decision-making situation.

The process of interpreting test results is a very delicate issue and one that can provoke strong opposition from various sectors

of the general public. Fitzgibbons (1974, 15-32) cites several instances in which the use of tests and test results created very serious problems for educators and school systems. His illustrations emphasize the responsibility that professional educators have to ensure that their interpretations of test scores rest upon a thorough knowledge of the attributes and uses of the different types of scores which tests report.

Although there are many methods by which test results may be reported, some of the more frequently used scores are the grade equivalent, the percentile rank and the stanine (Mehrens and Lehmann, 1973, 149).

Of all the scores reported to teachers, administrators, and parents, the most widely used and most frequently misinterpreted is the grade equivalent score (Hagan, 1961, 13). For the majority of these people, the grade equivalent score is viewed as indicative of the program grade level at which the student can function successfully (Massad, 1972, 290). This score, however, bears no direct relationship to the program level or to difficulty level of a basal reader (Carver, 1972, 300). Instead, its meaning is fixed in the norm population and simply signifies the median score obtained by a particular grade of students on a specific test at a definite point in time (Mehrens and Lehmann, 1969, 55). Although many writers have expressed grave reservations about the reporting of test results in terms of the grade equivalent score, test publishers continue to

include it among the scores a test reports. Very often there is no explanation in the test manual regarding the exact interpretation of the grade equivalent score. Its very misinterpretation seems to be the reason that so many educators use it.

Because of the misconceptions associated with the grade equivalent score and because it does not possess an equivalence of meaning across different tests, Hagan (1961, 14) suggests the use of the percentile rank score in preference to the grade equivalent score. Because of its fixed reference points of zero and one hundred, the percentile rank score makes it possible for the teacher to make an immediate, meaningful statement regarding the student's performance in relation to the norming group of the test. The teacher knows from the percentile rank score whether or not the student is above or below the average individual of the norming population. This type of conclusion cannot be made by looking at the grade equivalent score alone.

In stipulating standards for the construction of standardized tests, the American Psychological Association (1966, 34) considers the inclusion of a percentile rank set of scores as a "desired" quality of a test. Hamersma (1972, 52) states that, while the percentile rank score may sometimes be confused with the percentage of correct responses, its understanding does not involve an intricate explanation and can be explained quite simply.

While arguing the use of the percentile rank score in preference to the grade equivalent score, Hagan (1961, 15) also draws

attention to the fact that the percentile rank score is limited in that there is an uneven relationship between the number of raw score points that separate one percentile rank score from another.

The third type of score which the test user will often encounter is the stanine. Durost (1968, 125-36) explains the development and use of the stanine and states that its simplicity has led to its widespread use in standardized testing. Rost (1973, 20) asserts that one of the positive features of the stanine is that it represents a score band rather than a specific numerical score. This characteristic of the stanine helps to counteract the precision which teachers associate with exact scores and it also makes allowances for the problem of measurement error embodied in each test score.

Although the stanine enjoys some advantages over other types of scoring systems, it is subject to the same limitations as is the percentile rank score. Since it is related to the percentile rank system, the raw score differences between each stanine are equal to one another (Massad, 1972, 291). Another possible limitation of the stanine is that it is often described as an equal interval scale (Durost, 1968, 127). Educators and others may erroneously interpret this to mean that the performance difference between stanines one and two is the same as that between stanines five and six. The equal interval description, however, refers to the statistical construction of the scale itself and not to the levels of performance portrayed by each stanine. The discussion of the stanine in some standardized test

manuals (SDRT Manual, 1966, 13; Durrell Listening-Reading Series Manual, 1970, 26) does not seem to make this point very clear and teachers could be led to misinterpret the meaning of the stanine score.

INFORMAL READING INVENTORIES AND STANDARDIZED READING TESTS

In discussions of reading evaluation, one will often encounter the controversy surrounding the use of either the standardized reading test or the informal reading inventory as a means of diagnosing a student's reading ability. Mitchell (1968, 164) has criticised the diagnostic value of the standardized test on the grounds that it often lacks a sufficient number of test items on a specific skill to give a reliable measure of the student's true ability. He further points out that the extent of the student's strengths and weaknesses is not revealed by tests which have strict time limits. Wyatt (1968, 194) remarks that while two students may attain the same total reading score, there can be very salient differences in the skill needs of each student. Chall (1970, 53) maintains that standardized tests designed to cover a number of grade levels will frequently produce a distorted view of reading achievement; this is especially true with reference to students who are at both extremes of reading ability. The standardized reading test is generally too broad in its coverage of skills to provide an in-depth analysis of any one skill.

While acknowledging the criticism that global scores on

standardized tests may not afford assistance in formulating the day-to-day instructional program of an individual student, several writers argue that much valuable diagnostic information can be obtained from standardized tests. Ladd (1971, 305-11), Rost (1973, 319-20) and Cox and Sterrett (1970, 227-28) have suggested various methods for deriving this information. In each of their methods, the common element is the item analysis of the student's responses. This analysis, however, will only prove to be worthwhile when the test contains a truly representative sample of the skill behaviours and when there is a high degree of parallelism between the items of the test and the instructional objectives of the student's program. Such an analysis is likely also to be very time consuming. Chall (1970, 56) and Traxler (1958, 115) contend that the diagnostic value of the standardized reading test is realized when its results are examined in conjunction with other forms of reading assessment.

In comparing the standardized reading test to the informal reading inventory, Davies (1969, 548) criticises the standardized test because it overestimates the instructional reading level of most students. The high performance situation created by a standardized test, coupled with the guessing which occurs, tends to inflate the student's score. Kender (1968, 337-42) refers to separate studies by Betts and Killgallon which also indicated that the standardized reading test was less accurate than the informal reading inventory in establishing the instructional level of students. McCracken (1972, 273-77) conducted

a study with fifty-six grade six students using the Iowa Test of Basic Skills, Form I and an informal reading inventory in an effort to determine which provided the better indication of the students' instructional level. His findings revealed that the Iowa Test of Basic Skills overestimated the instructional reading level by an average of two grade levels. In an earlier study involving two hundred and two grade four students, Sipay (1964, 265-68), using three well-known standardized reading tests and an informal reading inventory with criterion levels of 96% word recognition and 60% comprehension, discovered that the three standardized reading tests significantly overestimated the instructional reading level of the students. When criterion levels of 90% word recognition and 60% comprehension were used, the standardized tests still overestimated the instructional level but not significantly.

In contrast to the above studies, Burgett and Glaser (1973, 71-74) quote the results of Burgett's doctoral research which showed that the Gates-MacGinitie Reading Test, Survey E, in particular the vocabulary subtest, could be used to establish a fairly accurate instructional level for a student. Burgett's study used the same criterion levels in his informal reading inventory as Sipay had used in his.

The review of the literature has shown that standardized testing plays a very prominent role in deciding the future of many facets of our educational system. Students are assigned to programs on the basis of test scores; existing programs are discontinued and new ones initiated because of test results; and public monies are allocated to

specific areas on the premise that a testing program can indicate where expenditures are most needed. The responsibilities associated with these kinds of decisions are extremely grave and can only be properly executed when educators are fully informed regarding the abilities to be measured and the nature of the evaluative instruments that are to be employed. This set of conditions, unfortunately, does not always prevail. Changes are being affected by educators who do not have the expertise to evaluate the foundations upon which their decisions rest.

The administration of standardized reading tests is a very costly, time-consuming undertaking that frequently results in far-reaching educational, emotional and social consequences. If the expenditures of time, money and human resources are to be justified, and if the decisions that result from such testing are to be valid and beneficial to society, the educator is obligated to make every effort to become professionally knowledgeable about the many aspects of standardized reading tests.

CHAPTER III

METHODOLOGY

OBJECTIVES OF THE INTERNSHIP

In undertaking this inservice program, the primary aim of the intern was the provision of information regarding reading evaluation that would enable teachers to critically analyze a standardized reading test so as to ascertain its worth in terms of the goals set for the administration of the test. In Chapter I it is maintained that standardized reading tests are often improperly used and that teachers are not adequately trained in the selection and use of such tests. Through this inservice program, the intern endeavored to improve teachers' understanding of standardized testing. The various aspects and functions of reading skills and reading tests were examined so that teachers would become more cognizant of the factors that constitute a good test of reading ability and of the necessity of relating the functions of a particular standardized reading test to the purposes for which testing is being done. As a result of this, the teachers would be able to select and use standardized reading tests more effectively and would be able to render accurate decisions regarding a student's reading capacity.

As a consequence of their participation in the analysis of standardized reading tests, the teachers would be able to improve their own informal tests. The examination of the advantages and

disadvantages of standardized reading tests was intended to provide the teachers with an insight into the elements that should be incorporated into any reading test, formal or informal. The inservice program was also designed to assist teachers in making more effective use of test scores given to them by others.

As a result of his previous teaching and administrative experience, the intern felt that most teachers were only familiar with a very limited range of standardized reading tests. By including several such tests in the program, the intern attempted to broaden the teachers' knowledge of the types of formal tests that are available to them. The intern's estimation of the teachers' familiarity with different standardized reading tests was confirmed by the results of the brief questionnaire (See Appendix B) given to the teachers at the first session of the inservice program.

A tertiary aim of the inservice program was that it might serve as a base from which the teachers could become involved in the development of a battery of informal reading tests that might be used in all Special Education classes. The inservice program, however, did not include this activity. The development of such a battery would be begun at a future date if the teachers wished to meet for that purpose.

The specific objectives of the inservice program were:

1. to enable teachers to analyze and evaluate standardized reading tests

2. to help teachers select the most appropriate test for their purposes

3. to assist teachers in the interpretation and use of test results

4. to provide teachers with an understanding of reading evaluation that would help them devise valid informal reading tests

5. to expose teachers to a number of standardized reading tests.

SUBJECTS

The internship was conducted with Special Education teachers from schools under the jurisdiction of the St. John's Roman Catholic School Board. The School Board serves forty-four schools and has a student population in excess of twenty-two thousand students. The area of Special Education comes under the guidance of the Board's Special Education Supervisor.

Special Education classes are located in all of the schools of the St. John's Roman Catholic School Board and at all the educational levels. The teachers who participated in the internship were working with students of primary, elementary and high school ages and were teaching in widely separated geographic areas of the school district. The majority of these teachers have had more than five years' teaching experience, with an average of just over two years' experience in Special Education.

At the first session of the internship, the teachers were requested to complete a short questionnaire (See Appendix B). The information obtained from the questionnaire is presented in Table I. In informal conversations, the teachers indicated that the subject of

Table I

Results of Teacher Information Questionnaire

N = 17 teachers	
Teachers who have completed three or more reading courses	35%
Teachers who have completed one or two reading courses	58%
Teachers who have not completed any reading course	7%
Teachers who have completed three remedial reading courses	7%
Teachers who have completed one or two remedial reading courses	58%
Teachers who have not completed any remedial reading course	35%
Teachers who have completed two measurement and evaluation courses	7%
Teachers who have completed one measurement and evaluation course	29%
Teachers who have not completed any measurement and evaluation course	64%
Teachers who have administered only the Gates-MacGinitie Reading Survey Test	50%
Teachers who have administered standardized reading tests other than the Gates-MacGinitie Reading Survey	15%
Teachers who have not administered any standardized reading test	35%
Teachers who have been involved in the development of a battery of informal reading tests	0%

reading evaluation was an area of great concern to them and that they did not feel secure in using standardized reading tests because they were not able to estimate the relative values of such tests in terms of the types of program decisions they had to make.

INSTRUMENTS AND MATERIALS

The materials used to conduct the inservice program consisted of xeroxed notes prepared by the intern and specimen copies of various standardized reading tests. The notes, which provided an explanation of the reading skills and statistical concepts in the inservice program, were based on reading and evaluation research. A copy of the notes can be found in Appendix E.

In addition to the notes, each teacher received a copy of each of the following tests:

Stanford Diagnostic Reading Test, Level I
" " " " Level II
Durrell Listening-Reading Series, Level I
" " " " Level II
Gates McKillop Reading Diagnostic Test
Gates-MacGinitie Reading Survey, Form D
Doren Diagnostic Reading Tests
Gray Oral Reading Tests
Silent Reading Diagnostic Tests
New Developmental Reading Tests

(See Appendix D for a more complete description of each test).

The tests were selected by the intern to illustrate various aspects of reading evaluation and to acquaint the teachers with a wider variety of standardized reading tests than they had previously known.

No effort was made by the intern to evaluate each test thoroughly. Rather, subtests were selected from the various tests for analysis, each subtest illustrating the evaluation of an aspect of standardized reading tests.

In addition to the above list of tests, the teachers also discussed and examined the operation of the Keystone Telebinocular and the Maico audiometer. An informal reading inventory was also used in conjunction with the analysis of the oral reading sections of the Gates McKillop and Gray tests. The IRI was also discussed in relation to the defining of the student's instructional reading level.

PROCEDURES

As previously stated in Chapter I, a letter was sent to each Special Education teacher within the schools of the St. John's Roman Catholic School Board explaining the aims of the inservice program and inviting the teachers to a preliminary inservice meeting. At this meeting, the teachers were given a more detailed explanation of the manner in which the program would be conducted. Any questions which the teachers had concerning the inservice program were answered by the intern.

The intern explained to the teachers that the program was designed to assist them in the analysis, selection and use of standardized reading tests. Several of the teachers were interested in the construction of their own informal reading tests and asked if the

inservice program would involve them in the construction of reading tests. It was pointed out to these teachers that, while the stated objectives of the program did not incorporate this activity, much of the material covered in the inservice program could be used by the teacher to design his own reading tests. Teachers who wished to become involved in the development of reading tests were assured that the services of the intern would be made available to them after the completion of the inservice program. The principal activity of the program, however, would entail the evaluation of formal measures of reading and reading related performances.

Because teachers attending the inservice program would be travelling from widely separated geographic areas, it was decided that the reading clinic at Memorial University would be the most central location for conducting the inservice meetings. The reading clinic was also a convenient location for the intern since some of the instruments used in the program, such as the Keystone Telebinocular and the Maico audiometer, were available there and could be used by the teachers. At the preliminary meeting the teachers were told that the program would run for a period of twelve consecutive weeks with one one-hour session each week. The sessions would be held in the afternoons after the regular school day. This meant that the teachers would be participating in the inservice program after having taught a full day. No other time period could be found when all the teachers were free to participate in an inservice program.

When the teachers met at the first formal session of the program, it was agreed that a decision would be made at each session regarding the day and the hour of the following week's meeting. This was designed to accommodate the timetables of the teachers who from time to time had to attend staff meetings or other professional functions. Of the twenty-six teachers who attended the preliminary meeting, seventeen elected to participate in the inservice program.

Since the aim of the program was to assist teachers in making more effective use of formal reading tests, the intern established a pattern for the sessions that would first involve a discussion of a major reading skill or measurement concept and then an analysis of that skill or concept as it appeared in a standardized reading test. To facilitate the discussions, the teachers were supplied with a set of notes related to the reading skills and measurement concepts. These notes contained such things as a simple definition of the skill, the functions of the skill during the act of reading, some appropriate techniques for evaluating the skill, problems that arise in standardized measures of the skill, and a series of questions which the teacher might use to evaluate a standardized measurement of the skill. A copy of the inservice notes is contained in Appendix E.

The teachers received the notes pertaining to a specific topic at least one week in advance of the session at which the topic was discussed. The more significant sections of the notes were highlighted by the intern and the teachers were asked to take special notice of

these in preparation for the ensuing discussions. The teachers were also requested to compare the explanations found in the notes with their previous knowledge of the topic and with their teaching experiences and to raise questions if there appeared to be any discrepancy between the notes and their previous knowledge or their teaching experiences.

After that portion of the weekly session dealing with the discussion of a specific skill was completed, the teachers were given a specimen test which contained a test of that skill. A number of questions were posed by the intern concerning the particular test and these formed the basis for the teachers' evaluation of the subtest. When time did not permit a full analysis of a subtest, the teachers were asked to take the questions home with them and to develop their own answers in time for the next session. At the next session each of the teachers was heard from and their answers were summarized. On two occasions, the intern requested the teachers to pass in a written evaluation of a test. The purpose of this exercise was to give the intern an indication of how well the teachers were able to use the notes to evaluate standardized tests.

The evaluation of subtests at the conclusion of a session also provided the intern with an opportunity to assist those who were finding it difficult to evaluate the tests. This oral evaluation produced a great deal of interaction among the teachers; the intern often had only to ask a few questions to keep the discussion moving in the

correct direction.

The topics of validity, reliability and test norms were not dealt with in an exhaustive manner since the majority of the teachers had not completed any course work in the area of tests and measurements. Some of the more salient features of these topics were examined and the teachers were instructed to read carefully the test manuals to ensure that complete information was provided with respect to the test's validity, reliability and norms. The inservice notes also contained additional material which the teachers could use at a later date when they would have more training in evaluation techniques.

The treatment of test scores involved the explanation, demonstration and application of the standard error of measurement. Discussion of this topic related to the chance error contained in a student's score, but more particularly to the use of cut-off points for the assigning of students to various school programs. The intern chose an arbitrary score as a cut-off point and then gave the teachers several student scores and asked them to assign the students to a reading program on the basis of one, two or three standard errors of measurement. As a result of these exercises, the teachers were made much more aware of the imprecision of individual test scores and the dangers of utilizing them as exact measurements of reading ability. The inservice notes also contained an explanation of the relationship of the standard error of measurement to reliability.

That portion of the inservice program related to the analysis

of test scores examined the strengths and limitations of the more commonly reported scores. Since most teachers receive only the student's total score on a test and have no chance to examine the individual items or responses, the teachers were presented with a grade equivalent, a percentile rank and a stanine score for the same raw score and asked to draw conclusions about the student's reading performance. This exercise quickly demonstrated the necessity for information concerning the nature of the test in order to make any kind of an accurate interpretation of the score. There was also discussion regarding the use of different scores for different types of audiences. This discussion entailed a close examination of the problems and advantages associated with the various types of normed scores.

Since it was the hope of the intern that the teachers would continue to analyze tests after the conclusion of the inservice program, the specimen copies of the tests were left with them. The intern also felt that this would be beneficial to the teachers since only 15 percent of them had been familiar with standardized reading tests other than the Gates-MacGinitie Reading Survey tests.

CHAPTER IV

EVALUATION OF INTERNSHIP

Upon the completion of the inservice program, the teachers were requested to answer a short questionnaire designed to elicit their opinions concerning the practicability and utility of the program. A five point scale--from strongly agree to strongly disagree--was used, with some statements being written in the negative so as to counteract a possible response-set. At the end of the questionnaire, space was provided for teachers to comment informally concerning the inservice program. A copy of the evaluation questionnaire is contained in Appendix C. Table II presents the responses of the teachers in terms of the percentages for each response category.

The results of the questionnaire revealed that the teachers felt that the program had been beneficial to them and that it did achieve its major objective of enabling them to evaluate reading and reading related tests. Although the program had been conducted after regular school hours and was fairly theoretical, the responses of the teachers were very positive.

Further evidence of the beneficial nature of the inservice program was found in the comments which teachers made at the end of the questionnaire. Several indicated that the program had given them a much better appreciation of standardized tests. One teacher stated that a similar program should be recommended for all teachers

because of the knowledge void that exists among teachers and administrators concerning the proper selection and use of standardized reading tests. Another teacher wrote that the inservice program had been "thoroughly enjoyable and instructional". Two teachers suggested that, while they found the program to be very helpful to them, they would have preferred to have had the program begin at the beginning of the new school year so that there would have been time for extended follow-up activities.

During the course of the inservice program, the intern also asked the teachers to voice any criticisms which they had concerning what was being done. At that time, the teachers signified that they were satisfied with the manner in which the program was being conducted.

Table II

Evaluation Questionnaire Results

Items	Percentage of Responses				
	SA	A	UD	DA	SDA
1. The objectives of the inservice program were clearly stated at the first session	59.0	41.0	0.0	0.0	0.0
2. The program achieved the objectives stated at the first session	33.0	66.0	0.0	0.0	0.0
3. The materials used in the program were not relevant to the objectives of the program	0.0	0.0	0.0	25.0	75.0
4. In view of the objectives of the program the topics dealt with were adequately explained	59.0	41.0	0.0	0.0	0.0
5. The individual sessions were not well organized	0.0	0.0	0.0	41.0	59.0
6. The materials and ideas presented in this program will benefit me in my future work as a teacher	47.0	53.0	0.0	0.0	0.0
7. The program provided sufficient guidelines to permit me to evaluate the worth of a standardized reading test	66.0	33.0	0.0	0.0	0.0

Table II (continued)

Items	Percentage of Responses				
	SA	A	UD	DA	SDA
8. As a result of the inservice program, I feel that I will be able to choose reading tests more wisely	71.0	29.0	0.0	0.0	0.0
9. The inservice program provided me with a better understanding of reading evaluation than I had prior to the program	75.0	25.0	0.0	0.0	0.0
10. If a similar inservice program were offered again, I would recommend other teachers to participate in it	82.0	18.0	0.0	0.0	0.0

SA = Strongly Agree
A = Agree
UD = Undecided
DA = Disagree
SDA = Strongly Disagree

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

In the opinion of the intern, the inservice program was a success. Besides their responses to the program evaluation questionnaire, there was other evidence during the course of the inservice sessions that indicated the teachers were deriving benefit from the program. On several occasions teachers remained behind after the conclusion of a session to further discuss elements of the program that had immediate relevance to their own testing procedures. These informal conversations revealed that the teachers were applying the concepts and ideas of the program to the construction of their classroom reading tests and that they were feeling much more secure concerning the validity of the results received. After the fifth session, it appeared that the teachers were beginning to lose the almost blind faith they had had in standardized reading tests and to become more confident in their ability to examine such tests. They had come to realize that even a highly recommended reading test will have its limitations.

The inservice program also convinced the teachers that one set of test results was generally not sufficient grounds for determining the reading potential of a student. They were aware that, because of

the reading skills measured, the testing methodologies employed, the item sample involved and the type of scores reported, two separate reading tests could yield fairly dissimilar student reading profiles. When several standardized tests had been reviewed, the teachers stated that they would be very hesitant to base their judgement of a student's reading ability on the results of one test. At the conclusion of the inservice program, the intern felt confident that these teachers would treat scores with the respect they warrant and that they would not likely make sweeping interpretations and generalizations regarding a student's reading ability as assessed by a standardized reading test.

While the teachers appeared to have lost the awe which they formerly had for standardized reading tests, they did not reject their use. They acknowledged that such tests did have some important purposes to serve. Knowing the pitfalls that could be created by their misuse equipped the teachers to make better use of the strengths of the standardized reading test. This understanding of the advantages and disadvantages of standardized reading tests seemed to have given some teachers a greater resolve to use them only at the appropriate times.

The intern also found the inservice program to be personally rewarding in that it afforded him an opportunity to see how teachers react to an inservice situation and to experience the problems of conducting an extended teacher education program. While university courses related to the development of inservice training programs are

beneficial, actual involvement in such a program is essential in order to combine theory with practice. The most significant learning experience for the intern was that of developing credibility with the teachers without jeopardizing rapport. The opinions of the teachers had to be respected; at the same time, the intern could not allow erroneous opinions that would be detrimental to proper reading instruction to prevail. The inservice program demonstrated to the intern the extent to which such an undertaking depends upon the development of good rapport and professional credibility.

Another insight which the intern derived from the program was the realization that the topic of reading evaluation functions exceptionally well as a focal point for a thorough discussion of any reading skill. The process of analyzing how a skill should be evaluated compels the teacher to review the functions of the skill, the circumstances affecting its maturation, methods of teaching and remediating it and worthwhile techniques for evaluating it. All of the factors that have any bearing on the development of a skill seem to have relevance for the manner in which it is assessed. In some inservice programs, it is very difficult to keep the teachers' attention focused on the topic under discussion. However, this problem did not arise in this inservice program since most of the comments that, on the surface, appeared not to be directly related to reading evaluation could fairly easily be incorporated into the discussions. This aspect of the program helped to develop and maintain good rapport between

the intern and the teachers.

RECOMMENDATIONS

Because the topic of reading evaluation permits such a wide range of discussion, the intern feels that the twelve week period allotted for the completion of the program was not sufficient time to allow an extensive treatment of the area of reading evaluation. Some of the program topics could have been pursued in more detail if time had been available. The intern feels that this type of inservice program might be more suitably conducted if it were extended over a period of fifteen to twenty weeks. This time allotment would provide for a greater exchange of comments and allow the teachers to compare the various standardized reading tests on the basis of actual administration as well as on the basis of oral analysis.

The topic of this internship could function very well as the basis for a school's inservice program. Due to its very comprehensive nature, the program could be conducted for a period of two or more years. It could involve the examination of standardized reading test use and interpretation, the construction and validation of informal reading tests and the designing of a testing program for the school. Such a program would most likely produce spin-off topics that could be run concurrently or become the foundation for future inservice programs. The evaluation of the present program indicated that teachers would benefit in many ways from an inservice program of

this nature.

In light of the sparsity of knowledge which teachers presently possess regarding the nature and types of standardized reading tests on the market, a school or school system could initiate an inservice program specifically designed to make its teachers aware of these tests. It would seem that the immediate value to be obtained from such a program would be the more enlightened use of tests and test results.

If the above cannot be accomplished, a school board should endeavor to provide each of its schools with an in-depth analysis of the reading tests that are currently in use in its schools. Although all reputable tests are accompanied by an examiner's manual, many teachers have not had sufficient training in the realm of evaluation to interpret some aspects of the manual. In addition to this, some manuals do not provide teachers with all the information they need to have in order to use the test correctly.

The area of reading evaluation is very important to successful reading instruction and should receive more attention in teacher training institutions than it presently is given. Unless teachers are capable of properly and accurately assessing the reading ability of their students, it is almost impossible to plan reading programs that will effectively utilize all the available resources. Teachers must become more knowledgeable about reading evaluation in order that effective and efficient reading instruction will occur and that decisions

based on test results will be educationally sound and beneficial to the student and society.

REFERENCES CITED

American Psychological Association. Standards for Educational and Psychological Tests and Manuals. Washington, D.C.: American Psychological Association, 1966.

Anastasi, Anne. Testing Problems in Perspective. Washington, D.C.: American Council on Education, 1966.

Beery, Althea, Thomas C. Barrett, and William R. Powell, eds. Elementary Reading Instruction: Selected Materials. 2nd ed. Boston: Allyn and Bacon, 1974.

Betts, Emmett A. Foundations of Reading Instruction, New York: American Book, 1957.

Bond, Guy L. and Miles A. Tinker. Reading Difficulties: Their Diagnosis and Correction. 3rd ed. New York: Appleton-Century-Crofts, 1973.

Bormuth, John R. "An Operational Definition of Comprehension Instruction, " in Althea Beery et al., eds. Elementary Reading Instruction: Selected Materials. 2nd ed. Boston: Allyn and Bacon, 1974.

Burnett, Richard W. "The Diagnostic Proficiency of Teachers of Reading," The Reading Teacher, 16 (January,1963), 229-34.

Burgett, R. E. and N. A. Glaser. "Appraising the Revised Standardized Reading Test," Elementary English, 50 (January,1973), 71-74.

Buros, Oscar. Sixth Mental Measurement Yearbook. Highland Park, New Jersey: Gryphon Press, 1965.

Carver, Ronald. "Reading Tests in 1970 Versus 1980: Psychometric Versus Edumetric," The Reading Teacher, 16 (December, 1972), 299-302.

Chall, Jeanne. "Interpretation of the Results of Standardized Reading Tests," in Roger Farr, ed. Measurement and Evaluation. New York: Harcourt, Brace and World, 1970.

Cox, Richard C. and Barbara G. Sterrett. "A Model for Increasing the Meaning of Standardized Test Scores," Journal of Educational Measurement, 7 (Winter,1970), 227-28.

Cronbach, Lee J. "Validation of Educational Measures," Paper presented at the Educational Testing Service Conference, New York, October, 1969.

Davies, William. "Functional Use of Standardized Reading Tests," in Althea Beery et al., eds. Elementary Reading Instruction: Selected Materials. Boston: Allyn and Bacon, 1969.

Durkin, Dolores. Teaching Them to Read. 2nd ed. Boston: Allyn and Bacon, 1974.

Durost, Walter. "The Characteristics, Use and Computation of Stanines," in Norman E. Gronlund, ed. Readings in Measurement and Evaluation: Education and Psychology. Toronto: The Macmillan Company, 1968.

_____. "Accountability: The Task, the Tools and the Pitfalls," The Reading Teacher, 24 (January,1971), 291-304, 367.

Durrell, Donald D. and Mary B. Brassard. Durrell Listening-Reading Series, Intermediate Level, Form DE, New York: Harcourt, Brace and World, 1969.

Ebel, Robert. "Standardized Achievement Tests: Uses and Limitations," National Elementary School Principal, 41 (September, 1961), 29-32.

_____. "Measurement and the Teacher," Educational Leadership, 20 (October,1962), 20-24.

Farr, Roger. "Reading Tests and Teachers," in Dorothy L. DeBoer, ed. Reading Diagnosis and Evaluation. Newark, Delaware: International Reading Association, 1970.

_____, and Virginia Brown. "Evaluation and Decision Making," The Reading Teacher, 24 (January,1971), 341-46.

Fitzgibbons, Thomas J. "Reading Tests and the Disadvantaged," in William E. Blanton et al., eds. Measuring Reading Performance, Newark, Delaware: International Reading Association, 1974.

Goodman, Kenneth, "Promises, Promises," The Reading Teacher, 24 (January,1971), 365-67.

Hagan, Elizabeth. "Analyzing Test Results," National Elementary School Principal, 41 (November,1961), 11-17.

Hamersma, Richard J. Educational and Psychological Tests and Measures. Dubuque, Iowa: Kendall-Hunt, 1972.

Harris, Larry A. "Evaluating a Reading Program at the Elementary Level," in Roger Farr, ed. Measurement and Evaluation. New York: Harcourt, Brace and World, 1970.

Hayward, Priscilla. "Evaluating Diagnostic Reading Tests," The Reading Teacher, 21 (March, 1968), 523-29.

Heilman, Arthur W. Principles and Practices of Teaching Reading. 3rd ed. Columbus, Ohio: Charles E. Merrill, 1972.

Hieronymous, A.N. "Evaluation and Reading: Perspective '72," The Reading Teacher, 26 (December, 1972), 264-67.

Hill, Walter R. "Reading Tests for Reading Evaluation," in William E. Blanton et al., eds. Measuring Reading Performance. Newark, Delaware: International Reading Association, 1974.

Hogan, Thomas. "Reading Tests and Performance Contracts," in William E. Blanton et al. eds. Measuring Reading Performance. Newark, Delaware: International Reading Association, 1974.

Julitta, Sister Mary. "Selection and Use of Standardized Reading Tests in Kindergarten Through Grade Three," in Helen M. Robinson, ed. Evaluation of Reading. Chicago: University of Chicago Press, 1958.

Karlsen, Bjorn and Richard Madden and Eric F. Gardner. Manual for Administering and Interpreting Stanford Diagnostic Reading Test, Level I, New York: Harcourt, Brace and World, 1966.

Kender, Joseph. "How Useful Are Informal Reading Tests," Journal of Reading, 11 (February, 1968), 337-42.

Killgallon, Patsy A. "A Study of Relationships Among Certain Pupil Adjustments in Language Situations," Unpublished doctoral dissertation, Pennsylvania State College, 1942.

King, Martha. "Evaluating Critical Reading," in Mildred A. Dawson, ed. Developing Comprehension, Newark, Delaware: International Reading Association, 1968.

Kress, Roy A. and Marjorie Johnson, "Evaluating Reading Progress," The Reading Teacher, 24 (January, 1971), 290.

Ladd, Eleanor M. "More Than Scores From Tests," The Reading Teacher, 24 (January, 1971), 305-11.

Letton, Mildred C. "Evaluating the Effectiveness of Teaching Reading," in Helen M. Robinson, ed. Evaluation of Reading. Chicago: University of Chicago Press, 1958.

Livingston, Howard. "What the Reading Test Doesn't Test – Reading," Journal of Reading, 15 (March, 1972), 402-10.

Maginnis, George H. "Measuring Underachievement in Reading," The Reading Teacher, 25 (May, 1972), 750-53.

Massad, Carolyn E. "Interpreting and Using Test Norms," The Reading Teacher, 26 (December, 1972), 286-92.

Masters, Howard. "Reading Tests Don't Tell It Like It Is," Reading Horizons, 13 (Winter, 1973), 59-63.

McCracken, Robert. "Informal Reading Inventories: Diagnosis Within the Teacher," The Reading Teacher, 26 (December, 1972), 273-77.

Mehrens, William A. and Irvin J. Lehmann. Standardized Tests in Education. New York: Holt, Rinehart and Winston, 1969.

_____. Measurement and Evaluation in Education and Psychology. New York: Holt, Rinehart and Winston, 1973.

Mitchell, Addie S. "Values and Limitations of Standardized Reading Tests," in J. Allen Figurel, ed. Forging Ahead in Reading. Vol. 12, pt. 1 Newark, Delaware: International Reading Association, 1968.

Mour, Stanley I. "Interpreting the Results of Standardized Tests," in J. Allen Figurel, ed. Forging Ahead in Reading. Vol. 12, pt. 1 Newark, Delaware: International Reading Association, 1968.

Powell, William. "The Validity of the Instructional Reading Level," in Robert Leibert, ed. Diagnostic Viewpoints in Reading. Newark, Delaware: International Reading Association, 1971.

Ramsey, Wallace. "The Values and Limitations of Diagnostic Reading Tests for Evaluation in the Classroom," in Leo M. Schell and Paul C. Burns, eds. Remedial Reading: Clinic and Classroom. 2nd ed. Boston: Allyn and Bacon, 1972.

Rankin, Earl F., Jr. "The Cloze Procedure - Its Validity and Utility," in Roger Farr, ed. Measurement and Evaluation of Reading. New York: Harcourt, Brace and World, 1970.

Rost, Paul. "Useful Interpretation of Standardized Tests," The Clearing House, 47 (January, 1973), 319-20.

Ruddell, Robert. "Preface," in Testing in Reading: Assessment and Instructional Decision Making. Richard Venezky, Urbana, Illinois: National Council of Teachers of English, 1974.

Sipay, Edward R. "A Comparison of Standardized Reading Scores and Functional Reading Levels," The Reading Teacher, 17 (January, 1964), 265-68.

Spache, George and Evelyn Spache. Reading in the Elementary School. 3rd ed. Boston: Allyn and Bacon, 1973.

Stake, Robert E. and James L. Wardrop. "Performance Contracts Test Errors," Reading Research Quarterly, 6 (Spring, 1971), 323-25.

Sutton, Marjorie Hunt. "Accountability and Teaching Reading," The Reading Teacher, 25 (March, 1972), 528-32.

Thorndike, Robert L. and Elizabeth Hagan. Measurement and Evaluation in Psychology and Education. 3rd ed. New York: John Wiley and Sons, 1969.

Traxler, Arthur E. "Values and Limitations of Standardized Reading Tests," in Helen M. Robinson, ed. Evaluation of Reading. Chicago: University of Chicago Press, 1958.

Wallen, Carl J. Competency in Teaching Reading. Toronto: Science Research Associates, 1967.

Wesman, Alexander G. "Reliability and Confidence," in Norman E. Gronlund, ed. Readings in Measurement and Evaluation: Education and Psychology. New York: The Macmillan Company, 1968.

Wiener, Morton and Ward Cromer. "Reading and Reading Difficulty: A Conceptual Analysis," Harvard Educational Review, 37:4 (1967), 620-43.

Wilson, Robert M. Diagnostic and Remedial Reading for Classroom and Clinic. 2nd ed. Columbus, Ohio: Charles E. Merrill, 1972.

Wyatt, Nita M. "Pinpointing Specific Skill Needs," in J. Allen Figurel, ed. Forging Ahead in Reading. Vol. 12, pt. 1. Newark, Delaware: International Reading Association, 1968.

Appendix A

LETTER TO

SPECIAL EDUCATION TEACHERS

March 17, 1975

Dear Teacher,

The School Board's Special Education Committee has been approached by Mr. T. Grace, a graduate student in reading at Memorial University, with a proposal for conducting an inservice program with our Special Education teachers on the topic of reading evaluation.

The inservice program is designed to provide the teacher with information about and experience in the evaluation of reading. The program would be of practical value to the teacher in the exercise of daily classroom responsibilities.

The primary goals of the inservice program are:

1. to analyze the major skills that contribute to the development of reading ability and to determine the most appropriate and valid methods for evaluating such skills

2. to familiarize teachers with the strengths and weaknesses of a number of formal reading tests so that teachers may be better able to select appropriate reading tests and be able to meaningfully interpret and use test scores.

3. to assist teachers in the construction of their own informal classroom reading tests.

The program will be conducted in a series of twelve one hour sessions. The sessions will be held once a week, after regular school hours. The specific time, day and location for the inservice program will be determined by what is most convenient for the teachers who wish to become involved in this program.

The Special Education Committee has been briefed concerning the proposed program and feels that it would be beneficial to any of our teachers and recommends that any teacher who would have the time become involved.

Teachers who would be interested in attending a preliminary session are asked to sign their names in the space below and to return this letter to the Supervisor of Special Education at the School

Board Office by Friday, March 21, 1975.

Those teachers who signify that they wish to attend this preliminary session are not committing themselves to participation in the program. If a teacher decides after this preliminary session that the program does not appeal to him or her, there is no obligation to continue any further with the program.

If you wish to attend the preliminary session, please sign your name and indicate your school in the spaces below.

Name: _____

School: _____

Teachers who may wish to obtain more information concerning this program may contact the Supervisor of Special Education at the School Board Office.

Sincerely yours,

Mrs. Alice Connolly
Supervisor of Special Education.

Thomas Grace
Graduate Student.

Appendix B

TEACHER INFORMATION QUESTIONNAIRE

Teacher Questionnaire

Name: _____

The following information is being sought so that the intern may obtain some indication of the background of the teachers with respect to their training in reading and evaluation.

1. How many reading courses have you completed? _____

Course	_____	date	_____	level	_____
"	_____	"	_____	"	_____
"	_____	"	_____	"	_____
"	_____	"	_____	"	_____

2. Have you taken any remedial reading courses? Yes No

Course	_____	date	_____	level	_____
"	_____	"	_____	"	_____
"	_____	"	_____	"	_____

3. Have you taken any courses in tests and measurements? Yes
No

Course	_____	date	_____	Level	_____
"	_____	"	_____	"	_____

4. What formal reading tests have you used in your classroom?

_____

_____

_____

5. With what other formal reading tests are you familiar?

_____

_____

_____

_____

6. Have you ever been involved in the development of a battery of informal reading tests? Yes No

7. In terms of its validity and reliability, what rating would you assign to oral reading as a method of evaluating the student's word recognition ability? 5 4 3 2 1

8. In deciding to choose a formal reading test, what would you set as your first criterion or consideration?

_____

_____

Appendix C

PROGRAM EVALUATION QUESTIONNAIRE

EVALUATION OF THE INSERVICE PROGRAM

The statements below are designed to assess your opinion with respect to the practicality and usefulness of this inservice program. Circle the response which best represents your feeling toward each statement.

1 = Strongly Agree 3 = Undecided 5 = Strongly Disagree
2 = Agree 4 = Disagree

1. The objectives of the inservice program were clearly stated at the first session	1	2	3	4	5
2. The program achieved the objectives stated at the first session	1	2	3	4	5
3. The materials used in the program were not relevant to the objectives of the program ...	1	2	3	4	5
4. In view of the objectives of the program the topics dealt with were adequately explained	1	2	3	4	5
5. The individual sessions were not well organized	1	2	3	4	5
6. The materials and ideas presented in this inservice program will benefit me in my future work as a teacher	1	2	3	4	5
7. The program provided sufficient guidelines to permit me to evaluate the worth of a standardized reading test	1	2	3	4	5
8. As a result of the inservice program, I feel that I will be able to choose reading tests more wisely	1	2	3	4	5
9. The inservice program provided me with a better understanding of reading evaluation than I had prior to the program	1	2	3	4	5
10. If a similar inservice program were offered again, I would recommend that other teachers participate in it	1	2	3	4	5

Additional comments:

Appendix D

A LIST OF READING TESTS USED IN THE INSERVICE PROGRAM

Doren Diagnostic Reading Test of Word Recognition Skills
Margaret Doren, Circle Pines, Minnesota: American Guidance Services, Inc., 1973.

Durrell Listening-Reading Series Primary level, Form DE
Donald D. Durrell and Mary Bassard. New York: Harcourt, Brace Jovanovich, Inc., 1970.

Durrell Listening-Reading Series Intermediate level, Form DE
Donald Durrell and Mary Bassard. New York: Harcourt Brace Jovanovich, Inc., 1970.

Gates-MacGinitie Reading Tests Survey D
Arthur I. Gates and Walter H. MacGinitie. New York: Teachers College Press, 1965.

Gates-McKillop Reading Diagnostic Tests Form 1
Arthur I. Gates and Anne S. McKillop. New York: Teachers College Press, 1962.

Gray Oral Reading Tests Form A
William S. Gray. New York: The Bobbs-Merrill Company, 1963.

New Developmental Reading Tests Intermediate level
Guy L. Bond, Bruce Balow and Cyril Hoyt. Chicago: Lyons and Carnahan, Inc., 1968.

Silent Reading Diagnostic Tests
Guy L. Bond, Bruce Balow and Cyril Hoyt. Chicago: Lyons and Carnahan, Inc., 1970.

Stanford Diagnostic Reading Tests Level 1 Form W
Bjorn Karlsen, Richard Madden and Eric F. Gardner. New York: Harcourt, Brace and World, Inc., 1966.

Stanford Diagnostic Reading Test Level 11 Form W
Bjorn Karlsen, Richard Madden and Eric F. Gardner. New York: Harcourt, Brace and World, Inc., 1966.

Appendix E

INSERVICE PROGRAM NOTES

EVALUATION IN READING

PURPOSES OF THE INSERVICE PROGRAM

This inservice program will attempt to provide teachers with an understanding of the concepts involved in the evaluation of reading ability so that they will be able to:

1. critically analyze and evaluate various types of reading tests
2. select the most appropriate test for the evaluative purposes that have been set
3. correctly interpret the results of reading tests
4. construct more effective informal tests.

This understanding of the concepts relating to the evaluation of reading ability will be achieved through discussion of the following topics:

1. The analysis of the major reading skills, their functions within the reading act, their relevance to the total reading program and suitable techniques for their evaluation
2. The analysis of the primary measurement concepts--validity, reliability, norms and the interpretation and use of scores
3. The critical examination of several standardized reading tests, using as criteria the information covered in topics 1 and 2.

ANALYSIS OF THE MAJOR READING SKILLS

The skills that will be analyzed are:

1. Visual Ability
2. Auditory Ability
3. Phonics
4. Structural Analysis
5. Contextual Analysis
6. Vocabulary
7. Comprehension
8. Oral Reading

The analysis of each skill will assume the following format:

A. a discussion of any prerequisite skill(s)

B. definition of the skill

C. description of the functions of the skill during the reading act

D. discussion of any special consideration related to the development of the skill

E. the prominence of the skill within the total reading program

F. listing of some techniques for evaluating the skill

G. a series of questions that may be used to analyze a test designed to measure a specific skill (There are six basic questions that may be asked of most skills tests.)

H. a statement of problems that may arise in some standardized skills tests.

Visual Ability

Visual ability is a reading related skill, and it is examined first because it is central to the entire reading process. Reading begins in the visual modality, and good visual ability is essential in order for a student to become a mature reader.

A. Prerequisite Ability

An understanding of the concepts of "same" and "different".

B. Definition

Visual ability refers to the power to see print and to discriminate between the various graphic features.

C. Functions of Visual Ability During the Reading Act

There are three primary functions: (1) acuity, (2) discrimination, (3) memory.

Visual acuity--the student's ability to perceive things as they really are. The student must be able to see the graphic symbols clearly.

Visual discrimination--the student's ability to detect similarities and differences among the various types of graphic features. The progression of this skill is as follows:

1. discrimination of gross and fine geometric figures
2. discrimination of individual letters
3. discrimination of common groupings or clusters of letters
4. discrimination of phrasal units.

Two other activities that may be placed under the heading of visual discrimination are "tracking"--following a line of print with only eye movement--and "return sweep"--returning to the left-hand side of the page after reading the previous line.

Visual memory--the student's ability to recall printed images. This facility is important in the development of phonics skills, vocabulary and comprehension.

D. Special Considerations

None.

E. Position of Visual Ability in the Total Reading Program

The development of good visual skill is essential to the success experienced by the student. After his acuity has been thoroughly examined in kindergarten and grade one, the functions of discrimination and memory should form an integral segment of the reading program, at least to the end of the elementary grades. As the student proceeds through the reading program, the visual exercises should become more

sophisticated so as to challenge his developing ability.

Diagnosis of visual ability should be completed first to ensure that the reading difficulty is not primarily a visual one.

F. Appropriate Techniques for Evaluating Visual Ability

Acuity--Keystone telebinocular or similar device.

Reading is basically a nearsighted activity, consequently, tests of farsightedness often do not identify most of the students who may have problems of acuity.

Discrimination--matching and differentiation exercises.

After the level of geometric shapes, these exercises should involve discrimination of orthographic features.

The test exercises must be equivalent in degree of difficulty as that required by the level of the program in which the student is at the time of testing.

Memory--It may be measured by presenting the student with a nonsense word and asking him to reproduce it.

Establishing a suitable exposure time for each age level can be a very difficult task.

G. Questions to Ask About Tests of Visual Ability

1. What visual skills are measured by the test?
2. What visual skills are not being measured?
3. Are the skills being measured in the same way they are used during the act of reading?

4. Is it possible to complete the test without using the skill that the test purports to measure?
5. What prerequisite skills must the student possess in order to perform the skill required?
6. Is the degree of difficulty equivalent to the degree of difficulty demanded by the student's reading level?
7. Is the test one of nearsightedness or farsightedness?

H. Problems of Standardized Tests of Visual Ability

1. Some tests of visual ability, such as the Snellen, are farsighted tests and frequently miss as high as 70 percent of the students who later experience some visual difficulties.
2. Tests of visual discrimination are often not based on orthographic features.
3. Visual tests sometimes do not test any aspect of visual memory.
4. Some visual tests are below the level of difficulty that is demanded by the student's reading level.

Auditory Ability

A. Prerequisite Ability

Exposure to a considerable amount of oral communication, in the form of both speaking and listening.

B. Definition

Auditory ability refers to the student's ability to hear sounds and to distinguish between the various phonemes.

C. Functions of Auditory Ability During the Reading Act

The functions listed below are not directly involved in the act of reading but are necessary for the acquisition of many reading skills: (1) auditory acuity, (2) auditory discrimination, (3) auditory memory, (4) auditory blending.

Auditory acuity--the ability to hear individual phonemes.

Auditory discrimination--the ability to distinguish one phoneme from another. It operates in the learning of phonics, the development of proficient spelling ability and the use of syllabication.

Auditory memory--the ability to recall sounds that have been heard. It is important in the learning and use of phonics, the development of listening vocabulary and auditory comprehension. A wide range of learning experiences involve oral communication and the following of explanations.

Auditory blending--the ability to hear separate phonemes and to put them together to form a word. It is necessary in the use of phonics and the application of structural analysis generalizations.

D. Special Considerations

1. Some individuals exhibit a deafness to a particular phoneme.
2. Dialect can affect the way in which a student may hear some phonemes.
3. Research indicates that very young children do not fully develop an auditory ability for the phonemes l, sh, th, v, s, z, j and r until the mid-primary level.

E. Position of Auditory Ability in the Total Reading Program

Good auditory ability is essential for the development of reading facility via the phonics method of teaching. It is an ability that is heavily relied upon throughout the student's life.

F. Appropriate Techniques for Evaluating Auditory Ability

Acuity--It may be examined with the Maico audiometer or a similar device.

Discrimination--The student may be presented orally with two words that differ by only one phoneme and be asked to state if they sound the same or different. Ex. cat - rat. It is preferable to use nonsense words rather than real words. The difference in the stimulus words may involve beginning, medial or final sounds. Students may also be required to distinguish between the beginning, medial or ending

sounds of polysyllabic words. A more advanced A-D test may require the student to mark the primary accent position of orally presented words.

Memory--The student may be presented with several sounds and be required to reproduce them. The series of sounds is made progressively longer. The difficulty with this testing technique is the establishing of suitable criterion levels of performance for the various age groups.

Blending--The student may be presented with a graphic set of several phonemes and be required to blend them orally to form a unified whole.

Ex. cl - ir - d = clird

It seems best to use nonsense words rather than real ones. When nonsense words are used, the examiner must make sure that the nonsense words pattern English orthography. Blending tests presuppose that the student has a strong knowledge of phonics generalizations.

G. Questions to Ask About a Test of Auditory Ability

1. What auditory skills are being measured?
2. What auditory skills are not being measured?
3. Are the skills being measured in a manner similar to the way in which the student uses them when learning

to read?

4. Is it possible to complete the test without using the auditory skill that the test is designed to measure?
5. What prerequisite skills must the student possess in order to complete the test?
6. Is the degree of difficulty equivalent to the degree of difficulty demanded by the student's reading level?
7. Could the test results be affected by a problem of student dialect?

H. Problems of Standardized Tests of Auditory Ability

1. Most auditory tests assume that the child's acuity is not impaired. In many instances students have not been given an acuity test; consequently, there is no basis for assuming that their acuity is good.
2. Orally presented blending tests can be severely confounded by the dialect of the examiner.
3. Some blending tests may be completed without using the skill of auditory blending.
4. Discrimination tests usually report only global scores and these are of very little use to the diagnostic teacher.
5. Orally presented minimal-paired discrimination tests can be confounded by the dialect of the student.

Phonics

A. Prerequisite Skills

The visual discrimination capacity to recognize individual letters or letter clusters within words.

The ability to discriminate auditorily the phoneme being taught. Before any phonics instruction is undertaken, the teacher should check to determine whether or not the student has the facility to discriminate auditorily the sounds that are being taught in the phonics program.

B. Definition

Phonics is based on the association of a particular grapheme with a particular phoneme. In reading phonics involves the process of applying these grapheme-phoneme relationships to the decoding of unknown words. Phonics is a word recognition technique used to obtain only the sound of a word.

C. Functions of Phonics During the Reading Act

Using a series of grapheme-phoneme generalizations, the student has to manufacture a word from a group of printed symbols.

The functions involved in this process are:

1. visual recoghition of individual letters or clusters of letters that constitute a phoneme; that is, blends,

diagraphs, vowels, etc.

2. translation of the grapheme into a phoneme

3. the blending of the graphemes to form a word

Ex. "blank" = recognition of the phonemes "bl" and "ank" and then their blending to form the word "blank".

4. the recognition of recurring letter patterns known as phonograms that always have the same sound whenever they appear

5. the division of a word through syllabication

D. Special Considerations

1. There are many words, many of them high frequency words, that do not conform to phonetic generalizations. Several utility studies have been completed to determine the most consistent of these.

2. Not all students learn easily via the auditory method.

3. Some students may have an auditory discrimination problem with respect to certain phonemes.

4. Dialect interference may hinder the development of some phonic generalizations as effective aids to learning to read.

Ex. Teaching the sound of "h" in some parts of the province is very difficult.

5. Overemphasis on the prolonged use of phonics can

produce word-by-word readers.

6. Phonics builds word recognition ability only, NOT comprehension ability.

7. The understanding and application of most phonics generalizations requires at least average conceptual ability.

E. Position of Phonics Within the Total Reading Program

Phonics is very important in the early stages of learning to read but its importance gradually decreases as more efficient skills replace it. However, some time should be allotted in the elementary program for the teaching and reviewing of phonic generalizations. Phonics is most effective when used in combination with the other word recognition skills.

F. Appropriate Techniques for Evaluating Phonics

1. Effective testing of grapheme-phoneme relationships should be done on an individual basis by means of selected oral reading that is taped and analyzed later.

2. When specific generalizations are to be measured, the tests should use nonsense words that pattern English orthography.

3. Tests should primarily examine the high utility generalizations.

4. Phonics tests should always be the translation of print

to sound.

G. Questions to Ask About Standardized Phonics Tests

1. What functions of phonics are being measured?
2. What functions of phonics are not being measured?
3. Is phonics being examined in the same way it is used during the act of reading?
4. Is it possible to complete the test without using the phonics skill that the test is designed to measure?
5. What prerequisite skills must the student possess in order to complete the test?
6. Is the degree of difficulty equivalent to the degree of difficulty demanded by the student's reading level?
7. Is the test concerned with sound-to-letter or letter-to-sound relationships?
8. Is it a group or individual test of phonics?

H. Problems of Standardized Phonics Tests

1. Many only measure sound-to-letter relationships.
2. There are frequently not enough test items to examine a generalization adequately.
3. Many phonics tests can be done without a good knowledge of phonics.
4. Some tests require a considerable amount of time to administer.

Structural Analysis

A. Prerequisite Skills

1. A good visual discrimination ability

2. Some degree of sight meaning vocabulary

B. Definition

Structural analysis is a word recognition technique based on the visual recognition of known segments of unfamiliar words.

C. Functions of Structural Analysis Within the Reading Act

In structural analysis the student attempts to decipher an unknown word by utilizing the meaningful part of the word that he has met previously. To do this, the student must be able to visually identify the known parts such as recognizing "some" and "thing" in the word "something". The process of recognition will also apply to inflectional endings and affixes. Having discriminated the known segments of the new word, the student has to be capable of putting these together to form a new word. It is a semantic blending of word parts. Thus the functions of structural analysis during the reading act involve:

1. visual recognition of the known segments
2. recalling the meaning of the known elements
3. blending the meanings to form a new word.

Although structural analysis is a skill that the student

uses to find the meaning of a word, the recognition of consistent letter clusters will also help him to obtain the pronunciation of the word. In this respect, we are referring to the recognition of common phonograms and affixes.

D. Special Considerations

1. The student can only employ the technique of structural analysis after some meaning-sight vocabulary has been developed.

2. When the student attempts to use structural analysis to obtain the pronunciation of a word, he should be shown that many letter clusters change their pronunciation when they are combined with other morphemes; for example, the word "know" has a different sound when it appears in the word "knowledge". The practice of teaching students to find small words inside large ones is of very limited utility and may actually hinder the students' ability to decode the word.

E. Position of Structural Analysis Within The Total Reading Act

Structural analysis is a very important reading skill because it helps the student understand the meaning of what is being read. As soon as the student has developed an adequate meaning-sight vocabulary, he

should be taught to use structural analysis to unlock the meaning of a new word. This skill should increase in importance as the student progresses through the reading program. Structural analysis should be combined with the teaching of phonics. The student who is strong in the use of this skill will become a good reader.

F. Appropriate Techniques for Evaluating Structural Analysis

1. The student may be required to find root words or affixes.
2. The student may be asked to give the meanings of various types of affixes.
3. The student may be given a number of morphemes and asked to form an English word.

G. Questions to Ask About Standardized Structural Analysis Tests

1. What functions of structural analysis are being measured?
2. What functions of structural analysis are not being measured?
3. Does the test examine structural analysis in a manner similar to the way in which it is used in the reading act?
4. Is it possible to complete the test without using the

structural analysis skill the test is designed to measure?

5. What prerequisite skills must the student possess in order to complete the test?

6. Is the degree of difficulty equivalent to the degree of difficulty demanded by the student's reading level?

H. Problems of Standardized Structural Analysis Tests

1. Some tests require the student to find small words within larger ones when the small word is not a morphemic element of the larger word, for example "in" in the word "kind".

2. Most tests do not test to determine if the student knows the meaning of the various parts of the word he has to identify.

Contextual Analysis

A. Prerequisite Skill

The student should be able to recognize 95 percent of the words in the passage he is reading.

B. Definition

Contextual analysis involves determining the approximate meaning of an unknown by using the meanings of the words that surround it. Contextual analysis may involve using the ideas of a passage to predict outcomes, make inferences or to check the logic of what has been

read.

C. Functions of Contextual Analysis During the Reading Act

1. As a vocabulary development skill, the student uses the surrounding text to determine the approximate meaning of an unknown word. There are a number of ways in which contextual analysis may function. The following are the more commonly used methods:

   a. Use of the appositive

   b. Use of a coordinate conjunction construction

   c. Use of nonrestrictive clauses

   d. Use of comparison or contrast

   e. Use of the mood of the other words in the sentence

   f. Use of inferences drawn from other parts of the passage

2. As a general comprehension skill, the student:

   a. uses the ideas that he has read to predict the outcome or to anticipate developments in a story;

   b. uses the ideas that he has read to check the logic of what he has been comprehending. This use of contextual analysis is a check on the student's word recognition.

D. Special Considerations

None

E. Position of Contextual Analysis Within the Total Reading Program

Contextual analysis is a very important reading skill and one that will be used very often by the mature reader to decode the meaning of new words. As the student progresses through the reading program, the teaching of contextual analysis should become more sophisticated and the student should be aware of the specific ways in which contextual analysis may function as an aid to his reading. A student who is deficient in the use of contextual analysis may suffer problems of comprehension and be working well below his potential.

F. Appropriate Techniques for Evaluating Contextual Analysis

To evaluate contextual analysis as a vocabulary skill, the student may be given a sentence in which an unfamiliar word appears along with one of the contextual devices and the student has to choose a word from four possibilities that he thinks has the same meaning as the unfamiliar word.

Ex. John was <u>adamant</u> about wanting to go to the game and he wasn't about to change his mind for anyone.

1. quick 2. firm 3. undecided 4. thinking

To evaluate contextual analysis as a comprehension skill:

1. the student may be asked to complete a sentence or passage by supplying the missing part

2. he may be given a complete sentence or passage and be asked to answer an inferential question or to make a prediction as to possible future events in the story. This type of testing can only be done on an individual basis.

G. Questions to Ask About Standardized Contextual Analysis Tests

1. What functions of contextual analysis are being measured?
2. What functions of contextual analysis are not being measured?
3. Does the test examine contextual analysis in a manner similar to the way in which it is used in the reading act?
4. Is it possible to complete the test without using the contextual analysis skill the test is designed to measure?
5. What prerequisite skills must the student possess in order to complete the test?
6. Is the degree of difficulty equivalent to the degree of difficulty demanded by the student's reading level?

H. Problems of Standardized Contextual Analysis Tests

1. Tests that employ the cloze technique may often measure syntactic knowledge as much as contextual

analysis.

2. The reading level of a contextual analysis test may be at or above the student's instructional reading level.

3. Many contextual analysis tests do not measure the skill as an aid to word meaning but only as an aid to inferential comprehension.

4. Some tests of contextual analysis are tests of reading meaning vocabulary.

5. Items of the tests are not leveled in terms of their degree of contextual difficulty.

Sight Vocabulary

A. Prerequisite Skills

1. Good visual discrimination

2. Good auditory memory

B. Definition

Sight vocabulary refers to the vocabulary that the student recognizes immediately.

C. Functions of Sight Vocabulary During the Reading Act

The above definition explains the function of a sight vocabulary. The ability to immediately recognize some words allows the student to devote more time to the act of comprehension.

D. Special Considerations

Some meaning-sight words may not be recognized immediately in isolation but will be when they are found in the context of a sentence.

E. Position of Sight Vocabulary Within the Total Reading Program

Sight vocabulary is an extremely valuable skill at all levels of the reading program. To effectively comprehend what is being read, the student cannot afford to spend time trying to decode each word. Time spent in this manner disrupts the student's line of thought and comprehension must suffer as a result. Consequently, considerable attention must be allotted to building the student's sight vocabulary. In order to employ his skills of structural and contextual analysis, the student will require a strong sight vocabulary since it is a prerequisite ability for the maturation of these reading skills. In developing the student's sight vocabulary, special attention should be given to the high frequency words, particularly those that are phonetically irregular.

F. Appropriate Techniques for Evaluating Sight Vocabulary

1. Flash card method--The student is shown a word for a second or two and is expected to give the pronunciation immediately.

2. Tachistoscopic method--This device exposes a word for a fraction of a second and the student must pronounce it.

G. Questions to Ask About Standardized Sight Vocabulary Tests

1. What were the criteria for the selection of words?

2. If the test is graded, how many words are in each grade list?

3. From what grade level have the words been chosen?

4. What is the manner of presentation?

H. Problems of Standardized Tests of Sight Vocabulary

1. No justification is given for the particular selection of words in the test.

2. If the words are divided into graded lists, there may be too few words in each list and dramatic grade score changes may occur as a result of one or two errors.

3. Some examiners permit the student to study the words for longer than a second or two. In such a presentation the student has time to apply his analytic decoding skills.

Meaning Vocabulary

The student has three types of meaning vocabulary: (1) speaking, (2) listening, (3) reading.

A. Prerequisite Skills

Listening vocabulary: (1) auditory acuity and discrimination, (2) varied experiential background.

Reading vocabulary: (1) average ability in the skills of phonetic, structural and contextual analysis, (2) good basic sight vocabulary, (3) varied experiential background.

B. Definition

Listening vocabulary refers to the body of spoken words for which the student has a meaning.

Reading vocabulary refers to the body of words that the student is able to read and for which he has a meaning.

C. Functions of Meaning Vocabularies in the Reading Act

Listening vocabulary. Although the student is not involved in listening while he reads silently, he will often use his listening vocabulary to help him decode or find the meaning of an unfamiliar word. By obtaining some of the sounds in a difficult word and combining them with the context of the sentence, the student may recall the word from his listening vocabulary. In this way, partial recognition of the word triggers a memory response and the student goes to his listening vocabulary

to find the correct word for the situation. Because of this association between the reading and listening vocabularies, it becomes important that efforts be made to increase the student's listening vocabulary through many types of oral language activities.

Reading vocabulary. It is the student's reading vocabulary that enables him to gain meaning from the printed medium. The student must be able to supply a meaning for all the visual patterns he encounters as he reads. However, the development of this ability is very dependent upon the student's facility in the previously discussed skills.

D. Special Considerations

In the primary grades, the student's listening vocabulary exceeds his reading vocabulary. However, as he progresses through the program, his reading vocabulary should surpass the listening vocabulary. This reverse in the sizes of the vocabularies usually occurs about grade seven or eight.

In the primary grades, the extent of the student's listening vocabulary may be used as a rough estimate of his reading potential at that particular time.

E. Position of Meaning Vocabularies Within the Total Reading Program

Both meaning vocabularies are essential components of the reading program, with the reading vocabulary being the more crucial in terms of printed communications. The entire reading program should be designed to develop both vocabularies for they are the two most basic components of comprehension.

F. Appropriate Techniques for Evaluating Meaning Vocabulary

Listening vocabulary--The student is given a word without context and he has to assign it to a particular category, for example the word "flammable" may be assigned to the category of "burning".

Reading vocabulary--A word is presented to the student in the context of a sentence and he has to find a synonym for it from a set of four words.

G. Questions to Ask About Standardized Meaning Vocabulary Tests

1. Is it a test of listening or reading vocabulary?
2. Are the words presented in context or in isolation?
3. What is the reading level of the context words in the reading vocabulary test?
4. What is the reading level of the multiple choice answer words?
5. Is the level of difficulty equivalent to the degree of difficulty demanded by the student's reading level?

H. Problems of Standardized Tests of Meaning Vocabulary

1. Some meaning vocabulary tests do not take into account the problem of multiple meanings when the words are presented in isolation.
2. Tests which use context may become tests of contextual analysis if too much context is provided.
3. The reading meaning vocabulary tests may not have controlled the reading level of the context words or the multiple choice answer words.
4. Most vocabulary tests test only for the most basic meaning of a word and do not test for depth of vocabulary meaning.

Comprehension

The student has two types of comprehension (1) listening (auditory) comprehension, (2) reading comprehension, and each of these can be subdivided into literal and inferential comprehension.

A. Prerequisite Skills

Listening comprehension (1) adequate auditory acuity and discrimination, (2) adequate auditory memory and vocabulary

Reading comprehension (1) adequate word recognition ability, (2) adequate reading meaning vocabulary

B. Definitions

Listening comprehension refers to the student's ability to

listen to a passage and to answer literal and inferential questions related to it.

Reading comprehension refers to the student's ability to read a passage and to answer literal and inferential questions related to it.

Literal comprehension refers to the student's ability to answer questions whose answers can be found in the factual material of the passage.

Inferential comprehension refers to the student's ability to use the factual information of a passage to make predictions about future events or to draw conclusions concerning events that are not explicitly stated in the passage.

C. Functions of Comprehension During the Reading Act

Because comprehension is a function of the purpose for which one reads, and of the type and difficulty of the material being read, one should not consider comprehension to be a singular ability that operates in the same manner in each learning situation. The type of comprehension that is required to find and use a telephone number is vastly different from that required to solve a math word-problem or that which is used to verify a hypothesis in science. It is essential to realize that different subject areas and different levels

of complexity within a subject area will demand a unique type of comprehension ability. When considering the topic of comprehension, one should view it in terms of the task to be accomplished. One should not readily and uncritically assume that a student's comprehension ability in one situation reflects what it will be in a different situation.

D. Special Considerations of Comprehension

In the earlier grades, approximately K to 6, the student's auditory comprehension will exceed his reading comprehension. However, the reading comprehension usually begins to match and overtake the student's auditory comprehension at about grade seven or eight. Auditory comprehension usually involves a considerable degree of auditory memory. If a student has a poor auditory memory, tests of auditory comprehension may test auditory memory more than comprehension in instances where the passages are rather lengthy.

The types of questions which a teacher asks has a significant influence upon the types of comprehension abilities that the student will develop.

In listening comprehension tests, the examiner's dialect and enunciation can influence the student's responses.

In the interpretation of the student's comprehension score, the teacher should consider the degree of familiarity which the student may have with the subject matter of the test and the relationship that exists between the syntactic structures of the passage and those that the student utilizes in his speech.

E. Position of Comprehension Within the Total Reading Program

From the readiness stage of reading through to the end of the program, comprehension should be the principal goal of all reading instruction. While particular instruction will be given to improve separate reading skills, the teacher's primary goal should be the development of these skills so that the student may better understand what he reads. Since comprehension is not a singular ability, teachers should ensure that they require the student to perform various types of comprehension activities. Teachers should be cognizant of the role which their oral questions can play in helping the student develop his comprehension ability.

F. Appropriate Techniques for Evaluating Comprehension

In a listening comprehension test, the student may be required to listen to a passage and then to answer

questions by giving oral responses or by marking an answer sheet. If the student must respond on an answer sheet, the examiner should read the multiple answers from which the student must select an answer.

In reading comprehension tests, the student reads a passage and has to answer questions that follow the passage. Some comprehension tests use a modified cloze procedure to test comprehension. However, it is very difficult to devise a literal comprehension test using this technique.

In either type of comprehension, the teacher may use a "free response" format. In this situation the student does not have to choose from a fixed set of responses but devises his own answer and it is recorded by the teacher. While this method will often provide the teacher with a great deal of diagnostic information, it is very time consuming since it must be done on an individual basis.

For the very young student and the retarded reader, the picture format must be used to measure comprehension.

G. Questions to Ask About Standardized Comprehension Tests

1. What aspect of comprehension ability is being measured?

2. Is the test one of literal, inferential or both types of comprehension?

3. How familiar is the student with the content of the comprehension test?

4. In the reading comprehension test, how significant a factor is word recognition when the test is designed to measure a particular type of comprehension?

5. Is the comprehension test timed or does the student have as much time as he needs to complete the test?

6. Is the type of comprehension demanded by the test equivalent to the type required by the student's school program?

H. Problems Related to Standardized Comprehension Tests

1. They frequently measure only literal comprehension.

2. They generally provide only a global score of the student's comprehension ability rather than measure specific types of comprehension skills. The global score does not provide the teacher with much diagnostic information.

Oral Reading

A. Prerequisite Skill

Adequate visual acuity and discrimination of print.

B. Definition

Oral reading refers to the student's ability to pronounce

accurately the words that he is reading and to do this in a manner that reflects the natural intonation patterns of the spoken language.

C. Functions of Oral Reading During the Reading Act

Oral reading is by definition the vocalized aspect of the reading act and is, therefore, considered a function of word recognition rather than the basis for the development of another reading skill. Oral reading is viewed as an output and not an input of the reading act. However, it is recognized that under certain conditions and for some readers oral reading may function as an aid to comprehension or to a better understanding of English intonation patterns as they pertain to printed communication.

D. Position of Oral Reading Within the Total Reading Program

In and of itself, oral reading has only minimal value in terms of its contribution to the development of reading ability. It may be used to indicate to the student that the flow of printed material is based on the rhythm which the ideas would have if they had been spoken by the author rather than written. Round-robin oral reading sessions often hinder the development of reading ability more than they help it. Oral reading should be used primarily to diagnose the student's

word recognition and intonation abilities.

E. Special Considerations for Oral Reading

1. Oral reading diagnosis only provides an indication of the student's word recognition ability.

2. Oral reading passages used for testing purposes should be 100 or more words in length and usually taken from unseen reading material.

3. In scoring the student's oral reading, some reading researchers suggest that sporadic repetitions and hesitations should not be counted as errors since they may actually represent good reading ability. However, if the student is experiencing frequent hesitations or repetitions, the teacher may assume that the passage is very close to his frustration level.

F. Appropriate Techniques for Evaluating Oral Reading

Select a number of paragraphs of 100 words or more from various grade levels. These paragraphs should be taken from unseen material and represent reading achievement at about the midpoint of the various grade levels. Begin the testing by having the student read a paragraph or two that the teacher feels is at least one year below the student's apparent instructional reading level. Continue the reading until the student encounters considerable reading difficulty. For purposes of

scoring, it is advisable to record on tape the student's reading for it is very difficult to note all the student's errors and their possible causes.

The following are considered oral reading errors or miscues:

1. mispronunciations (Dialectic pronunciations are scored as correct pronunciations.)
2. omissions
3. insertions
4. substitutions
5. reversals
6. repetitions and hesitations. (These are considered oral reading errors on some tests.)

The analysis of the student's miscues may reveal a pattern of weakness in phonics, sight vocabulary, use of context, structural analysis or visual discrimination. With this information, the teacher will be in a position to plan appropriate remedial instruction to overcome the student's weaknesses.

By counting the number of miscues, the teacher is able to determine the student's independent, instructional and frustrational reading levels. The percentage of correct responses for each level is as follows:

independent reading level = < 99%

instructional reading level = 95%

frustration reading level = below 90%

By attaching ten comprehension questions to each passage, the teacher may obtain similar levels with respect to the student's comprehension ability. These questions should involve both literal and inferential comprehension. The percentage of correct responses for each comprehension level is as follows:

independent comprehension level = 90%

instructional comprehension level = 70%

frustration comprehension level = below 50%

If the teacher wishes, he may also obtain an estimate of the student's silent reading comprehension levels. In this case the student reads a different set of paragraphs silently and then the teacher asks the ten comprehension questions. The comprehension percentages above are also used for establishing silent reading comprehension levels.

From an analysis of the oral reading and comprehension levels, the teacher should be in a position to decide the type of reading material that may be used to instruct the student.

The results of an informal reading inventory should be used in conjunction with other diagnostic measures.

G. Questions to Ask About Oral Reading Tests

1. Are repetitions and hesitations considered as miscues?
2. Are the paragraphs at least 100 words in length?
3. From where were the paragraphs taken?
4. How many comprehension questions are there?
5. Do the questions require inferential as well as literal comprehension?
6. Does the manual that accompanies the test provide possible explanations concerning the various types of miscues and suggestions for remediation?

H. Problems Related to Standardized Oral Reading Tests

1. The comprehension questions are only literal.
2. The reading passages are too short.
3. Repetitions and hesitations are considered as miscues.
4. The accompanying manuals provide very little information concerning the possible causes of miscues.

ANALYSIS OF SOME STATISTICAL CONCEPTS IN STANDARDIZED TESTING

Validity

Validity refers to the ability of a test to measure what it is designed to measure. Before passing judgement on the validity of a test, one must be fully aware of the purposes for which the test is designed and the purposes for which the test is being used in a given situation. These two sets of purposes must coincide in order for the test to have any validity. A particular test may be quite valid in one situation but quite invalid in another when it is being used to achieve purposes for which it was not designed. Complete awareness of the goals of the testing situation and the goals for which a test was designed is crucial to the proper utilization of any test.

Our discussion of test validity in reading will center around the aspects of content validity and the validity of the procedure used to measure a reading skill. The topic of score validity will be treated under the heading of reliability.

If a teacher is to obtain an accurate assessment of the student's strengths and weaknesses in a particular reading skill, it is imperative that the items of the test adequately represent all facets of the skill being measured. If a test lacks items on a specific aspect of a reading skill, it does not have full content validity and the teacher has no means of determining the student's proficiency with regard to that

particular area of the reading skill. In order for a test to have complete content validity it must contain items that test all facets of the reading skill. The teacher, however, must first know the various behaviours that constitute a skill before he is able to ascertain whether or not the test adequately represents a full examination of that skill. Teachers should examine the test manual to see if the authors of the test provide any explanation of which items have been included in the test. When the manual does not provide such information, it may be possible that decisions made on the basis of test results could be inaccurate. If a test does not adequately measure the skill it was designed to measure then the score obtained is meaningless to the teacher.

After the teacher has concluded that the test possesses the prerequisite content validity, he should concern himself with an examination of the procedure used to measure the skill. Since the teacher wishes to determine whether or not the student is capable of utilizing the skill in an actual reading situation, it is important that the test measure the skill in a manner that is as close to its real application as is possible and practical. Some reading tests do not measure the skill in the manner it is used during the reading act. Most paper-and-pencil tests of phonics examine the sound-to-letter relationship and not the letter-to-sound relationship. A score on the former is not an indication of the student's ability to apply phonics generalizations to new words. To help in deciding the question of

procedural validity, teachers are referred to Section C of the notes dealing with the reading skills. This section briefly outlines the manner in which a skill functions during the reading act. When a test examines a skill in a manner that is different from the way in which the skill is used in the process of reading, the test manual should supply a justification for the procedure.

Even though a test may appear to have adequate content validity and to examine the skill in a legitimate manner, the teacher cannot unquestioningly accept the student's score as a perfect estimate of his facility in that skill. There may be a number of other factors than the student's ability that may contribute to the score obtained. Some of these will be explained in the following section.

Reliability

The concept of reliability relates to the score which an individual obtains on a test. By definition, reliability refers to the power of a test to provide approximately the same score if the test were administered several times. It is a measure of consistency.

The degree of concern which a teacher has regarding the reliability of a test will be determined by the nature of the decisions that will be made on the basis of the results obtained. The more crucial the decision to be made, the more concerned should the teacher be regarding the test's reliability.

Due to the many different factors that may contribute to the

production of a reliability coefficient, it is very difficult to state a specific reliability level appropriate for all testing situations. When making decisions regarding the grade or program placement of a student, it is recommended that the teacher select a test which has a reliability coefficient of .85 or better. If, however, the reliability coefficient has been obtained via the equivalent form method with a time lapse between the two forms, the teacher could also accept a .75 reliability coefficient. If, for reasons of practicality, the teacher is unable to use a test with one of the above levels of reliability; then the test with the highest reliability should be chosen from those available.

In order for a particular reliability coefficient to have complete applicability to a given testing situation, the group with whom the teacher plans to use the test must be similar in nature to the students who formed the population for the reliability study. Significant differences between the characteristics of these two groups will mean that less confidence can be placed in the coefficient of the test. When critical decisions are to be made, it is vital that some consideration be given to the similarities and differences between the two groups of students.

There are several techniques that may be used to determine the reliability of a test. The following are the most frequently used methods.

Test-retest. In this method the student is given the same test on two different occasions with a time interval of about one week

between testings. This technique accounts for errors that may arise as a result of the administration procedures and the test situation.

Split-half. This form of reliability is based on one administration of the test. The items of the test are divided, after administration, into two halves on an odd-even basis and the two total scores for each half are correlated to produce the reliability coefficient. Through this technique the test constructor attempts to account for errors that may arise as a result of the differences between the items of the test. There are some serious limitations to this method of determining reliability and the coefficients derived may often be spuriously high. A teacher should also consider the number of items that constitute each half. The greater the number of items in each half, the more dependable will be the reliability coefficient.

Kuder-Richardson formula. This technique is very similar to the split-half method and accounts for the same type of chance error. It is also subject to the same type of criticisms.

Equivalent form. Of the more commonly used methods of determining reliability, this one is the most suitable for most testing situations. In this method there are two equivalent forms of the test, each having the same number of items, covering the same subject matter and being of equal difficulty. The students are administered one form of the test and about a week or two later they are given the second form. The scores of the two forms are then totalled and correlated. The technique attempts to account for errors that may

arise as a result of item sampling, variations in the administration procedures, and changes in the student from one testing situation to another. Because it accounts for the greatest number of sources of chance error, the equivalent form method is considered the best way to determine the reliability coefficient of a test. Since it does account for several sources of chance error, the coefficients produced are often lower than what might be reported by similar tests using another reliability technique.

Standard Error of Measurement

An individual test score is not entirely the result of what the student knows about a specific skill. Some portion of the score can be accounted for by such factors as guessing, the student's familiarity with tests, distractions during the test, the student's physical and emotional state at the time of testing and several other conditions of testing. Since these circumstances are not predictable, they are termed forms of chance error and they may serve to add to the student's true score or to detract from it. Thus the portion of the student's score produced by these factors is called chance error score. To indicate the degree of chance error associated with a test, the test manual usually reports a statistic known as the standard error of measurement, or abbreviated as the "Sem". The standard error of measurement is usually given in terms of raw score points. Since the standard error of measurement may have a positive or negative influence upon the student's score, the use of it will result in a score

band within which the teacher can assume will fall the student's true score. The size of the score band is related to the degree of probability which the teacher wishes to accept when trying to estimate the true score.

1 Sem = 68% probability

2 " = 95% "

3 " = 99% "

Example: If a student obtained a raw score of 27 on a 40 item test, it could be stated that the score of 27 contains an estimate of what the student really knows about the subject as well as some chance error. The standard error of measurement for our test is 3 raw score points. It may now be said that with

68% probability the student's true score lies somewhere between 24 and 30

95% probability the student's true score lies somewhere between 21 and 33

99% probability the student's true score lies between 18 and 36.

The degree of probability which the teacher wishes to use will depend upon the nature of the decision that has to be made. The more serious the decision the more certain the teacher should be of knowing where the student's true score may be.

The standard error of measurement is directly related to the reliability of the test. The higher the reliability of the test, the lower

is the size of the standard error of measurement. It is an inverse relationship.

Some tests do not report a standard error of measurement. Consequently, the teacher has no means of estimating how much of the student's score may be due to chance factors.

Norms

When the teacher is interested in making comparisons between students or between an individual student and a group, he will require norms. If the teacher, however, is concerned with the assessment of the student's strengths and weaknesses in a particular discipline area, then norms will be of very limited usefulness. Teachers should only be concerned with the use of test norms when they are primarily interested in comparing the achievement of the student with that of another student or with that of a group.

Norms are not to be taken as representing some form of absolute standard of achievement in a skill. They are only an indication of how well a specific group of students performed on a test at a given point in time. They are not meant to demonstrate the most desirable level of achievement. If the teacher feels that the norming group is comparable to the type of students he is dealing with, then he may wish to accept the performance of the norming group as a standard for his students. It seems, however, more advisable to use the norms as a point of reference rather than as a set of achievement standards.

The utilization of the test norms is only valid when the group of

students tested is similar in nature to the students who constituted the norming population. If there are major differences between the characteristics of the two groups, the norms cannot be used because there is no longer a common basis for valid comparison. It is very important that the teacher examine the relationship between the two groups of students before making any comparison of scores. Other points to consider regarding the norming population are its size and the date when it served to produce the norms. The larger the norming population, the more likely it will be that the norms will represent the population for whom the test was designed. In some disciplines, the content of the discipline will change over a period of time as a result of new discoveries and insights. Changes may also occur in the nature of the population for whom a test was designed. These conditions are most likely to prevail when a test that was constructed ten or more years ago is used with a current group of students. Because of the changes that have occurred in that period of time, the norms of the test will not be representative of achievement in that discipline for present-day students. Changes in the nature of the population for whom a test was designed may occur in a much shorter time period as a result of a major alteration in teaching methods or sudden exposure to a new source of information.

Test Scores

A great deal of confusion and misunderstanding prevails among educators with respect to the interpretation and use of the various

types of scores reported by standardized tests. In this section, we will deal with some of the more commonly used scores and attempt to outline their strengths and weaknesses.

Grade equivalent score. Of all the scores reported in standardized tests, the grade equivalent score is probably the most widely used and the most frequently misinterpreted. Many teachers feel that the grade equivalent score is meant to indicate the grade level at which the student is functioning. The grade equivalent score cannot be interpreted in this manner. To understand the meaning of this score, one needs to know how it is derived. If the test norms range from grade 2 to grade 12, it is assumed that the test constructor has administered the test to a group of students from each of these grades. The scores for each grade are then rank ordered and the raw score that divides a grade group in half is considered the grade equivalent score for that grade level. If a raw score of 34 divided the grade four sample in half, 34 would become the raw score for the grade equivalent score of 4.0. The same procedure is used for each grade included in the test norms. The months between each grade equivalent score are statistically computed and assigned to specific raw scores. Thus, it can be seen that a grade equivalent score does not refer to the subject matter of a particular grade level but only to the average score obtained by a group of students on one test. If a grade three student achieves a grade equivalent score of 6.4, it means that he did as well on this test as the average grade six student in his

fourth month of school. It does not mean that the grade three student is capable of coping satisfactorily with subject matter at the grade six level. Unfortunately many teachers interpret the grade equivalent score in the latter sense. Frequently the test constructor has not administered the test to all the grades for which norms have been established. A portion of the norms will be based on actual test results and the rest of the norms will be statistically arrived at. Several authorities on testing have stated that there are a number of serious limitations associated with the use of grade equivalent scores. Because of the gross misinterpretations that occur with the use of this score and because of its almost total uselessness in diagnostic teaching, teachers would be well advised to avoid reporting test results in terms of a grade equivalent score.

Percentile score. The percentile score does not represent the percentage of correct responses obtained by a student on a test. The percentile score relates to the percentage of students that are at or below a given raw score and indicates the position held by a student with respect to other students to whom the test was administered. If a grade five student obtained a raw score of 43 and this raw score translated to a percentile score of .74, it would mean that the student had surpassed 73 percent of the students on whom the test was normed and was himself positioned at the 74th percentile rank in terms of the norming group. The percentile rank score reveals where the student stands with regard to the scores obtained by the students on whom the

test was normed. Depending upon the appropriateness of the test being used, the percentile score indicates to the teacher whether the student is doing well or poorly in the particular subject matter. The percentile score can also be used to compare the student's scores on different subject matter tests as long as the norming populations for the tests are the same. The grade equivalent score cannot be compared in this manner. A student with a reading grade equivalent score of 3.4 and a math grade equivalent score of 5.8 would appear to be a more capable student in math than in reading. It is possible, however, that the percentile score in reading might be .84 while that for the math might be only .70. By the use of the percentile score it can be seen that the student is achieving more in reading than in math. The percentile score provides a more understandable basis for making comparisons. Parents and teachers are also less inclined to misinterpret a student's test result when it is reported as a percentile score.

Stanine score. The stanine score is a nine point scale that is closely related to the percentile score scale. In the case of the stanine score, the percentile range of 0 to 100 has been divided into nine units. The following table lists the stanine, the number of percentile ranks in each stanine and the percentile band that each stanine represents:

stanine 1 = 4 = 0-4%	stanine 2 = 7 = 5-11%
stanine 3 = 12 = 12-23%	stanine 4 = 17 = 24-40%

stanine 5 = 20 = 41-60% stanine 6 = 17 = 61-77%

stanine 7 = 12 = 78-89% stanine 8 = 7 = 90-96%

stanine 9 = 4 = 97-100%

The advantage of using the stanine system of scoring is that it gives a score band that allows for the chance error that will arise in the administration of the test. Like the percentile score, the stanine score indicates to the teacher the position of the student relative to his peers in the subject under consideration. Although some test manuals state that the stanine system contains nine equal steps, the teacher is not to interpret this to mean that each stanine represents the same amount of achievement. The description of the stanine scale as a system of nine equal steps pertains to the statistical nature of the scale and not to the level of achievement represented by each stanine. Generally the first three stanines are considered to indicate below average achievement, the middle three to indicate average achievement and the top three to indicate above average achievement. The utility of this classification system for a given group of students will depend upon how closely the students resemble the norming group and how well the content of the test matches the content which the teacher wishes to examine. The stanine scale may be effectively used as a method of completing an initial grouping of students.

Questions to Ask About a Test

Skills

1. What skills does the test purport to measure?
2. Does it really measure these?
3. How is each skill evaluated?
4. Is it evaluated in the same manner in which it is used in the reading act?
5. Does each subtest have enough items in it to give a valid indication of the student's ability in that skill?
6. Does the manual explain how the test items were selected?
7. Are there strict time limits for the subtests?

Scores

1. What type of scores are reported?
2. Is there a separate score for each subtest?
3. Could guessing be a major factor in the student's score?
4. How reliable is the test?
5. Is there a standard error of measurement reported for each subtest?
6. Does the manual provide an interpretation of each subtest score?
7. Does the student have to complete the entire test or,

subtest before a score can be obtained?

8. How difficult is the test to score?

Norms

1. When were the norms developed?

2. What type of students were used to form the norms?

3. How many students were in the norming group?

Other

1. Will this test give me the type of information I want concerning the student?

2. For what grade levels is the test intended?

3. How long does it take to administer?

4. How difficult are the student's directions?

5. Is it a group or individual test?

6. How much does the test cost?