THESIS

This is to certify that the

thesis entitled

Differences Between Learning Disabled and Low Achieving Students on Standardized Test Performance

presented by

William W. Wiseman

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Special Education

Eugene Pernell, Jr.
Major professor

Date 2-16-84

O-7639

DIFFERENCES BETWEEN LEARNING DISABLED AND LOW ACHIEVING STUDENTS ON STANDARD TEST PERFORMANCE

By

William W. Wiseman

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology,
and Special Education

1984

ABSTRACT

DIFFERENCES BETWEEN LEARNING DISABLED AND LOW ACHIEVING STUDENTS ON STANDARDIZED TEST PERFORMANCE

By

William W. Wiseman

The ability and achievement test performance of 404 students selected from a sample of 1,552 students from 45 school districts throughout Michigan was analyzed to determine if methods proposed in the literature could assist in formulating useful decision-making rules for classification purposes. Two hundred thirty students classified as learning disabled were compared with 174 low achieving students selected on the basis of their achievement scores on a test of word recognition. The methods investigated in this study for diagnostic-utility purposes have been proposed in the literature as criteria for differentiating between handicapped (learning disabled) and nonhandicapped students. Methods of z-score discrepancy levels, ability-test subtest scatter, verbal-performance ability differences, and general ability and achievement characteristics have been proposed as differentiating between classifications. Previous studies comparing learning disabled with low achieving students have been limited in sample size and have suffered from inadequate controls.

In the present study, the question of diagnostic utility was of primary concern. Any method used to compare students that failed to differentiate at an 85% success rate was rejected as not useful. None of the methods investigated met this criterion for diagnostic utility and useful decision-making rules. The methods used to calculate discrepancy levels and ability-test subtest scatter indicated statistically significant differences at the .05 level of confidence. However, these methods were found not to be useful for practical diagnostic purposes in differentiating between learning disabled and low achieving students. Conclusions and recommendations were discussed in terms of public policy, clinical practice, directions for educational research, and limitations of the study.

ACKNOWLEDGMENTS

Each of the members of my guidance committee has contributed from his separate area of expertise to the completion of this work. I greatly appreciate their efforts.

Harvey Clarizio, my dissertation chairman, has for a long time been able to understand my strengths and weaknesses and appeared always, at the right moment, to provide direction, criticism, and encouragement. His patience and expertise in the completion of this research are without equal.

Donald Melcer, in his objective and direct communication style, has assisted me with editorial comments and always responded to questions with direct and practical answers.

Lawrence Lezotte is appreciated for his efforts in helping me understand statistical aspects of this research and practical applications.

Eugene Pernell served as my program advisor and with openness and sincerity affected my completion of program requirements.

Two very special people are my daughter and my wife. My daughter, Lydia Wiseman, has been patient and understanding when circumstances dictated efforts on this work. Her interest, enthusiasm, and excitement have made it worthwhile. More than anyone else, my wife, Judith Wiseman, has made this work possible. Her positive attitude,

support, love, and understanding are always present. Her questions, support, assistance, and belief in my abilities have enabled me to finish this work and, more important, affect all areas of my life.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Classification of Students

A major purpose and concern of school systems and educational personnel is classifying students. Students are classified on the basis of chronological age (by grade) in elementary school and by age and subject matter at the secondary level (Shinn, Ysseldyke, Deno, & Tindal, 1982). Schools also classify students according to a variety of special needs (i.e., mental impairment, speech and language disorder, emotional impairment, physical impairment, and learning disability). Classification by handicap is intended to benefit the student and is based on premises similar to those used in classifying regular students. Providing homogeneous disability groupings theoretically makes it possible to teach compensating techniques. It enables the teacher to focus on processes that help students acquire information and allows school administrators to provide equitable and appropriate programs for the handicapped. Although the practice of formal classification of handicapped children was originally intended for exclusion, when Binet was commissioned to "find a way to locate those who could not learn so that teachers would not be charged with failure on their account" (Maurer, 1972), the intention of present diagnostic practice is to benefit all students.

Identification of the Handicapped

Individuals and professional groups have for some time been striving to arrive at specific classification schema for the handicapped. A major focus of this effort has been to identify essential characteristics that differentiate handicapped student populations so that specific curricula can be applied to them. This type of sorting process has resulted in two broad applications: classification by disability (e.g., mental subnormality versus learning disability) and classification by severity (e.g., mild versus severe). In both cases, diagnosticians have used standardized-test results to help differentiate disability groups. Identification of some disabilities is a relatively straightforward process. For example, classifying a student as blind depends on the absence of sight. Other handicapping conditions pose more difficult diagnostic challenges. Mental retardation can be differentiated from learning disabilities (Frame, Clarizio, Porter, & Vinsonhaler, 1982); however, differentiating those with specific learning disabilities from low achievers is more difficult. Indeed, the literature reviewed in the next chapter suggests that methods used to differentiate a student with a learning disability from one with other handicaps and particularly from a student who is found not to be handicapped but who achieves at a low academic level may not be of diagnostic utility.

Defining the Learning Disabled

In the 1940s, Strauss and Lehtinen developed procedures to identify and teach exogenous and endogenous mentally retarded students. In the 1950s and early 1960s, Cruickshank and his colleagues developed procedures to identify brain-injured students. In the early 1960s, Kephart developed procedures to identify and teach "slow learners," and Kirk coined the term "learning disabilities" to refer to students of essential average intelligence who failed to make adequate progress in school. In the mid-1960s, the federal government established the category "learning disabled" as a handicapping condition.

All along, the presumption has been that a specific group of students who were failing in school could be readily identified and that identifying them would lead to the development of appropriate educational programs. These students were presumed somehow to be different from other students who were failing because of low achievement.

The federal government has established rules for determining a learning disability, in an attempt to develop guidelines for serving students with this type of handicap. Now, a team of professionals may determine that a student has a specific learning disability if there is one or more low areas of achievement, if there is a discrepancy between ability and achievement, and if the student exhibits an implied processing disorder (this requirement is no longer mandatory, although it is included in many school definitions).

To comply with the federal definition of learning disability, diagnosticians and school personnel continue with predefinition emphasis on three major factors or characteristics in the assessment of students, using primarily standardized tests. The existence of a process disorder, significant nonachievement in one or more academic areas, and a discrepancy between ability and achievement are purported to establish the presence of a learning disability. These characteristics may exist singularly or in combination, as long as the discrepancy criterion is met.

Theoretical Basis for Diagnosis

Although both learning disabled and low achieving students may meet the criterion for low achievement, learning disabled students have been thought to be unique in how they score on ability and achievement tests. Researchers have concentrated on identifying behavior patterns exhibited on tests that could reliably be associated with the learning disabled student (Sheppard, Smith, & Vojir, 1983). Two focal points with regard to identification have been prevalent: process disorders and ability-achievement discrepancy.

Process disorders. The concept of process disorders is based on the assumption that students possessing a learning disability are deficient in cognitive ability and that standardized intelligence tests assess accurately the type and functional degree of such ability. For example, some authorities have argued that, if a student exhibits a discrepancy between verbal and performance IQ abilities, there is a likelihood that a learning disability exists (Bateman, 1965). The student may also exhibit a degree of scatter between subtests in the verbal and performance ability areas, which, depending on the degree of variance, implies that the student has a learning disability (Clements, 1966; Matarazzo, 1972). This approach to identification was influenced by research that emphasized brain-behavior relationships and neurological dysfunctions. If different parts of the brain mature at diverse rates (Bender, 1963; Chusid, 1979; Cowan, 1979; Geschwind, 1964), a pattern of skills and deficits results. Additional concepts such as developmental imbalances (Gallagher, 1966) have provided further impetus for the development of verbal- versus performance-ability comparisons and subtest scatter analyses.

Ability-achievement discrepancy. The field of learning disabilities paralleled the remedial-reading field in its attempt to differentiate learning disabled students from "slow learners." In 1932, Monroe noted that "a child may fail to learn to read and yet be of adequate intelligence" (p. 1). Bond and Tinker (1957) also highlighted the concept of underachievement, stating that "the disabled reader is a child . . . who is not living up to his potential as a learner" (p. 83). Reading specialists emphasized identifying the "disabled reader" (achievement less than computed expectancy) rather than the "poor reader" (low achievement but supposedly achieving up to computed expectancy) (Epps, McGue, & Ysseldyke, 1982).

Bateman (1965) and Ysseldyke and Algozzine (1979) emphasized the use of ability-achievement discrepancy with regard to identifying students with learning disabilities. Myklebust (1968) and Algozzine, Forgnone, Mercer, and Trifiletti (1979) developed formulas to be used in calculating ability-achievement discrepancies. The current federal formula was designed to identify those students who were doing poorly in school but yet were of average intelligence. No consensus exists among authorities about whether the ability-achievement discrepancy method is useful in differentiating the learning disabled from the low achieving student who has not been classified as handicapped. Recent research (Algozzine & Ysseldyke, 1982b; Epps, Ysseldyke, & Algozzine, 1982a) has indicated that learning disability may be a category of low achievement. Epps, Ysseldyke, and McGue (1982) focused on the use of standardized achievement scores without the use of a formula to identify students who are doing poorly in school, regardless of their intelligence.

Magnitude of the Problem

Despite the sizable number of students being declared eligible for learning disabilities services and the implicit appeal of the term (Divoky, 1974; Hobbs, 1975), there is little agreement on the definition of learning disabilities. Vaughan and Hodges (1973) reported as many as 38 different definitions of learning disability, and Mercer, Forgnone, and Wolking (1976), in a survey of 42 state departments of education, found considerable variation in definitions of the term. Thurlow and Ysseldyke (1979) reported that Child Service Demonstration centers for learning disabled students varied considerably in how students within those programs had been identified as handicapped. This variance of definitions has been accompanied by a wide range of estimates as to the prevalence of the handicap in the school population. National estimates range from 1 to 30% (Lerner, 1976), to as much as 70% (Tucker, Stevens, & Ysseldyke, 1982).

The actual number of students served in learning disabilities programs has increased dramatically in the past decade. During the 1969-70 school year, .02% of the total school population was served in programs that were analogous to current learning disabilities programs. By 1977 that figure had risen to 5.2% (Epps, Ysseldyke, & Algozzine, 1982b). Tucker (1980) reported that in one state the percentage of students identified as learning disabled rose almost 44% from 1970 to 1977. During the 1981-82 school year, the percentage of the total school population served in learning disabilities programs was 3.01%. Although the overall percentage of students in learning disabilities programs has decreased since 1977, 1,455,135 students were served in 1982, which was 34.7% of the total handicapped population (USOE, 1982).

In contrast to the 3.01% incidence figure for students with learning disabilities, it is believed that there is a large population of low achieving students who are not identified as handicapped but who exhibit test characteristics similar to those of learning disabled students. Up to 33% of the school population may be perceived as low achievers (Warner, Schumaker, Alley, & Deshler, 1980). A strong possibility of misdiagnosis exists when these two populations cannot be specifically differentiated from each other through the use of currently popular methods.

Of equal importance are considerations of teacher training, diagnostic and placement costs, and curriculum modifications, which need to be addressed in providing educational programs for this large but undiagnosed or misdiagnosed population. School personnel may need to focus their attention on broader issues than the student's academic performance. Situational variables such as the teacher, peer group, educational content, and method of instruction used in the student's classes may need to be examined more systematically (Lidz, 1981). The system of allocating funds from a categorical basis, along with major changes in legislation and regulations to allow school districts flexibility in demonstrating prevention techniques and in solving problems of the entire school population, may need encouragement (Reynolds & Wang, 1981).

Synopsis of the Present Investigation

The validity of differentiating the learning disabled from the low achieving student who has been deemed not handicapped is far from established. Using general ability and achievement-test performance, differences between verbal and performance ability levels, degree of discrepancy between ability and achievement, and analyses of subtest scatter may not be accurate methods with credible diagnostic utility (Epps, Ysseldyke, & Algozzine, 1982a). If such is the case, using these methods to validate the existence of a learning disability apart from low achievement is counterproductive to the intention of diagnosis and intervention.

The primary concern of this investigator was to determine whether there is practical value in using test results in the aforementioned ways as a diagnostic rationale in identifying students as learning disabled and not identifying others who share a common characteristic--low academic achievement. Current studies that have made efforts in this direction have defined low achievement as scoring below a specific point on achievement tests. Epps, Ysseldyke, and Algozzine (1982a) used scores below the 25th percentile, as did Algozzine and Ysseldyke (1982a) and Shinn et al. (1982). In a study involving junior and senior high school students, Warner et al. (1980) found sufficient numbers of students apart from special education programs scoring below the 33rd percentile and defined them as low achievers.

The present study compared students identified as learning disabled with those who had been identified as not possessing any discernible handicap that qualified them for special education services. The group of students who did not qualify for special education services scored below the 26th percentile on a test of word recognition and are designated in this study as low achievers (LA). The learning disabled and low achieving groups were compared using two achievement standards: the learning disabled (LD) with the low achieving (LA) group who scored below the 26th percentile (LA<26th) and the learning disabled (LD) with the low achieving (LA) group who scored below the 17th percentile (LA<17th).

The students included in this study comprised a broad spectrum of grades and ages. All students were referred for diagnostic evaluation after experiencing difficulty in school and were judged by a team of professionals to be either learning disabled or not eligible for special education services. The sample size was sufficiently large ($\underline{N} = 404$) to offset statistical errors produced by exceedingly small samples. The influence of moderator variables on any differences between groups will be reported. Using different methods to assess group differences in ability-test performance may indicate which method is most accurate in affirming classification decisions. The extent to which the methods used in this study affirm classification decisions should prove of value to diagnosticians and educational personnel responsible for making such decisions. This study may also assist in delineating test-performance characteristics that would be helpful in establishing realistic diagnostic criteria for classification decisions.

CHAPTER 2

LITERATURE REVIEW

Overview

The preceding chapter outlined a history of attempts to verify the existence of learning disabilities apart from other handicaps and the nonhandicapped students. The general trend of using discrepancy models to calculate ability-achievement differences, verbal-performance ability differences, and ability-test subtest scatter analyses was presented. Of particular note was the fact that the definition of learning disabilities established by the federal government and used by school districts contains a strong element of low achievement. This chapter reviews the literature in the following areas: similarities and differences between populations, characteristics of the learning disabled, general ability and achievement characteristics, use of a discrepancy model, effectiveness and reliability, verbal-performance ability differences, and ability-test subtest scatter comparisons. The focus is on studies comparing LA with LD students. The final section of the literature review points to the direction of this study.

Similarities and Differences Between Populations

Diagnostic models designed to differentiate between the LD and the LA student appear to be predicated on the belief that actual differences in ability and achievement exist and/or that the LD student

uses abilities and skills in different ways from other students. Traditional methods of assessing students for classification purposes have relied on standardized test results, with the notion that if differences exist they can best be measured by tests that reflect both student ability and academic skills learned.

Similarities and differences between LD and LA populations have been studied in a number of ways. Warner et al. (1980), in calculating discrepancy levels between achievement and ability with junior and senior high school students, reported that no significant differences were found between LD and LA populations. Others (Algozzine & Ysseldyke, 1982a; Epps et al., 1982b; Shinn et al., 1982) have indicated a lack of consistent direction for distinguishing between the two populations in terms of processing information, scoring on ability subtests, degree of ability-achievement discrepancy, and verbal-performance IQ scores. As Epps et al. (1982) indicated, "a major difficulty with studies attempting to identify the salient characteristics that distinguish the learning disabled from other students is that they compare with populations that are obviously different" (p. 209). The LD population represents a heterogeneous group characterized by a variety of different behavioral characteristics (Lerner, 1976; Mercer, 1979). Algozzine and Ysseldyke (1982b) indicated that the "only common school related characteristic among many learning disabled children is low achievement of some form" (p. 117).

A major study that compared such similarities was conducted by Ysseldyke, Algozzine, Shinn, and McGue (1982). Comparison of 50 fourth-grade students who had been identified as LD with 49 LA students at the same grade level on a number of analyses indicated few differences. Students were administered a test battery including the WISC-R, the PIAT, subtests of the Stanford Achievement Test, a perceptual motor test, and a self-concept scale. Methods for comparing the two groups included measuring the discrepancy between ability and achievement, comparing subtest scores in an overlap analysis, and contrasting performance in the domains of cognition, achievement, perceptual-motor functioning, and self-concept. The results of the study indicated that no psychometric differences of practical utility were observed. The researchers concluded that "there is considerable misclassification in identification of LD students" (p. 83). They failed to identify psychometric differences between the two groups on test performance.

Characteristics of the Learning Disabled

Analyses of the characteristics of LD students have been conducted, and the results have shown considerable variation among students who have been identified as LD. Norman and Zigmond (1980) analyzed data from the files of 1,966 students, ages 6 to 17 years. The case studies were selected from populations of LD students from 24 school districts in 22 states. The focus of the study was on examining characteristics of ability and achievement in terms of variability in range and level, as well as indications of what percentage of the sample did not demonstrate empirical evidence of a learning disability. Two methods were used to support evidence that a student could be classified as LD: the federal formula used to calculate a severe discrepancy and a learning-efficiency model (Zigmond, 1978). The learning-efficiency model used a formula that yields a percentage score after actual achievement is divided by expected achievement (C.A. x IQ/100-5.0). Students were considered to be correctly classified if their severe discrepancy level (federal formula) was 50% or greater or if their learning efficiency was less than 50%.

The authors found that age appeared to be negatively related to IQ level. In a majority of the school districts (54%), there were students classified as LD who had reading and math scores almost at grade level. Below-grade-achievement discrepancies of two or more grades were not evident until a chronological age of 12. Learning Efficiency Rates (percentage of actual achievement divided by expected achievement; C.A. x IQ/100-5.0) of the total sample were 68.1% in reading and 74.5% in math. Norman and Zigmond found that less than half of the populations studied were underachieving.

More recently, Shepard et al. (1983) studied the test results of 790 LD students selected from 22 special education administrative units in Colorado. The students' files were used to obtain IQ data, and the degree of discrepancy between ability and achievement in math and reading was calculated. The authors used a standard-error-of-difference correction with both a lenient and a strict criterion to calculate discrepancy between ability and achievement scores and the degree of variation between students' verbal-performance IQ scores. Results of the study indicated variation in IQ levels with 28.5% of the population having IQs below 90 and 8.3% having IQs below 81. Thirty-five percent of the students had no significant discrepancies between ability and achievement. However, in 45.2% of the cases, the files contained insufficient data for these analyses. Achieving below grade level as a consistent factor in differentiating LD students from the rest of the population was not validated in the study. In preschool and kindergarten, the achievement average was above grade level, whereas in primary grades fewer than half of the students studied met the below-grade-level criterion. In summary, Shepard et al. stated that less than half of the sample had characteristics consistent with the definitions of LD students cited in federal statutes and in professional publications.

General Ability and Achievement Characteristics

Studies comparing the overall general characteristics of the LD population with those of a LA population appear to have been limited to students at the elementary-school level. The major exception is the research conducted by Warner et al. (1980), who compared the test results of junior and senior high school students in an attempt to discover if differences existed on tests of ability and achievement. In reporting the results of their research comparing 234 LD students with 220 low achievers, the authors noted several differences between the two groups. Scores on the reading, mathematics, and written-language subtests of the Woodcock Johnson Psychoeducational Battery indicated that the LD group scored significantly lower than the low achievers at both junior and senior high school levels. Scores from selected subtests of the Wechsler Battery (used to predict estimated IQs) indicated that the LD group scored significantly below the LA group at both junior and senior high school levels. Both the LD group means (senior high--93.02, junior high--87.9) and the LA group means (senior high--99.08, junior high--98.99) were below average. The authors indicated that "the traditional characteristics of LD students as having average ability would not characterize a substantial proportion of the students in our sample" (p. 31). They noted that the results, particularly in the achievement area, may have been due to sample selection. The LA group had been selected on the basis of scoring below the 33rd percentile because it had been difficult to find a sufficient number of students who scored below the 25th percentile on achievement tests.

Taylor, Satz, and Friel (1979) compared the test performance of 80 LD students with 80 students who were not diagnosed as handicapped but who exhibited poor performance in reading on a locally normed teacher judgment scale and the IOTA (Monroe, 1928). They used Peabody Picture Vocabulary Test (Dunn, 1965) results to control for IQ level in a sample selected from 570 second-grade boys. Taylor et al.'s results indicated that the nonlabeled students could not be distinguished from the students labeled LD.

Ysseldyke et al. (1982) compared the performance of students labeled LD with students who scored below the 25th percentile on the Iowa Test of Basic Skills. Test results from the WISC-R, the Woodcock Johnson Tests of Cognitive Ability, the Peabody Individual Achievement Test, the Stanford Achievement Test, and the Woodcock Johnson Tests of Achievement were compared for 50 LD and 49 LA fourth graders. Although there were statistically significant differences between groups on tests of achievement, the authors noted that the real difference between the means was only 1.26 points, which was the result of answering one more math problem correctly. When identical scores were computed, at least one-half the number of both groups compared had the same scores.

Shinn et al. (1982) compared the performance of 34 LD students with that of 37 students who were not identified as handicapped but who scored below the 26th percentile on the Iowa Test of Basic Skills. These 71 fifth graders were administered measures of reading (word lists), spelling (word lists), and written expression (vocabulary words), which were constructed and used in identical formats with both groups over a 5-week period. The test results were compared for overall differences, test-area differences, and learning rates on a week-by-week basis for each test area. The non-LD group significantly outperformed the LD group in reading words correctly. Reading differences between the groups ranged from 23.6 to 26.1 words per minute. However, there was no difference in the groups' learning rates in reading. On an average, the LD group increased at a rate of 1.74 words per week, whereas the non-LD group increased at a rate of 2.35 words per week. In spelling, the non-LD group spelled more words correctly, with a 6.7 to 8.6 word advantage. Written expression differed significantly between groups; the non-LD group outperformed the LD group by an average of approximately seven words per week. Absolute growth in all academic areas indicated no differences between the groups. Contrary to all other findings, the LD group excelled in the learning rate of written expression. They gained an average of one word per week, whereas the non-LD group lost an average of one-half word per week.

Of particular note in comparing the results of the Shinn et al. (1982) and the Ysseldyke et al. (1979) investigations is that both studies used the same student samples. Shinn et al. (1982) found differences in word-recognition, spelling, and written-expression skills between the LD and the LA groups (measured by nonstandardized tests), whereas Ysseldyke et al. (1979) found no differences between the groups on standardized test performance. Shinn et al. suggested that such differences in findings might have been a result of the indirect measurement of skills done by standardized tests and that, indeed, differences between LD and LA students do exist. Given their findings, the authors seemed to indicate that standardized testing fails to substantiate differences between LD students and others who may be doing poorly in academic tasks.

The issue of competency versus performance may arise when standardized tests are used to measure differences between groups for the purposes of classification. If differences between LD and LA students are not found on standardized test scores but are seen by the classroom teacher in competency-based tasks, the methods used for diagnostic purposes need clarification. Are students identified as LD on the basis of performance, or, as discussed elsewhere in this chapter, are they identified by focusing on the characteristics of the diagnostic process? If so, focusing on the characteristics of the diagnostic process may be due, in part, to the belief that what matters is not actual differences in ability and skills but in how students perform in the classroom.

Operationalizing the Discrepancy Model

Although Birch and Belmont (1964), Browning (1967), and Schwartz and Bryan (1971) were among those who emphasized the importance of the IQ as the major determinant of learning disabilities, a number of researchers have endorsed the notion of comparing differences between ability and achievement. As early as 1932, Monroe noted that a child may fail to learn to read and yet have adequate intelligence. Bond and Tinker (1957) also highlighted the concept of underachievement, stating that the "disabled reader is a child . . . who is not living up to his potential as a learner."

Bateman (1965) expressed the view that a realistic quantifying method is necessary to clarify the extent of learning disability, and such authors as Myklebust (1968) and Algozzine et al. (1979) have developed formulas similar to those used in calculating reading expectancies.

Various methods have been used to compare ability and achievement. Schere, Richardson, and Bialer (1980) proposed a formula that averaged grade scores from group- and individually administered achievement tests. Lerner (1981) compared four methods of quantification and cautioned that different ages give different indications of a discrepancy.

The present uses of discrepancy models vary greatly, depending on school districts' policies for establishing special-program cutoff points. However, all such models attempt to meet the federal requirement that achievement be significantly lower than ability, regardless of other criteria used in determining whether a learning disability exists.

At present there is no empirical evidence that a discrepancy model can differentiate between the LD student and the low achiever. Brenton and Gilmore (1976) reviewed the discrepancies between ability and achievement of 60 students and found that 67% were identified as handicapped. They noted that considerable caution should be exercised when classifying children, especially females, as LD. A recent investigation by Epps et al. (1982) comparing the performance of 48 LD and 96 nonidentified students on 10 different discrepancy formulas had varied results. The investigators compared student performance using six ability-achievement methods based on test results and four below-grade-placement achievement methods. The most successful method of affirming classification decisions was the below-grade-placement method, which used a standard score cutoff of below 86 points on achievement tests, including the Woodcock Johnson Test of Achievement subtests (reading, mathematics, and written expression) and the Peabody Individual Achievement Test subtests (mathematics, reading recognition, and reading comprehension). Seventy-eight percent of the LD and 35.2% of the nonidentified group scored below the cutoff point. Application of the federal formula to test results indicated that 25% of the LD and 4.4% of the nonidentified group met the criterion of a significant discrepancy between ability and achievement. The authors concluded that there is considerable doubt that school personnel can accurately and reliably identify LD students using the methods included in their study. None of the definitions used was in high agreement with the schools' classification of students as LD or non-LD. Approximately 18% of the schools' identified LD students were not identified by more than one definition. The authors indicated that: "As it is presently conceptualized, the category of learning disabilities is an ill-defined disorder with little consistency among definitions to allow for reliable prediction of LD classification" (p. 20).

Attempts to validate discrepancies between ability and cognitive functions, other than those assessed by achievement tests, have yielded inconsistent results. Burns (1977) studied the relationship between abilities assessed by the ITPA and found there was a nonlinear relationship that lends little support to the credibility of this method.

Other researchers have compared ability with achievement in various ways to ascertain if previously diagnosed LD children are different from the LA population and from randomly sampled populations that are comparable in age, sex, and socioeconomic status. Algozzine and Ysseldyke (1982a) compared 50 children diagnosed as LD with 49 students who were low achievers. The students' test results were compared on 16 discrepancy models, including the current and alternative federal formulas, statistical ceiling models, standard score cutoff models, and learning-quotients models. The findings indicated that no one model or formula was capable of differentiating the LA student from the LD. The results indicated a range of differential classification from 0 to 15%; no child was classified on all definitions. In addition, Algozzine and Ysseldyke found that 92% of the LA population met the criterion for LD classification on at least one definition.

Epps et al. (1982) used a larger sample that included randomly selected students who were not referred for evaluation, students who scored below the 25th percentile on a standardized group achievement test, and students diagnosed by school districts as LD. The results offered little support for the usefulness of the federal definition. Comparison of the populations in grades 3 through 5 focused on six forms of ability-achievement discrepancy and four forms of grade-placement discrepancy. The results indicated there is a substantial lack of agreement and uniformity in applying the federal model, with no clear direction toward a definition that distinguishes the LD from other students.

Discrepancy-Model Effectiveness

Of major concern in evaluating the effectiveness of a discrepancy model are the direction and ceiling at which the majority of the given population fits the assigned criteria. Algozzine and Ysseldyke (1982b) compared 40 LD students and a matched number of students scoring below the 25th percentile on a standardized achievement test on three alternate methods of diagnosing LD students using the current federal definition (cutoff points at the 1.0, 1.5, and 3.0 standard deviation levels). Using any one subtest from the Woodcock Johnson Achievement Test or the Peabody Individual Achievement Test (subtests--mathematics, reading, spelling) and the WISC-R Full Scale IQ, the authors found inconsistent results. Of the students in the sample, eight met no identification criteria on any definition. The criteria for identification were met by 92% of the LA group on at least one formula. More LD students met the criteria on four definitions, yet more LA students met the criteria on two other definitions. Of the entire sample, 73% met the identification criteria on the least-stringent ability-achievement model with a standard score difference of 10 or more points.

Warner et al. (1980) reported similar results in their comparison of adolescent LD and LA students. Although their cutoff level for identifying LA students was higher than that used in other studies (33rd percentile or below), the results were similar to those reported by Algozzine and Ysseldyke (1982a).

Discrepancy-Model Reliability

Of particular importance is the statistical reliability of any method used for differential diagnosis. Shepard (1980) commented on the four most commonly used diagnostic methods and advocated the regression method as the most statistically reliable, stipulating that scores be secured from co-normed tests. She indicated that the Harris (1970) method (determining expected grade equivalent in reading based on ability by subtracting 5 years from the student's mental age) fails to consider the less-than-perfect correlation between ability and achievement. Because of regression effects, the Harris method tends to identify more bright than dull students as LD. The Bond and Tinker (1967) method (years in school x IQ/100 + 1.0) assumes that the IQ score obtained from standardized tests is a ratio scale of measurement. The method does not consider that the variance in grade-equivalent performance increases with grade level. The *z*-score method proposed by Erickson (1975) assumes that achievement and IQ are in the same relative position as measured by standardized tests. This method (the *z*-score is obtained by subtracting achievement discrepancy from ability discrepancy after converting by: raw score - group mean divided by standard deviation) is based on the expectation that the achievement score is in about the same relative position in the distribution as the ability score and that regression of measurement is certain if a one-to-one correspondence does not exist. Shepard (1980) listed several advantages of the regression method over other methods. As previously mentioned, she stipulated that both the achievement and ability tests used to calculate discrepancies should be co-normed to avoid errors due to differences in standardization populations. However, the practical value of this provision is questionable. At present, diagnostic tests that do not meet this criterion (WISC-R, PIAT, Woodcock Johnson, WRAT, and others) are widely used.

Reynolds (1981) affirmed Shepard's (1980) criticism of grade-equivalent methods, saying that they ignore the dispersion of scores around the mean and that the regression between grade and test score is not equivalent across grades or school subjects. Reynolds did indicate, however, that the *z*-score method is satisfactory for indicating that real differences between ability and achievement exist, as long as the reliability of the test instrument is included as a factor in the calculations. Correction for differences in test reliabilities adjusts for differences due to chance or error of measurement, and calculated *z*-scores indicate real differences in performance. Reynolds indicated that the real differences between *z*-score performance are divided by the square root of combined test reliabilities subtracted from one:

$$\frac{\text{Absolute difference between } z\text{-score (ability)} - z\text{-score (achievement)}}{\sqrt{(1 - \text{reliability}) + (1 - \text{reliability})}} = z\text{-score discrepancy}$$

The resultant *z*-score discrepancy is then compared to the normal curve for statistical significance. "The values of *z* required at three commonly employed levels of significance for a one-tailed test are: 1.65, $p = .05$; 2.33, $p = .01$; 3.08, $p = .001$. For a two-tailed test, the values are 1.96, $p = .05$; 2.58, $p = .01$; 3.28, $p = .001$" (p. 354). When the test reliabilities of the instruments used fall within certain ranges, Reynolds indicated that specific cutoff standard deviation levels that indicate discrepancy levels meet statistical reliability as long as the standard deviations for each test instrument are equated by comparison of equal (or converted to equal) standard scores. For example, if the test reliabilities from which scores are obtained fall between .90 and .95, a *z*-score difference of .66 standard deviations is significant. Reynolds presented a conversion chart that facilitates the equating of standard scores when these scores are not equal.

Despite psychometric cautions, the emphasis on quantification of differences between ability and achievement appears to be gaining momentum. Warner et al. (1980) and Algozzine and Ysseldyke (1982b) indicated a trend by school districts to refer and place those students who are the lowest achieving but not exhibiting any other discernible handicaps.

The question of performing comparisons to differentiate between LD and LA students still needs to be addressed, particularly with a method that has the most statistically reliable basis. If achievement is a function of learning disabilities, an examination of differences between ability and achievement may help make appropriate classifications. When statistically significant differences between ability and achievement exist, it would appear to be important to note the percentage of the nonhandicapped group identified as having significant differences, as compared with the LD group. Also, the method used to determine differences between ability and achievement should be practical. Reynolds' (1981) method would appear to meet both of these criteria because it offers a statistically reliable calculation that can be used to indicate the significance of discrepancies. The method also adjusts for the differences between standardization distributions.

Obtaining a statistically significant discrepancy between ability and achievement does not necessarily mean the student needs special education intervention. The literature indicated that there may be a sizable population of students who are low achieving (achievement not commensurate with measured ability) who have not been referred for diagnostic evaluation but who, upon evaluation, demonstrate significant discrepancies between ability and achievement (Epps et al., 1982; Warner et al., 1980).

Comparisons of Verbal-Performance Ability Differences

Early attempts to differentiate the LD from the rest of the population on the basis of Verbal-Performance IQ test results appeared promising. Diagnoses were based on specific differences in test scores. Ackerman, Peters, and Dykman (1971) indicated that "discordance" among a LD sample was more frequent than in a "normal" sample. Anderson, Kaufman, and Kaufman (1976) reported statistically significant differences in test scores between children diagnosed as LD and the "normal" population. However, they did not find that children in their sample exhibited unusually large discrepancies, despite the statistical significance obtained. Piotrowski (1978) conducted observations that apparently verified Kaufman's (1976b) findings, although it became apparent that such analyses would have to acknowledge the degree of discrepancy in the normal population. When Schooler, Beebe, and Koepke (1978) compared the normal and LD populations, they found there were few Verbal-Performance IQ differences between the two groups.

Recent studies investigating significant differences between LD students, those with other handicaps, and the nonhandicapped population may be characterized by two general methods of analysis: the use of statistical cutoffs for different confidence levels and the comparison of actual differences between experimental and control groups. Veres (1982) indicated LD students were likely to exhibit differences between verbal and performance scores, although the direction of the difference was not significant. Further differentiation of age and IQ levels in comparing the LD group with the normal group (i.e., referred for evaluation but not classified) appeared to offer no consistently reliable guidelines or diagnostic rules.

Epps et al. (1982b) attempted to verify the consistency of Verbal-Performance IQ differences as a diagnostic tool through the discriminant analysis of 48 LD and 96 nonhandicapped students at three levels of statistical confidence. Although the sample included children randomly selected from the third through fifth grades, no operational rules or guidelines for classification decisions were presented.

Indications that the diagnostic utility of comparing verbal-performance ability differences may be less than reliable have been presented by a number of researchers. Ysseldyke et al. (1982) compared the results of 49 subtests, including verbal-performance analysis, with a population of 50 LD and 49 LA students at the fourth-grade level. No psychometric differences of practical utility between the groups were observed.

The degree to which Verbal-Performance IQ differences are important appears to be a judgment by school personnel making diagnostic decisions. Epps, McGue, and Ysseldyke (1982) used 18 judges trained in school psychology and special education to evaluate the test results of 99 fourth-grade students, 50 of whom had previously ben classified as LD and 49 of whom had scored below the 26th percentile on the Iowa Test of Basic Skills but had not been classified as LD or handicapped. Verbal-performance difference was one of 48 criteria used in deciding whether a student was classified as LD or not LD. Judges were evaluated on their degree of leniency; the most lenient judges classified all students as LD. Factors used in influencing judges' decisions were rank-ordered on the basis of average mean differences of the two groups. For the lenient judges, verbal-performance differences were ranked 10 out of a possible 48. For the nonlenient judges, verbal-performance differences were ranked below 10 (actual degree not reported). On the basis of this study, verbal-performance difference was found not to be as important in making classification decisions as were ability-achievement discrepancies or other test characteristics such as individual ability and achievement scores.

In summary, there appears to be little evidence that assessing the degree or direction of Verbal-Performance IQ differences helps differentiate the LD from those with other handicaps and the nonhandicapped but LA population. Of particular note is the lack of studies concerning the nonhandicapped but LA population who share a common characteristic with the LD (low achievement). Although the Institute for Research on Learning Disabilities at the University of Minnesota has completed several studies, sample size and age restrictions on these studies have been limited. The Institute has asserted a need for further studies with a focus on operationality.

Comparisons of Ability-Test Subtest Scatter

In Chapter I the concept was stated that subtest scores represent the degree of cognitive functions. Atypical patterns in subtest-score scatter were thought to demonstrate atypical learning patterns, which were thought to characterize a learning disability.

Kaufman (1976a) indicated that the variability of range scatter differed between LD and normal children; children with higher overall IQs exhibited more range than did those with lower IQs. In further research efforts, Tabachnick (1979), Gutkin (1979), and Thompson (1980) indicated similar differences between LD and normal children. However, they advised caution in using the results for interpretive purposes.

Clarizio and Bernard (1981) found that attention-factor scores differentiated the LD from the educable mentally impaired and nonclassified groups but not from the emotionally impaired group. This study supported the contention that, as a group, the LD score relatively lower on subtests categorized under attention-concentration. In reviewing the differences between profiles, however, Clarizio and Bernard indicated that these scores were not useful in discriminating between groups or in making accurate predictions for diagnoses.

Veres (1982) compared the scatter indices of a group of children classified as LD with a nonclassified population and compared recategorized subtests, similar to the study by Clarizio and Bernard (1981). Veres' analyses found no useful diagnostic rule, even when factors of age and IQ level were cross-tabulated.

Algozzine and Ysseldyke (1982a) investigated whether ability-subtest scatter is a useful diagnostic tool in differentiating nonhandicapped LA from LD students. When the results of 99 fourth-grade children (50 identified as LD and 49 labeled LA) were compared, no significant differences were found. Using a cutoff level of 10 points between the highest and lowest subtest scores, 21% of the LD and 77% of the LA samples were found to exhibit a significant degree of subtest scatter. Ysseldyke et al. (1982) reported little variance when their sample of LD students was compared with nonhandicapped LA students. Using mean-difference scores, which included analysis of ability-subtest scatter, the authors concluded that "we do not yet have good enough data regarding the extent to which identification as LD . . . is beneficial." They emphasized that further evaluation of similar populations (i.e., LA students) should be considered.

Epps, Algozzine, and Ysseldyke (1982) compared the performance on ability tests of randomly selected students, LD students, and nonhandicapped LA students. The study included 144 students in the third through fifth grades. Several operational definitions were explored, including the degree of subtest scatter. The results indicated that, in correlating degree of subtest scatter with other definitions including ability-achievement discrepancy and verbal-performance difference, the degree of scatter had a negative relationship with identification of LD students. That is, when the greatest number of LD students were identified by any other definition, the degree of subtest scatter was small.

In summary, researchers have used recategorized subtests and analyzed degree of scatter and mean deviation with LD students and those with other handicaps, as well as with nonhandicapped but LA students. There is some indication that plausible differentiation may result from using subtest-scatter analyses. However, a number of researchers have contrary opinions about this.

Direction of the Present Study

This study was directed toward comparing LD with LA students to evaluate methods proposed in the literature as being useful in differential diagnosis. The investigator adjusted for limitations found in previous studies that had left results open to interpretation. The major assumption was that:

> Comparisons of LD with LA students on standardized test results will not lead to decision rules that are useful in differential diagnosis when general ability and achievement characteristics, ability-achievement discrepancies, verbal-performance ability differences, and degree of ability-subtest scatter are used as the basis for comparisons.

Limitations in the Literature

Various writers have suggested that misidentification of LD and LA students (false positive) is prevalent in current practice (Hallahan & Kaufman, 1978; Ysseldyke & Algozzine, 1979). Although some studies have addressed this issue (Veres, 1982; Ysseldyke et al., 1982), there is a general lack of evidence concerning the usefulness of the differential-diagnosis methods examined in the present research. Studies that have evaluated these methods have produced results that are open to various interpretations. Small sample size, grade restrictions, inadequate controls, the influence of discriminant functions (moderator variables), and opposing findings have characterized the studies reported in the literature.

Sample size and composition. With the exception of Warner et al. (1980), who reported difficulty finding enough students below the 33rd percentile, studies have been limited primarily to fourth- and fifth-grade students (Algozzine & Ysseldyke, 1982b; Epps et al., 1982; Shinn et al., 1982). Algozzine and Ysseldyke, Epps et al., and Shinn et al. reported that some students included were referred, others randomly selected, and yet others included whose basis for classification decisions was not clear (Algozzine & Ysseldyke, 1982a; Epps et al., 1982; Norman & Zigmond, 1980; Shepard, 1983).

Inadequate controls. In previous studies, assessment results have been compared without controlling for the differing validities and reliabilities of tests. Also, correlations between student performance on achievement tests have not been explored (Epps et al., 1982; Shinn et al., 1982; Ysseldyke et al., 1982).

Discriminant functions. The influence of moderator variables such as sex, age, grade, referral reason, and IQ level on study results appears to be limited. Ysseldyke et al. (1982) used a stepwise linear discriminant function model to assess the influence of test scores on performance differences between LD and non-LD students. Researchers, however, have not reported the influence of moderator variables partially because the necessary demographic information was unavailable (Shinn et al., 1982; Ysseldyke et al., 1982).

Opposing findings. The use of ability-achievement discrepancy formulas has been explored in a number of studies, but there was little consensus about their statistical reliability for diagnostic purposes. Shepard (1980) indicated that the regression method is statistically reliable (if used with co-normed tests), whereas Reynolds (1981) advocated the *z*-score method for determining statistically significant differences. Support for either method has not been reported; some investigators (Algozzine & Ysseldyke, 1982a; Shepard, 1983) have examined the effectiveness of other methods.

Contributions of the Present Study

The results of this study may demonstrate the diagnostic utility of certain methods that are currently used for differential diagnosis. The percentage of false-positive and false-negative diagnoses that do not affirm the decisions of Educational Placement and Planning Committees (EPPCs) may be demonstrated.

The sample size of this study was substantial (N = 404) and comprised a broad spectrum across grades (first through twelfth) so that method effectiveness could be compared across grade ranges.

The same standardized tests (dependent on age) were used to assess all students included in this study. Therefore, no variability in performance may be attributed to different test measures. The LA group was selected on the basis of specific achievement-score levels. Selected diagnostic methods were applied to alternate comparisons in an attempt to indicate differing usefulness. The relationship between student performance and achievement subtests was examined.

The influence of discriminant functions (moderator variables) was analyzed for their effect on group differences when these differences were found to be statistically significant.

The students selected for this study were relatively typical of those referred for psychological evaluation in rural and suburban areas. All students had been referred for evaluation after attracting attention in the school environment and were classified by EPPCs as either LD or not eligible for special education services. Equal probability of classification in either category was reasonably assured by the nature of the EPPCs. These multidisciplinary teams of professionals followed specific procedures for identifying the handicapped, as mandated by the state of Michigan (effective since 1973). Consideration of ecological factors other than test scores was part of the classification procedures, and specific steps were taken to ensure parental involvement. These professional teams were also responsible for initiating educational plans for students classified as handicapped.

The results of this study should provide information on the usefulness of methods used in differential diagnosis that are absent in the current literature. The comprehensiveness of the study in terms of sample size and selection should lend credibility to the results, which can be generalized to the referred population when LD and LA classification decisions are the issue.

CHAPTER 3

METHOD

Source of Data

The data used in this study were collected during the 1975-76 school year by the Michigan Department of Education, Special Education Services. Bernard (1978) described many of the essential features of these data.

Of the 645 school districts in Michigan, 45 agreed to participate in the data-collection project. Each of the school districts provided data on all children considered by EPPCs regarding their eligibility for special education programs and services. The data submitted by school psychologists encompassed 57 factors, including reason for referral, WISC-R or other IQ scores, grade placement, chronological age, sex, WRAT or other achievement scores, and the decisions of the EPPCs.

Data were collected on a total of 1,552 students with a grade range from 0 through 12.9 (reported in tenths) and a mean grade placement of 4.8. The age range of the students was from 1.3 through 24.9 years, with a mean age of 10.5.

Of the 1,552 student evaluations submitted, 356 were reevaluated. The data derived from these reevaluations were not included in this study, to avoid bias in classification decisions and also because no information was available about previous classifications or length of time the student had been receiving special education services.

The EPPCs assigned the students to five categories: learning disabled (417), educable mentally impaired (228), emotionally impaired (137), physically or otherwise handicapped (138), and those students not found eligible for special education (422). Classification decisions were not available on 210 students.

The Sample

Data germane to this study were available for 230 LD students and 174 students determined not eligible for special education services who had been included in the data-collection project. Students were included in the present study if they were classified by EPPCs as LD or not eligible for special education services and if their scores from the Wechsler battery, word recognition subtest scores from the Wide Range Achievement Test, and grade levels were reported. Students not eligible for special education were included if they scored below the 26th percentile on the word recognition subtest. Of this group, 125 students scored below the 17th percentile.

The sample used in this study is not purported to be representative of the special education or general populations in Michigan at the time of data collection. School districts participated voluntarily in the data-collection project. Of the 645 school districts in Michigan, 45 agreed to participate. Suburban and rural districts reported data for 84.9% of the students included in the study. A sizable percentage of the students (38.8%) were evaluated by 11% of the participating districts. A marked percentage (86.6%) of students were male; females accounted for only 13.4% of the sample. This representation is disproportionate to the number of females in the general population, but is consistent with sex ratios in LD populations. It is unknown whether the reasons for referral are characteristic of the school-referred population. Students referred for academic reasons constituted 87.6% of the sample; those referred for behavioral and physical reasons accounted for 10.1% and .9% of the sample, respectively. The socioeconomic status of this sample was weighted, with 1% from high levels, 51.5% from middle levels, and 22% from low levels. Characteristics of the sample are reported in Table 1.

Procedures

The students included in the study were those who had been classified by EPPCs as LD or not eligible for special education services with no discernible handicap. It is unclear what factors the EPPCs used in making these decisions. However, it is known that all students had attracted attention in the school environment and were referred for evaluation because of academic, behavior, or physical reasons. The majority of students (87.6%) were referred for academic reasons.

The researcher assumed that the EPPCs followed guidelines for the identification of handicaps as mandated by the state of Michigan (effective since 1973) and, more important, that the classification decisions were accurate. Thus it was assumed that the effectiveness of methods used in the present study to affirm EPPC classification decisions could be evaluated.

Students who were classified by EPPCs as LD were included in the study, regardless of their level of ability or achievement scores, if

Table 1. Sample characteristics

	Group					
	LD		LA<26th		LA< 17th	
	%	N	%	N	%	N
Total number		(230)		(174)		(125)
Sex						
Male	81	(186)	72	(164)	76	(95)
Female	19	(43)	28	(49)	24	(30)
Socioeconomic Status						
High	2	(3)	0	(0)	0	(0)
Middle	69	(123)	69	(85)	65	(56)
Low	29	(52)	31	(38)	35	(30)
Type of District						
Urban	12	(26)	9	(15)	10	(12)
Suburban	27	(57)	18	(31)	18	(23)
Rural	61	(132)	73	(123)	72	(90)
Type of Referral						
Behavior	8	(17)	14	(24)	8	(10)
Academic	91	(206)	86	(148)	92	(114)
Physical	1	(3)	1	(1)	0	(0)
IQ Level						
0 through 74	11	(26)	9	(15)	11	(14)
75 through 89	41	(94)	50	(87)	56	(70)
90 through 110	42	(97)	37	(65)	30	(37)
111 through 130	3	(7)	1	(2)	1	(1)
131 through 150	3	(6)	3	(5)	2	(3)
Mean IQ (SD)	88.00	(12.38)	86.69	(10.01)	84.93	(9.58)
Age Ranges						
5- 8	32	(74)	35	(61)	31	(39)
9-11	28	(64)	36	(62)	32	(40)
12-14	25	(58)	18	(31)	25	(31)
15-17	13	(30)	10	(18)	10	(13)
18-22	1	(3)	0	(1)	1	(1)
Mean Age (SD)	10.67	(3.27)	10.22	(2.87)	10.50	(2.83)
Grade Ranges						
1- 3	39	(89)	47	(81)	44	(55)
4- 6	20	(47)	22	(39)	22	(27)
7- 9	26	(59)	19	(33)	22	(28)
10-12	15	(35)	12	(21)	12	(15)
Mean Grade (SD)	4.99	(3.01)	4.43	(2.78)	4.62	(2.76)

Note: Totals vary because of missing data. Where totals vary, percentages are of the available data.

data germane to this study were reported. The 230 students in the LD group had reported scores from the Wechsler Battery, word recognition subtest scores from the Wide Range Achievement Test (Jastak & Jastak, 1965), and grade-level indicators. Students who were classified as not eligible for special education services with no discernible handicap were included if there were scores from the Wechsler Battery, grade-level indicators, and if their scores from the word recognition subtest fell below the 26th percentile. These 174 students constituted the LA group. Of this number, 125 students scored below the 17th percentile on the test of word recognition.

In the data analysis conducted for the present study, comparisons were made between the 230 LD students and the 174 LA students who scored below the 26th percentile (LA<26th). Identical analyses were conducted between the LD group and the 125 LA students who scored below the 17th percentile (LA<17th).

Dependent Variables

General ability and achievement characteristics of both groups were compared by computing the degree of variance between group means. The *z*-score discrepancy method (Reynolds, 1981) was used for calculating ability-achievement discrepancies between full-scale IQ scores and word recognition subtest scores. As noted in Chapter 3, Reynolds reported that a *z*-score discrepancy model would be useful in ascertaining whether there was statistically significant variation between ability and achievement scores on standardized tests. Reynolds indicated that when test reliabilities from which scores are derived fall within the range of .90 to .95, an absolute difference of .66

standard deviations or more indicates that a significant difference exists. The reliabilities of the tests used in this study exceeded this range; therefore, the results represent a conservative estimate of statistically significant differences. Reynolds' method was applied in two ways: absolute differences between ability and achievement scores and directional differences (mathematical signs used in calculations).

The variance in achievement subtest scores (word recognition, spelling, and reading) was calculated by two methods: range (highest minus lowest (Algozzine & Ysseldyke, 1982a; Epps et al., 1982) and mean deviation (difference between the student's overall achievement mean and the most variant achievement score.

The degree of scatter (variance) on ability tests was calculated by three methods: range (highest minus lowest scaled score; Algozzine & Ysseldyke, 1982a; Epps et al., 1982), range (NDEV, variance by three or more scaled score points from the student's mean; Kaufman, 1976a; Veres, 1982), and mean deviation (difference between the student's IQ mean and scaled scores which exceeds that required at the .15 confidence level).

<u>The Utility of Diagnostic Decision Rules</u>

Although statistically significant differences may be obtained in some of the analyses included in this study, they do not necessarily offer direct evidence on the question of utility. Diagnostic rules are useful if they aid in differential diagnosis. That is, the rule can be applied to student characteristics and can result in appropriate classification decisions that do not exceed an acceptable error level.

Deno and Mirkin (1977) suggested that a student should be performing at half the rate of the average student to require intervention from special education services. This rule is relatively straightforward, with a criterion establishing the need for services. However, the method of implementation (how achievement is assessed) may contain errors that result in misdiagnosis.

The diagnostic criteria and methods reviewed in Chapter 2 and used as dependent variables in this study were applied to student test performance results. They can be considered useful in differential diagnosis if they affirm the EPPC classification decisions and differentiate at least 85% of the LD from the LA students. Although this level of usefulness is better than chance, it is an arbitrary criterion proposed in this study. It is, however, more generous than the criterion established by interpretation of the federal regulations for the identification of learning disabilities, which allows no error in classification. The percentage of students in each group who would have been misdiagnosed by the criteria and methods used are reported as false-positive and false-negative decisions and are compared with the 85% level to validate differential diagnostic usefulness.

Statistical Treatment

For purposes of this study, any statistical procedure that resulted in variance between groups at a .05 or greater level of confidence was judged to be significant. The Statistical Package for the Social Sciences (Hull & Nie, 1981) was used for analyses.

Correlation analyses were conducted on achievement subtest scores, dependent variables, and moderator variables. Pearson product-moment correlations and resultant probability levels were used to determine whether to conduct further analyses assessing the influence of these variables. When the variables were highly correlated ($p \leq .05$), it was assumed that a degree of interdependence existed, and analyses were conducted using the most inclusive variable.

The significance of variance between groups on general ability and achievement characteristics was derived from pooled t-tests because variance was assumed to be unequal unless otherwise found.

Univariate analyses of variance were calculated between groups on all dependent variables. These analyses were conducted on an overall design with all discriminant functions included. A full factoral model adjusting for unequal cell frequencies set initial levels of variance with which resultant discriminant function analyses could be compared. A model of univariate variance was used to calculate the degree of influence moderator variables had on the variance of dependent variables.

The percentage of false-positive and false-negative diagnoses was calculated to indicate the usefulness of dependent variables in terms of diagnostic utility.

Hypotheses

This study was designed to evaluate the effectiveness of criteria and methods used for differential diagnosis of LD as compared to LA students. Comparisons to EPPC classification decisions were essential to this evaluation. The hypotheses tested in this study are as follows:

Hypothesis 1: Comparisons of the general intelligence characteristics of LD and LA students will not lead to a useful decision rule in differential diagnosis.

Hypothesis 2: Comparisons of the general achievement characteristics of LD and LA students will not lead to a useful decision rule in differential diagnosis.

Hypothesis 3: Comparisons of ability-achievement discrepancy levels of LD and LA students will not lead to a useful decision rule in differential diagnosis.

Hypothesis 4: Comparisons of ability-test scatter indices of LD and LA students will not lead to a useful decision rule in differential diagnosis.

Hypothesis 5: Comparisons of verbal-performance ability test results of LD and LA students will not lead to a useful decision rule in differential diagnosis.

CHAPTER 4

RESULTS

Overview

The primary focus of this research was to analyze the performance of LD and LA students on standardized ability and achievement tests to determine if useful diagnostic decision-making rules could be validated. Of particular importance was the distinction between statistical and applied practical value. Although some analyses may provide statistical evidence affirming theoretical concepts, their ability to differentiate between handicapped (LD) and noneligible (LA) students should be of concern to those individuals making educational decisions.

Analyses were conducted on two groups of students: 230 LD students (grades 1 through 12) and 174 LA students scoring below the 26th percentile on a test of word recognition. Of the latter group, 125 students scored below the 17th percentile on the word recognition test. Comparisons were made between the LD students and both groups of LA students. Sample characteristics were discussed in Chapter 3 and were presented in Table 1.

In this chapter, each hypothesis is restated in the testable null form. The results of the statistical analyses are presented and discussed. A final section includes findings that are of general interest but are not necessarily related to the hypotheses investigated.

Comparisons of Ability and Achievement Test Performance

Ability Test Performance Comparisons

Hypothesis 1 was formulated to compare the ability test performance of the two groups of students (LD and LA).

> Hypothesis 1: Comparisons of the general intelligence characteristics of LD and LA students will not lead to a useful decision rule in differential diagnosis.

The means, standard deviations, and *t*-test *F*-ratios and probability levels comparing the means are given in Table 2 for each group and each comparison. Casual inspection of the means and standard deviations indicates a degree of difference that would be of little value to diagnosticians. The differences between means ranged from 0.24 points when the LD and LA<26th groups were compared on performance abilities to 3.07 points when the LD and LA<17th groups were compared on verbal abilities. Given the format of the WISC-R and the WAIS and the degree of measurement error (3.19), such differences would be of little diagnostic value.

When *t*-tests were used to calculate the significance of group mean differences, two comparisons indicated statistical significance. The difference between the LD and LA<17th groups for Full Scale means was 2.82 points, which was significant at the .01 level. The Verbal IQ difference between the LD and LA<17th group means was 3.07 points, which was significant at the .01 level. Are these differences in test results useful for purposes of diagnostic utility? Table 3 indicates the degree of misidentification that would occur if ability test means

Table 2. T-test F-ratios and probability levels for significant differences between subsamples' ability and achievement scores

	LD and LA<26th						LD and LA<17th					
	LD		LA<26th				LD		LA<17th			
	Mean	SD	Mean	SD	F	Prob.	Mean	SD	Mean	SD	F	Prob.
WISC-R IQ (FS)	89.55	11.19	88.49	10.06	.97	.33	89.55	11.19	86.73	9.45	2.36	.01
WISC-R IQ (V)	88.00	12.38	86.69	10.01	1.04	.30	88.00	12.38	84.93	9.58	2.28	.02
WISC-R IQ (P)	92.82	12.59	92.58	11.70	-0.24	.81	92.82	12.59	91.21	11.38	.71	.48
WRAT Word Recognition	2.49	1.82	2.71	1.72	-1.31	.19	2.49	1.82	2.57	1.58	-0.48	.63
WRAT Spelling	2.37	1.52	2.53	1.66	-1.15	.25	2.37	1.52	2.38	1.38	-0.25	.81
WRAP Arithmetic	2.86	1.57	2.86	1.61	-0.10	.92	2.86	1.57	2.87	1.54	-0.22	.83

Table 3. Summary of use of ability and achievement characteristics to classify students

Measure	Correct Classification		Incorrect Classification		Percent False Positive	Percent False Negative	Percent Incorrect Overall
	LA	LD	LA	LD			
Use of Ability Characteristics							
LD and LA<26th							
Full Scale IQ	102	120	72	110	41	48	45
Verbal IQ	90	119	84	111	48	48	48
Performance IQ	83	128	91	102	52	44	48
LD and LA<17th							
Full Scale IQ	82	117	43	113	34	49	44
Verbal IQ	78	132	47	98	38	43	41
Performance IQ	70	131	55	99	44	43	43
Use of Achievement Characteristics							
LD and LA<26th							
Reading Recognition	74	149	100	81	57	35	45
Spelling	81	127	93	103	53	45	49
Arithmetic	100	140	74	90	43	39	41
LD and LA<17th							
Reading Recognition	50	149	75	81	60	35	44
Spelling	52	127	73	103	58	45	50
Arithmetic	72	90	53	140	42	61	54

were used as diagnostic criteria to classify students in the three groups. Forty-three of the LA<17th group and 113 of the LD group obtained scores on Full Scale IQ measures that were closer to the other group's mean than to their own group's. The degree of overall incorrect classification that would result if group means were used as a basis for diagnosis was 44%. Similar results were obtained for Verbal IQ mean comparisons, with an overall error rate of 41%. Applying this method to determine diagnostic utility where differences between group means were not statistically significant indicated an even higher degree of misidentification; the highest overall incorrect percentage was between the LD and LA<26th groups on measures of Verbal and Performance IQ. In all analyses, the degree of misidentification that would occur if observed differences between groups on the ability test were used for classification exceeded the level set in this study as acceptable for diagnostic utility (15%). Therefore, Hypothesis 1 was not rejected.

<u>Achievement Test Performance Comparisons</u>

Hypothesis 2 was formulated to compare the achievement test performance of the two groups of students (LD and LA).

> Hypothesis 2: Comparisons of the general achievement characteristics of LD and LA students will not lead to a useful decision rule in differential diagnosis.

The means, standard deviations, and t-test F-ratios and probability levels comparing the means are given in Table 2 for each group and each comparison. Inspection of these indices indicated that differences between the groups were too small for diagnostic purposes. Actual differences between achievement test means of the LD and LA<26th groups ranged from 0.0 to .22 for word recognition results. With the exception of arithmetic, the LA groups obtained means that were higher than those of the LD group. It should be noted that the higher mean (2.87) obtained by the LA<17th group, as compared to the mean obtained by the LA<26th group (2.86) on arithmetic was not due to statistical error. The LA group had been subdivided on the basis of their scores on reading recognition.

Analyses of variance computing group differences in achievement score range did not yield statistically significant results. The results of these analyses for both methods are presented in Tables A1 and A2 in the Appendix.

Do the differences in achievement test results offer any diagnostic value for classification purposes? Table 3 indicates the degree of misidentification that would occur if achievement test group means were used to classify students. The degree of overall incorrectness that would result if proximity of the student's score to the group mean score was used for classification exceeded the level for diagnostic utility (15% incorrect). The lowest overall incorrect percentage was 41%, which was obtained for LD and LA<26th group comparison on arithmetic results. In this comparison, 74 students in the LA<26th group and 90 LD students had scores that were closer to the other group's mean. Based on the degree of misclassification that would occur if the degree of difference between group means on achievement test results was used to classify students, Hypothesis 2 was not rejected.

Ability Achievement Discrepancy Model Comparisons

Hypothesis 3 was formulated to compare LD and LA students in terms of discrepancies between ability and achievement test results.

> Hypothesis 3: Comparisons of ability-achievement discrepancy levels of LD and LA students will not lead to a useful decision rule in differential diagnosis.

Absolute *z*-Score Discrepancy-Level Differences

Using Reynolds' (1981) formula, the numbers and percentages of each group were calculated for various *z*-score discrepancy levels. The results are presented in Table 4. The results indicated that no group met the criterion for diagnostic utility (85% affirmation rate), as defined in this study.

Of the LD group, 35% had discrepancy levels that fell within the nonstatistically significant range (0.0 through 0.65 standard deviations), as reported by Reynolds. When discrepancy levels were reviewed by grade ranges, the figures remained fairly consistent with the overall percentages for grades 1 through 12. The highest percentage of students in the LD group who were identified as being correctly classified by this method were in grades 10 through 12; 69% of these students exhibited statistically significant discrepancies. Discrepancy-level calculations for the LA groups indicated that approximately 50% of both groups, overall and in specific grade ranges, had discrepancies within the statistically significant range. If the discrepancy method is useful for differentiating between the LD and the LA student, it was not verified by the results reported in this study.

Table 4. Percentages and numbers of students scoring at *z*-score discrepancy levels

Discrepancy Levels (Standard Deviations)	Group					
	LD		LA<26th		LA<17th	
	%	*N*	%	*N*	%	*N*
Grades 1 Through 12						
2.00 through 4.00	10	(23)	6	(11)	6	(7)
1.50 through 1.99	11	(25)	4	(7)	5	(6)
1.00 through 1.49	27	(61)	11	(20)	12	(15)
0.66 through 0.99	17	(40)	24	(42)	24	(30)
0.00 through 0.65	35	(81)	54	(94)	54	(67)
Grades 1 Through 3						
2.00 through 4.00	8	(7)	7	(6)	5	(3)
1.50 through 1.99	10	(9)	5	(4)	5	(3)
1.00 through 1.49	24	(21)	11	(9)	13	(7)
0.66 through 0.99	19	(17)	20	(16)	18	(10)
0.00 through 0.65	39	(35)	57	(46)	58	(32)
Grades 4 Through 6						
2.00 through 4.00	6	(3)	3	(1)	0	(0)
1.50 through 1.99	11	(5)	0	(0)	0	(0)
1.00 through 1.49	32	(15)	10	(4)	11	(3)
0.66 through 0.99	17	(8)	31	(12)	30	(8)
0.00 through 0.65	34	(16)	56	(22)	59	(16)
Grades 7 Through 9						
2.00 through 4.00	12	(7)	9	(3)	11	(3)
1.50 through 1.99	14	(8)	9	(3)	11	(3)
1.00 through 1.49	27	(16)	6	(2)	7	(2)
0.66 through 0.99	15	(9)	27	(9)	29	(8)
0.00 through 0.65	32	(19)	48	(16)	43	(12)
Grades 10 Through 12						
2.00 through 4.00	17	(6)	5	(1)	7	(1)
1.50 through 1.99	9	(3)	0	(0)	0	(0)
1.00 through 1.49	26	(9)	24	(5)	20	(3)
0.66 through 0.99	17	(6)	24	(5)	27	(4)
0.00 through 0.65	31	(11)	48	(10)	47	(7)

Note: Percentages are approximate and are reported to the nearest whole number.

Directional *z*-Score Discrepancy-Level Differences

Reynolds (1981) was concerned with the absolute difference between ability and achievement test results when calculating the degree of discrepancy (for statistical purposes). On the other hand, the directionality of the *z*-score when mathematical signs were not ignored was surprising. When the difference between the IQ score and the IQ test mean was greater than the difference between the achievement test score and the achievement test mean, a negative *z*-score was obtained. When such calculations resulted in a negative *z*-score, the student was achieving better than what would be expected on the basis of the ability (IQ) assessment. The percentages and numbers of students in each group who scored at negative *z*-score discrepancy levels is reported in Table 5. Of the LD group, 12% achieved better than expected at a statistically significant level, and 25% had actual *z*-scores that indicated this directionality. Given the degree of correlation between achievement subtest scores (reported under general findings), it seems doubtful that low achievement was a major factor in classifying at least 12% of the LD group (25% if statistical significance was ignored). This finding may explain, in part, the classification of the LA students as not eligible for services. If better-than-expected achievement was part of the diagnostic process and if no other contributing factors were identified that influenced classification as LD, at least 8% of the LA<26th group and 6% of the LA<17th group appeared to be doing better than expected on school-related tasks.

As indicated in Chapter 2, a number of formulas have been developed to quantify ability-achievement discrepancies. The current

Table 5. Percentages and numbers of students scoring at *z*-score discrepancy levels with negative directionality

Discrepancy Levels (Standard Deviations)	Group					
	LD		LA<26th		LA<17th	
	%	*N*	%	*N*	%	*N*
Grades 1 Through 12						
-0.66 through -4.0	12	(28)	8	(14)	6	(7)
-0.01 through -4.0	25	(58)	29	(50)	27	(34)
Grades 1 Through 3						
-0.66 through -4.0	19	(17)	10	(8)	5	(3)
-0.01 through -4.0	34	(30)	30	(24)	29	(16)
Grades 4 Through 6						
-0.66 through -4.0	11	(5)	5	(2)	0	(0)
-0.01 through -4.0	19	(9)	21	(8)	15	(4)
Grades 7 Through 9						
-0.66 through -4.0	8	(5)	6	(2)	7	(2)
-0.01 through -4.0	20	(12)	39	(13)	36	(10)
Grades 10 Through 12						
-0.66 through -4.0	3	(1)	10	(2)	13	(2)
-0.01 through -4.0	20	(7)	24	(5)	27	(4)

Note: Percentages are approximate and are reported to the nearest whole number federal definition specifically states that a condition of learning disabilities is: "The child does not achieve commensurate with his or her age and ability levels in one or more of the areas listed (USOE, 1977, p. 65082). The results presented in this study indicate that a sizable percentage of the LD group did not meet that criterion.

Discrepancy-Level Variance Between Groups

To assess the degree of variance between groups, univariate analyses of variance were calculated between groups in five grade ranges, with *z*-score discrepancy levels as the dependent variable. The results of these analyses appear in Table A1 and A2 in the Appendix. Table 6 indicates the variance of discrepancy levels between grade ranges when both the LD and LA<26th groups were combined. This analysis was conducted to assess if particular grade ranges could be identified in which the degree of discrepancy was most prominent. Homogeneous grade ranges (not differing by more than the degree required for statistical significance) appeared to be grades 4-6, 7-9, and 10-12. Grades 1-3 appeared to be the most variant; the mean for these grades was 6.03, which varied from the next closest grade range (grades 7-9) by .48 points.

To assess for sources of variation within grade ranges, analyses of variance were conducted for the LD and LA<26th groups at grade ranges in separate analyses. The degree of variation between grade ranges for the LA<26th group was not significant (p = .91), as indicated in Table 7. All grade ranges were homogeneous; none differed by the range required for statistical significance at the .05 level.

Table 6. Analysis of variance--discrepancy levels at grade ranges: LD and LA<26th groups

Source of Variation	df	Mean Square	F-Ratio	p
Between grades	3	14.97	4.12	.01
Within grades	400	3.63		

Means

Grades 1-3 = 6.03	Grades 7-9 = 6.51
Grades 4-6 = 6.67	Grades 10-12 = 6.89

Grand Mean = 6.40

Table 7. Analysis of variance--discrepancy levels at grade ranges: LA<26th group

Source of Variation	df	Mean Square	F-Ratio	p
Between grades	3	.53	.19	.91
Within grades	170	2.85		

Means

Grades 1-3 = 6.06	Grades 7-9 = 6.27
Grades 4-6 = 6.26	Grades 10-12 = 6.19

Grand Mean = 6.16

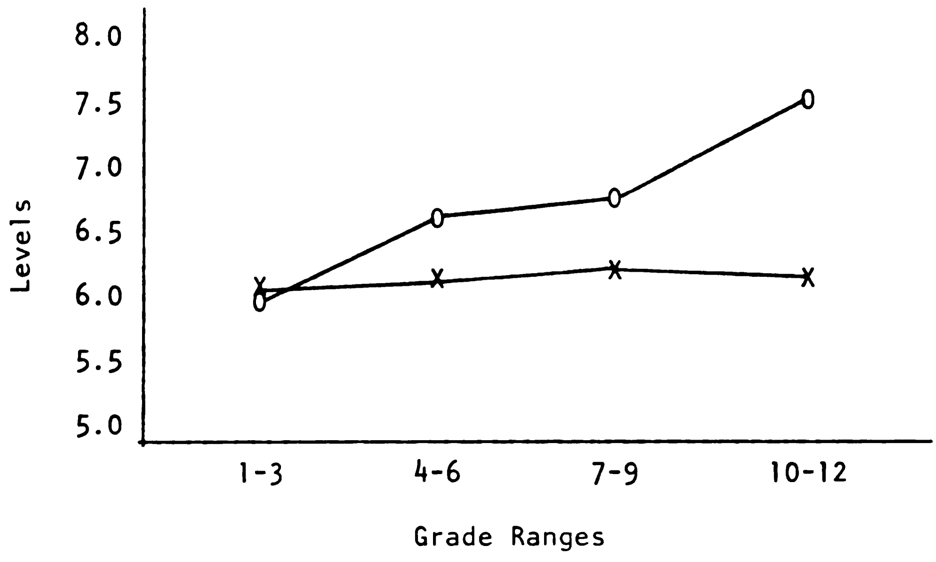
Significant variation between grade ranges was found to exist in the LD group. Table 8 indicates the source of variation as grades 10-12 being significantly different from grades 1-3. Figure 1 shows this variation, as well as the variation between groups at all grade ranges other than grades 1-3.

Table 8. Analysis of variance--discrepancy levels at grade ranges: LD group

Source of Variation	*df*	Mean Square	*F*-Ratio	*p*
Between grades	3	18.58	4.47	.00
Within grades	226	4.16		

Means

Grades 1-3 = 6.00
Grades 4-6 = 6.72
Grades 7-9 = 6.89
Grades 10-12 = 7.31
Grand Mean = 6.58



0 = LD
X = LA<26th

Figure 1. Discrepancy level means at grade ranges for LD and LA<26th groups.

Univariate analyses of variance with discrepancy level as the dependent variable were conducted between the LD and LA<17th groups. As reported in Table A2, no statistically significant variance was found to exist between groups (p = .20). When the groups were combined, there was significant grade-range variability (p = .03). Table 9 indicates the combined-group means; Figure 2 illustrates the variance.

Table 9. Analysis of variance--discrepancy levels at grade ranges: LD and LA<17th groups

Source of Variation	df	Mean Square	F-Ratio	p
Between grades	3	11.32	3.10	.03
Within grades	354	3.65		

Means

Grades 1-3 = 6.12
Grades 4-6 = 6.60
Grades 7-9 = 6.74
Grades 10-12 = 6.92
Grand Mean = 6.48

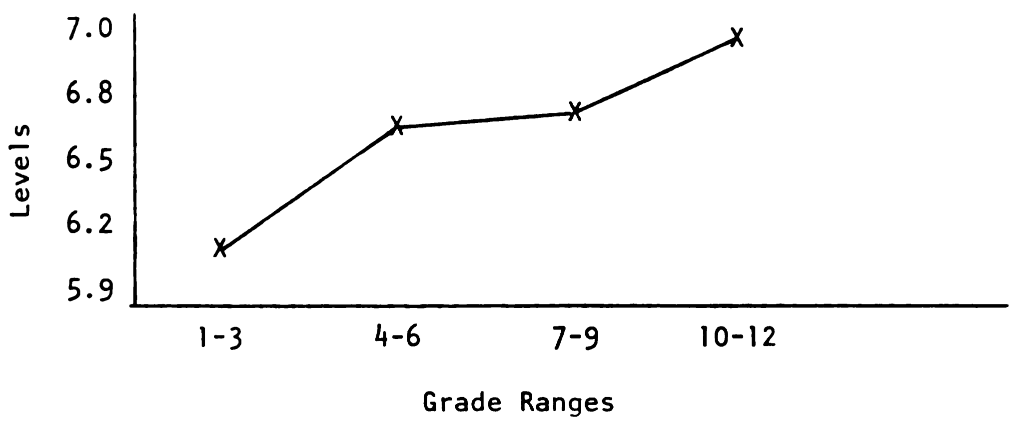


Figure 2. Discrepancy level means at grade ranges for LD and LA<17th groups.

In summary, the LD group was found to exhibit statistically significant variance in discrepancy levels, whereas no statistically significant variance was noted in the LA<26th or the LA<17th groups. There did appear to be statistically significant variance between the LD and LA<26th groups for all grade ranges except grades 1-3. Variance between the LD and LA<17th groups was found not to exist; however, when grade ranges of both groups were combined, statistically significant differences did exist between grade ranges.

z-Score Discrepancy-Level Diagnostic Utility

Although the univariate analyses between groups using discrepancy levels as a dependent variable indicated statistical significance, they shed no light on the question of diagnostic utility. Using Reynolds' (1981) formula for absolute differences between ability-achievement performance, discrepancy-level scores were used to compare groups. If a LD student exhibited a discrepancy level equal to or exceeding .66 standard deviation, he/she was judged to be correctly classified (even though a reliable difference between ability and achievement does not necessarily indicate a need for special education services). Conversely, if a LA student exhibited a .66 standard deviation or greater discrepancy, he/she was judged to be incorrectly classified. The results of these analyses are presented in Table 10.

From inspecting the data presented in Table 10, it is apparent that the *z*-score discrepancy-level method does not satisfy the criterion for diagnostic utility, as set in this study. Although 65% of the LD group were correctly classified using this method, 46% of the LA groups were incorrectly classified. For the practitioner who wishes to be correct 60% and 59% of the time, such a method would be useful. Of the LA<26th group, 80students exhibited discrepancy-level scores that met the criterion for statistical significance and would subsequently be diagnosed as LD (false positive) if this method was used for classification. Of the LA<17th group, 58 students met the same criterion, whereas 81 of the LD group would be judged as non-LD (false negative). The overall incorrect percentages (40% and 39%) reflect the error when both false-positive and false-negative indicators are combined. Based on the results of these analyses, Hypothesis 3 is not rejected.

Table 10. Summary of the use of *z*-score discrepancy levels to classify students

	LD	LA<26th
Correct classification	149	94
Incorrect classification	81	80
Percent false positive	..	46
Percent false negative	35	..
Percent incorrect overall--40		

	LD	LA<17th
Correct classification	149	67
Incorrect classification	81	58
Percent false positive	..	46
Percent false negative	35	..
Percent incorrect overall--39		

Ability-Test Subtest Scatter Comparisons

Another proposition presented earlier concerned the amount of scatter in ability subtest scores of the LD as compared with the scores of the LA student. Hypothesis 4 addressed these comparisons.

> Hypothesis 4: Comparisons of ability-test scatter indices of LD and LA students will not lead to a useful decision rule in differential diagnosis.

Ability-test subtest scatter was used to compare groups and was calculated in three ways: range (highest minus lowest scaled score), Kaufman's NDEV measure (variance from the student's mean by three or more points by scaled scores), and mean deviation (variance from the student's mean by the most variant scaled score). This within-person variance was calculated in numerical quotients from which mean comparisons were made. Only those scaled-score deviations from the student's mean that exceeded the difference required at the .15 level of confidence were used in calculating mean deviated scatter.

Table 11 contains the group means for the three measures of calculated scatter. Casual inspection of the data in this table indicates that no differences are apparent that would be of diagnostic utility.

Table 11. Group means for three measures of IQ scatter

Group	Range	NDEV	Mean Deviation
LD	10.83	1.23	1.67
LA<26th	9.97	.05	2.52
LA<17th	9.97	.64	2.34

Subtest Scatter Variance Between Groups

Analyses of variance using each method of calculating scatter as a dependent variable were performed to ascertain whether significant group differences existed. The results of these analyses are reported in Tables A1 and A2. Scatter calculated using the NDEV method was found to be significantly variant between the LD and LA<26th groups (p = .02). Table 12 reflects the degree of variance between group means when grade-range comparisons were not made.

Table 12. Analysis of variance--IQ scatter (NDEV): LD and LA<26th groups

Source of Variation	df	Mean Square	F-Ratio	p
Between groups	1	132.27	5.213	.02
Within groups	402	26.33		

Means

LD = -1.23 LA<26th = -0.05

Grand Mean = -0.713

To assess for causal factors, analyses of variance were performed for each group at grade ranges. Tables 13 and 14 illustrate the degree of variance within groups at different grade ranges. Figure 3 illustrates the degree of variance between groups at the four grade ranges.

The degree of variance between grade ranges within the LA<26th group was found not to be significant. All grade-range means were homogeneous; variance did not exceed the range required at the .05

level of confidence. Similar variance was found to exist within the LD group. When the groups were compared, however, statistically significant grade-range differences were found to exist, particularly at grade range 1-3, with a variance of 2.38 points between the LD and the LA<26th groups.

Table 13. Analysis of variance--IQ scatter (NDEV) at grade ranges: LD group

Source of Variation	df	Mean Square	F-Ratio	p
Between groups	3	71.47	2.05	.11
Within grades	226	34.88		

Means

Grades 1-3 = -2.27 Grades 7-9 = -0.29

Grades 4-6 = -1.42 Grades 10-12 = -0.13

Grand Mean = 1.23

Table 14. Analysis of variance--IQ scatter (NDEV) at grade ranges: LA<26th group

Source of Variation	df	Mean Square	F-Ratio	p
Between grades	3	4.01	.28	.84
Within grades	170	14.55		

Means

Grades 1-3 = -0.11 Grades 7-9 = 0.02

Grades 4-6 = -0.31 Grades 10-12 = 0.60

Grand Mean = -0.05

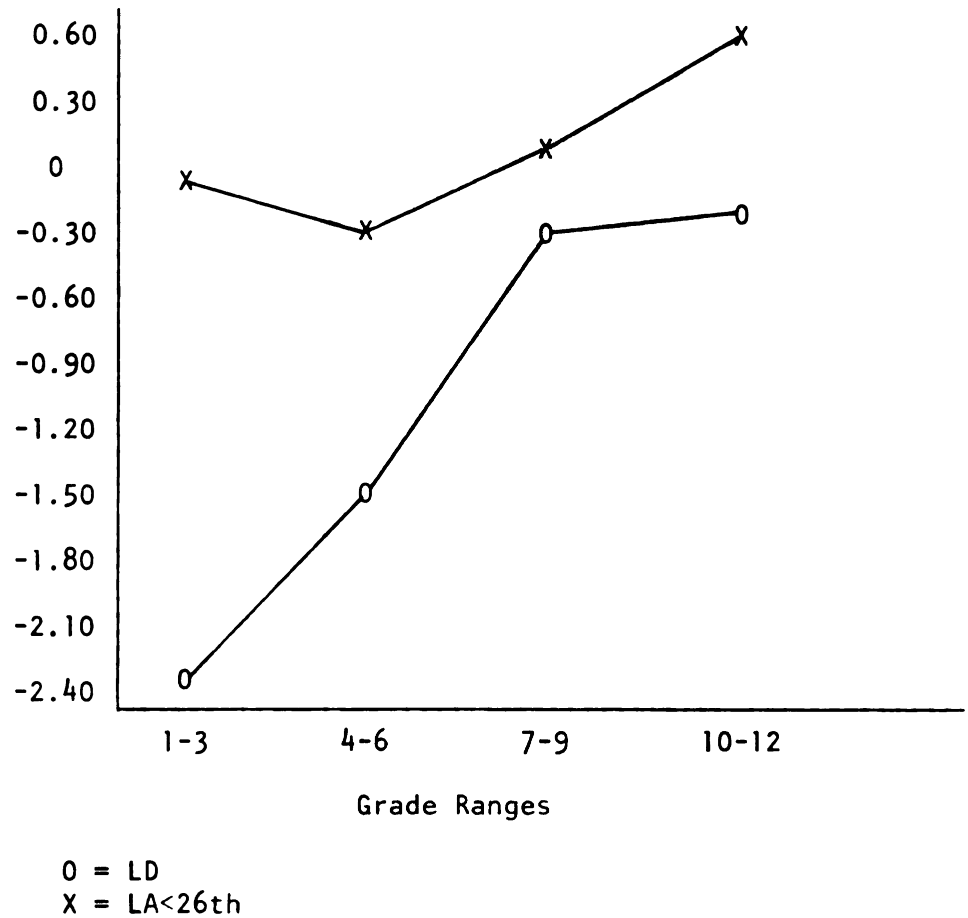


Figure 3. IQ scatter (NDEV) means at grade ranges: LD and LA<26th groups.

Subtest Scatter Diagnostic Utility

Does this statistically significant variance between groups have value for decision-making rules? Table 15 presents the results of classification decisions based on the degree of scatter for the three methods. Using range and mean-deviation methods, the proximity of the student's score to his/her group mean was used as the basis for classification. When the score was closer to the other group's mean than to the student's own mean, the method used misclassified the

Table 15. Summary of use of scatter methods to classify students

Method	Correct Classification		Incorrect Classification		Percent False Positive	Percent False Negative	Percent Incorrect Overall
	LA	LD	LA	LD			
LD and LA<26th Groups							
Range	54	161	120	69	70	30	47
NDEV	116	52	58	178	33	77	58
Mean Deviation	145	45	29	185	17	80	53
LD and LA<17th Groups							
Range	40	161	85	69	68	30	43
NDEV	69	52	56	178	45	77	66
Mean Deviation	102	45	23	185	18	80	59

student. Using the NDEV method, students were classified according to the degree of scatter. A student was correctly classified as LD if his/her scaled scores varied three or more points from the mean of his/her overall ability test performance. A LA student was correctly classified if his/her scaled scores varied less than three points from the mean of his/her overall ability test performance.

Inspection of the data presented in Table 15 indicates that the LD group would be misclassified to a greater extent than the LA groups, when the NDEV and mean-deviation methods are used. The range method (highest minus lowest scaled score) is the most promising for correct classification of the LD group, although it still does not meet the criterion set in this study for diagnostic utility. No one method was able to differentiate between groups with a degree of misclassification of 15% or less. When percentages of misclassification were combined, the error rate ranged from 43% to 66%. Based on the analyses conducted and the degree of misclassification that would occur using methods for calculating IQ scatter, Hypothesis 4 is rejected.

Comparisons of Verbal-Performance Ability Differences

As discussed in Chapter 3, researchers have postulated a difference between verbal-performance ability test scores that may distinguish the LD student from others who do not possess any discernible handicap. As Kaufman (1976b) noted, a difference of 12 points between these scores is required for statistical significance at the .05 level of confidence. Hypothesis 5 states that observed differences between the groups included in this study will not assist in the formulation of useful decision rules.

Hypothesis 5: Comparisons of verbal-performance ability test results of LD and LA students will not lead to a useful decision rule in differential diagnosis.

To evaluate Hypothesis 5, students included in the study were compared using two methods of analyzing verbal-performance IQ score differences: Kaufman's difference of 12 or more points and a mean deviated method in which each scaled score was included in the calculations if it differed from the student's overall ability test mean by the difference required at the .15 confidence level (reported in the test manual).

Actual Group Differences

The percentages and numbers of students for each group as they scored on the two methods are reported in Table 16. Inspection of the data indicates that no marked percentages of students in any group scored at one level on either method or the directional aspects of one method. Forty-eight percent of the LD students in grades 1 through 12 obtained scores that equaled their performance ability scores on the mean deviation method. Similar results (35% and 33%) were indicated for the LA groups. The method using 12 or more point differences failed to differentiate any group at the level set in this study for useful decision rules (85%). Fifty-eight percent of the LD group had verbal-performance differences of less than 12 points, as did 70% of the LA<26th group and 67% of the LA<17th group.

Variance of Verbal-Performance Ability Differences Between Groups

Analyses of variance were conducted using two methods as dependent variables. The results of the analyses are reported in Tables A1 and

Table 16. Percentages and numbers of students scoring at mean deviated and actual differences between Verbal-Performance IQs

	Mean Deviated			Actual			
	V<P	V=P	V>P	V<P	V=P	V>P	V≠P<12
	% N	% N	% N	% N	% N	% N	% N
LD							
All grades	19 (44)	48 (110)	33 (76)	22 (51)	5 (12)	15 (34)	58 (133)
Grades 1-3	19 (17)	45 (48)	27 (24)	18 (16)	7 (6)	20 (18)	55 (49)
Grades 4-6	29 (12)	47 (22)	28 (13)	17 (8)	6 (3)	19 (9)	57 (27)
Grades 7-9	20 (12)	39 (23)	41 (24)	24 (14)	3 (2)	7 (4)	66 (39)
Grades 10-12	9 (3)	49 (17)	43 (15)	37 (13)	3 (1)	9 (3)	51 (18)
LA<26th							
All grades	22 (38)	35 (61)	43 (75)	6 (10)	4 (7)	20 (35)	70 (122)
Grades 1-3	22 (18)	41 (33)	37 (30)	6 (5)	6 (5)	11 (9)	77 (62)
Grades 4-6	23 (9)	28 (11)	49 (19)	5 (2)	3 (1)	31 (12)	62 (24)
Grades 7-9	21 (7)	36 (12)	42 (14)	6 (2)	3 (1)	27 (9)	64 (21)
Grades 10-12	19 (4)	24 (5)	57 (12)	5 (1)	0 (0)	24 (5)	71 (115)
LA<17th							
All grades	22 (28)	33 (41)	45 (56)	7 (9)	2 (2)	24 (30)	67 (84)
Grades 1-3	24 (13)	38 (21)	38 (21)	4 (5)	2 (1)	7 (9)	73 (40)
Grades 4-6	19 (5)	30 (8)	52 (14)	4 (1)	0 (0)	33 (9)	63 (17)
Grades 7-9	25 (7)	29 (8)	46 (13)	7 (2)	4 (1)	29 (8)	61 (17)
Grades 10-12	20 (3)	27 (4)	53 (8)	7 (1)	0 (0)	27 (4)	67 (10)

Note: V=P column in Mean Deviated category indicates any difference less than what is required for significance at the .15 level.

Percentages are approximate and are reported to the nearest whole number.

A2. No variance approaching the .05 level of confidence was found to exist for either method.

Diagnostic Utility of Verbal-Performance Differences

To assess the question of diagnostic utility, comparisons were made between the classifications by EPPCs and the effectiveness of these methods to confirm such classifications. Table 17 presents the results of the comparisons. Using the actual-difference method (12 or more points), 63% of the LD group were incorrectly classified. Five percent of the LD group had equal verbal and performance scores, and 58% had verbal and performance scores that were not equal but were less than 12. These false-negative classifications far exceeded the level for useful decision-making rules (15%). In comparison, 26% of the LA<26th group and 31% of the LA<17th group were classified as false positive (having differences of 12 or more points). The mean-deviation method offered more congruence of classification decisions for the LD group, yet it still markedly exceeded the level for usefulness (15%). Of the LD group, 48% were incorrectly classified (false negative) using this method. The overall percentage error rates for both methods ranged from 47% to 55%, which offers little support for the diagnostic utility of these methods for useful decision-making rules. Based on the analyses conducted in this study, Hypothesis 5 cannot be rejected.

General Findings

The findings presented in this section relate to the hypotheses and methods presented in Chapter 3. Some of the findings might not focus directly on the usefulness of decision-making rules, but they are

Table 17. Summary of use of verbal-performance differences to classify students

Method	Correct Classification		Incorrect Classification		Percent False Positive	Percent False Negative	Percent Incorrect Overall
	LA	LD	LA	LD			
LD and LA<26th Groups							
Actual	129	85	45	145	26	63	47
Mean deviation	61	120	113	110	65	48	55
LD and LA<17th Groups							
Actual	86	85	39	145	31	63	52
Mean deviation	41	120	84	110	67	48	55

presented here to clarify certain aspects of the groups studied and the methodology used. The tables referred to in this section are found in the Appendix.

Achievement Test Performance Correlations

Pearson product-moment correlations were computed between the achievement test results of all groups included in this study. Between-group correlations were also calculated. The results of the correlational analyses indicated that variability in the two groups' performance on the three subtests (word recognition, spelling, arithmetic) was minimal. The performance of the LD group and that of the LA groups on these measures were highly correlated. The coefficients ranged from .552 to .923, all of which were significant at the .001 level of confidence. The correlations of achievement performance between the groups indicated similar results. The coefficients ranged from .638 to .849, and all were significant at the .001 level of confidence. Hypothesis 2 stated that comparison of general-achievement characteristics would not lead to useful decision-making rules for differentiating between groups. The results presented here would appear to verify this assertion. Performance on the achievement test included in this study indicated no statistically significant variance between the two groups of students.

Dependent Variable Correlation

With the exception of discrepancy-level calculations, the dependent variables used in this study were assessed in two or more ways. Two methods were used to evaluate Verbal-Performance IQ differences, three methods were used with ability-subtest scatter, and two methods were used to evaluate achievement variability. Pearson product-moment correlations were computed between the methods, to assess the statistical relationship. The methods used within each analysis were highly correlated with each other. The coefficients ranged from .19 to .93, all of which were significant at the .00 level of confidence.

Covariate Correlations

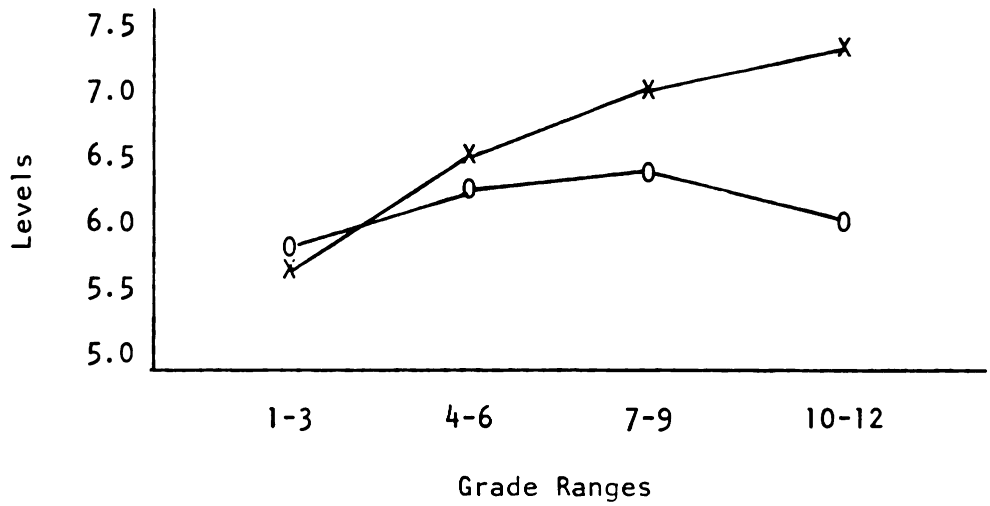
The degree of correlation (Pearson product moment) between covariates served as a basis for exclusion in the analyses of discriminant functions reported in this section. In Chapter 3, it was indicated that the influence of discriminant functions (moderator variables) would be investigated when overall analyses of variance were statistically significant. The selection of covariates included in these analyses was based on the relationship between covariates, as determined by correlation significance. Table A5 presents the results of the correlations between covariates. Covariates of IQ level, sex, and primary referral reason were found to be correlated at a statistically significant level with all other covariates but not with each other. Therefore, these covariates were found to be independent of each other in terms of the degree of influence on variance between groups.

Covariate Analyses

A univariate analysis with covariates uses a linear regression format that adjusts for the variance between groups caused by the discriminant function (moderating variable). The resultant *F*-ratios and significance levels are derived after these adjustments are made. Comparing the resultant statistical significance with the overall analysis (covariate included) gives an indication of the effect of the covariate on the degree of variance observed in the overall analysis. Covariates of IQ level, sex, primary referral reason, type of district, socioeconomic status, and age were defined as covariates in this study. Grade range was included in all overall analyses. IQ level, sex, and primary referral reason were found to be significantly related to all other covariates, but they were independent of each other. The influence of these covariates on statistically significant between-group variance was evaluated.

Discrepancy-level variance. To ascertain the degree of influence covariates had on the variation between groups, separate univariate analyses with covariates were conducted. The results of these analyses are reported in Tables A6 and A7. Primary referral reason and sex were not found to be significant in producing the observed variation between groups in the overall analysis. The significance levels of *t*-values indicating covariate effect did not approach the .05 level of confidence, and there was no significant change in the variation when compared with the overall analysis. IQ level appeared to have a significant effect in producing the variation in the overall analysis. A *t*-value of 19.25 with significance at 0 for the LD and LA<26th group comparison and a *t*-value of 18.10 with significance at 0 for the LD and LA<17th group comparison indicated the IQ-level effect. The adjusted means with variance not influenced by IQ level are depicted in Figures

4 and 5. Comparison of these means with the nonadjusted means shown in Figure 1 indicates a high degree of similarity in both mean differences and overall direction. The effect of IQ level, while statistically significant, did not appear markedly to influence group differences. IQ level appeared to have the same effect for the LD and LA<17th group comparisons.

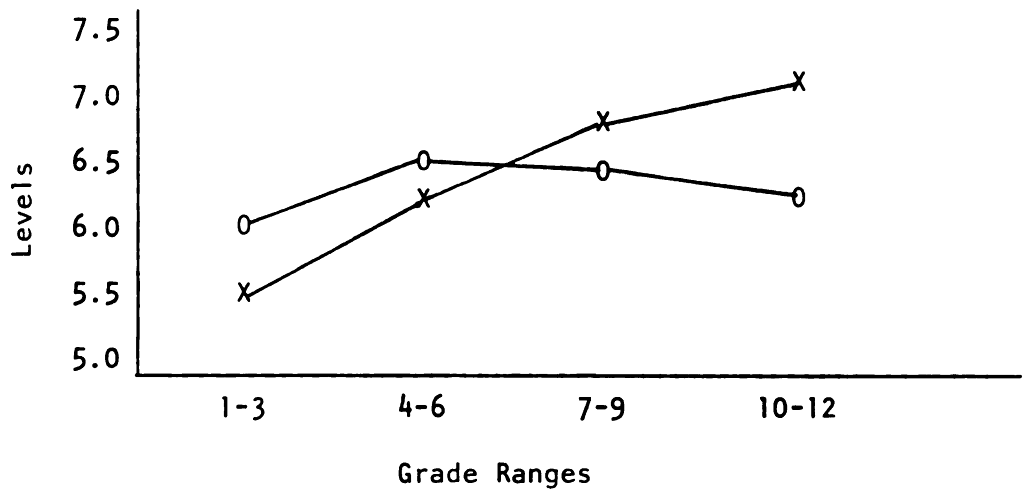


X = LD
0 = LA<26th

Means

LD	LA<26th
Grade Range 1-3 = 5.67	Grade Range 1-3 = 5.91
Grade Range 4-6 = 6.58	Grade Range 4-6 = 6.34
Grade Range 7-9 = 7.10	Grade Range 7-9 = 6.50
Grade Range 10-12 = 7.31	Grade Range 10-12 = 6.30

Figure 4. Adjusted discrepancy level means at grade ranges: LD and LA<26th groups.



X = LD
0 = LA<17th

Means

LD		LA<17th	
Grade Range 1-3	= 5.67	Grade Range 1-3	= 5.91
Grade Range 4-6	= 6.58	Grade Range 4-6	= 6.34
Grade Range 7-9	= 7.10	Grade Range 7-9	= 6.50
Grade Range 10-12	= 7.31	Grade Range 10-12	= 6.30

Figure 5. Adjusted discrepancy level means at grade ranges: LD and LA<17th groups.

IQ scatter (NDEV) variance. To ascertain the degree of influence covariates had on the significant variation between groups, separate univariate analyses with covariates were conducted. The results of these analyses are reported in Tables A6 and A7. None of the covariates (IQ level, sex, primary referral reason) was found to be significant in producing the observed variation in the overall analysis. The significance levels of *t*-values indicating covariate effect did not approach the .05 level, and there was no statistically significant change in the variation when compared with the overall analysis.

CHAPTER 5

DISCUSSION

Summary

The ability and achievement test performance of 404 students selected from a sample of 1,552 students from 45 school districts in Michigan was analyzed to determine if methods proposed in the literature could assist in formulating useful decision-making rules for classification purposes. Two hundred thirty students classified as LD were compared with 174 LA students selected on the basis of their achievement scores on a test of word recognition. Methods calculating *z*-score discrepancy levels, ability-test subtest scatter, verbal-performance ability differences, and general ability and achievement characteristics did not meet the criterion set in this study for usefulness in differential diagnosis. Methods were deemed useful if they affirmed classification decisions made by EPPCs for 85% or more of the students in the groups studied. The methods used to calculate discrepancy levels and ability subtest scatter indicated statistically significant differences at the .05 level of confidence. However, they were not found to be useful for practical diagnostic purposes.

In this chapter, conclusions and recommendations are discussed under the following areas: clinical practice, public policy, directions for educational research, and limitations of the study.

Conclusions and Recommendations for Clinical Practice

Comparison of the general ability and achievement characteristics of the LD and LA students included in this study offered little direction to the diagnostician who wishes to use the results for classification purposes. Although statistically significant variation was found to exist between the LD and LA<17th groups, this variation offered little in terms of formulating useful decision rules. When proximity to the student's mean was used as the basis for classification, the degree of overall error was 44% for Full Scale IQ comparisons and 41% for Verbal IQ comparisons. The high degree of correlation ($p = .001$) between achievement test results, both within groups and between groups, spoke little to any variation in terms of achievement. Group characteristics of IQ levels offered little for the purposes of differentiation. Forty-two percent of the LD group and 30% of the LA<26th group had IQ levels in the range of 90 to 111. Six percent of each group had IQs above 110; 52% of the LD group and 67% of the LA group had IQs below 90.

The range of ability and achievement scores was found not to vary between samples on any of the methods used. In this study, the diagnostic utility of using ability and achievement performance to differentiate between groups was found to be highly unproductive. Using ability and achievement characteristics for differential purposes was found to produce an unacceptably high error rate, with a range from 41 to 54%.

The use of ability-test scatter as a diagnostic tool offers just as little promise. The most promising method appeared to be the range method (highest minus lowest scaled score), which identified only 70% of the LD and approximately 30% of the LA students. Simply put, this method, though most successful, leaves much to be desired for realistic classification purposes.

Verbal-performance differences using two methods were unsuccessful in differentiating between the LD and LA students in this study. Although the actual 12-point-difference method was most successful in identifying classifications, the percentage of false-positive and false-negative diagnosis far exceeded an acceptable level.

The only discriminant factor (covariate) that produced changes in significant overall analyses was found to be IQ level. Even with that effect, the general trend of discrepancy levels among groups was quite consistent.

In conclusion, none of the diagnostic methods tested in this study was found to have an acceptable level of utility (85%). It would seem that, to the extent that the sample of students used in the study was representative of the population referred for educational diagnoses, the use of general achievement and ability characteristics, z-score discrepancy levels, ability-test scatter, or ability-test verbal-performance differences for classification may be well intentioned but not accurate or realistic.

Educational practitioners who use the standardized test data included in this study and who focus on the methodology studied here should be advised that these methods have little practical value for classification purposes. Neither the literature nor the present study supports these practices. This statement is not intended to discourage judgments based on individual student processes observed in assessment, but rather to discourage strongly the practice of serving some students and denying others special programs on the basis of criteria that have no realistic basis.

Conclusions and Recommendations Concerning Public Policy

The methods used in this study failed to distinguish between a classification of LD and a nonhandicapped classification of low achievement. The data presented here suggest that considerable confusion may result when an attempt is made to distinguish school-defined LD students from their LA peers. In fact, this study failed to do so using student samples that were more characteristic of the referred population at large than those used in recent studies comparing the LD with the LA student. Public policy should disregard the notion that a label is considered a valid and separate category if it serves only a portion of the school population with no discernible impairment other than low achievement. A sufficiently large population of students is believed to exist, if this study's samples are representative of the general referred population, who are not being served. The reasons for this condition are still unclear. Public policy makers should be charged with identifying student characteristics and more clearly specified classification, without which neither clinical practice nor research can advance very far from the present state. Of central concern to this study is the question, Does the handicap of learning disabilities exist as it is presently defined in the federal definition? The results of this study suggest a negative response. Is the concept of learning disabled, as a separate and distinct group from the general population who is experiencing difficulty in the school environment, realistic? Again, the results of this study suggest a negative answer. The possibility exists that direction of efforts is realistic in creating a category of educationally handicapped and assisting educational personnel to develop the flexibility needed to address the needs of a sizable portion of the school population. Researchers and practitioners should lend their expertise to policy makers, whenever possible, in an attempt to define appropriate categories for services. Active participation by both parties may help ensure congruence between practical educational practices and legislative intention.

The number of students who would qualify for special education services under the current federal definition, even with an achievement-ability discrepancy level set marginally low, did not constitute a two-thirds majority in any group studies. Sixty-seven percent of the LD group and between 46 and 50% of the LA groups demonstrated a reliable discrepancy between ability and achievement, a figure too low for diagnostic utility. Even more surprising was the number of students scoring with negative directional discrepancies: 12% of the LD and 8 and 6% of the two groups of LA students. Even though between-group variance was found to exist, the number of students affecting such variance would well have been due to chance. The degree of misidentification that would result if the discrepancy method was applied for classification purposes exceeded, by far, an acceptable level. Thirty-five percent of the LD and 46% of the LA students would receive either a false-positive or a false-negative diagnosis. Given the federal definition for the determination of a learning disability, it is surprising that a sizable number of the students included in the LD group in this study exhibited no significant discrepancy between ability and achievement. The stringent requirements of the definition indicate that children may only be determined to have a learning disability if they do not achieve commensurate with their age and ability in one or more areas. The results of this study may well indicate that congruence between the nature (validity and purpose) of standardized tests and the uses to which they are applied does not exist. Using standardized tests to affirm teacher judgment is prevalent and highly accepted (Shinn et al., 1982), but the ability of such tests to differentiate between handicapping conditions may well be questioned.

Conclusions and Recommendations for Further Research

None of the analyses performed in this study met the criterion for usefulness in differential diagnosis as set in this research. Certainly the overall error rate was more generous than what is presently required in interpreting the federal regulations and than what would be acceptable to practitioners making educational decisions. Research efforts should be directed toward clarifying the nature of handicaps, the factors educators view as important to educational progress, and the appropriate measurement of these factors before realistic educational schema can be developed.

The fallibility of diagnostic decisions is an area of research that needs to be explored. The accuracy of methods used for differential diagnosis should be of concern to practitioners and those responsible for legislating standards for identifying the handicapped. The degree to which present special education populations meet the criteria mandated by state and federal regulations may provide an information base from which realistic criteria can be formulated. Aligned with these efforts should be an information base regarding the factors that are instrumental in initial placements of students. The degree to which test scores and ecological factors, as well as daily classroom behavior, influence the decisions of placement teams may provide information relative to criteria for the presence of a handicap. School personnel seem best able to assess the qualities necessary for academic progress. Using such information as diagnostic directions in the assessment process may assist in formulating practical definitions of the handicapped.

The majority of students included in this study were referred for academic reasons. Although this may not be representative of the entire referred population, the factors influencing teacher referrals should be pursued. What criteria do teachers use in deciding to refer a student for evaluation? Teacher awareness of individual differences and the degree to which students are viewed as requiring intervention may assist in developing appropriate definitions of handicaps.

Although it appears highly unlikely that results of the analyses performed in this study will vary upon replication, research may well be continued on samples that are as representative of the general population as the samples included in this study were of the referred population. It seems likely that large samples of students with specified low achievement scores can be obtained across the spectrum of grades. Analyses of characteristics and comparative studies with identified LD students may well offer validations that provide insight into recategorization of classifications (McKinney, 1984).

Other research efforts should be directed toward comparing functional differences between the LD and LA populations. Abandoning current practices that are not valid does not mean simply relying on teacher judgment. Informal measures of assessment may allow for efficient, expedient quantification of functional classroom performance as the basis for selecting students who should receive LD services. Comparative studies using representative samples across the school-age population appear to be nonexistent, but such studies might help classify and serve students experiencing academic difficulty. For example, although the male-female ratio in this study is characteristic of the general LD population, the issue of bias in referral and placement of females versus males needs exploration. Theoretically, there are few differences in the ability and achievement levels of males and females in the general population. If this is true, there may be a population of LD students who are not being referred and served in special programs.

Diagnostic-utility questions related to the specificity of handicapping conditions need to be addressed. Until this is accomplished, assessing group differences is redundant.

Limitations

Certain limitations were inherent in the study, and they should be noted when the results are generalized to the school-age population. Limitations regarding sample selection, methodology, and interpretation of results are of particular importance.

Sample Selection

1. As noted in Chapter 3, the classification decisions of the EPPCs were accepted as accurate. Uncertainty about the factors used in making these decisions is of concern when this population of LD and noneligible students is represented as being characteristic of the general or special education populations.

2. All of the students included in this study were referred for evaluation after attracting attention in the school environment. Generalizing the results to nonreferred students may not lead to accurate classification decisions or assessment techniques.

3. The source of the sample (mainly suburban and rural) and socioeconomic level (predominantly middle level) indicate biases that might affect generalization of the results. There might be a broader population of LD and LA students to whom the results of this study cannot be generalized.

4. Although the presence of a discrepancy between ability and achievement was a consideration in classifying LD students, there is no assurance that this was a primary consideration.

5. The data reported in this study were collected during the 1975-76 school year. Curriculum and methodological changes, as well as the nature of special education services, may well influence the

variability and composition of present LD students as compared with those included in this study.

6. The variability of psychometric competency used in data collection is unknown. Although all of the data used in this study were collected by certificated school psychologists, the variability of assessment situations might have produced inconsistent results.

Methodology

1. The results of this study were derived from standardized test performance. Analyses of group differences on informal or competency-based tests may yield significantly different results.

2. The achievement-test results used in this study were limited to three areas that might not reflect the academic (competency-based) tasks required in the general school environment. Hence generalization of the results to academic assessments other than those included in this study may be invalid.

Results

1. A primary concern in current diagnostic practices is the use of *z*-score discrepancy formulas to indicate practical (functional) differences between ability and achievement. The method used in this study is purported to measure statistically significant differences. Generalizing that a calculated statistically significant discrepancy indicates a need for special education services is strongly discouraged.

APPENDIX

Table A1. Univariate analyses for LD and LA<26th Groups

	Group by Grade Range Effect				Group Effect			Grade Range Effect		
	Within Cells MS	Between Group MS	F	p	Between Group MS	F	p	Between Group MS	F	p
Achievement Range (scatter)	331.36	449.17	1.36	.26	1.91	.01	.94	328.65	.99	.40
IQ Range (scatter)	44.17	34.69	.78	.50	73.12	1.66	.20	105.31	2.38	.07
IQ Range (mean deviation)	21.94	16.51	.75	.52	66.77	3.04	.08	55.56	2.53	.06
Achievement Range (mean deviation)	9.00	6.96	.77	.51	1.67	.18	.67	20.48	2.27	.08
IQ Scatter (NDEV)	26.16	20.52	.78	.50	137.27	5.25	.02	54.95	2.10	.10
IQ V-P Difference (actual)	.93	1.04	1.12	.34	2.77	2.98	.09	.36	.39	.76
IQ V-P Difference (mean deviation)	.55	.21	.39	.76	.54	.98	.32	1.01	1.85	.14
Discrepancy levels	3.60	5.65	1.57	.20	17.25	4.79	.03	13.47	3.74	.01

Table A2. Univariate analyses for LD and LA<17th Groups

	Group by Grade Range Effect				Group Effect			Grade Range Effect		
	Within Cells MS	Between Cells MS	F	p	Between Group MS	F	p	Between Group MS	F	p
Achievement Range (scatter)	334.91	467.68	1.40	.24	10.95	.03	.86	314.39	.94	.42
IQ Range (scatter)	43.52	37.53	.86	.46	60.24	1.38	.24	67.14	1.54	.20
IQ Range (mean deviation)	24.31	18.98	.78	.51	35.35	1.45	.23	57.56	2.37	.07
Achievement Range (mean deviation)	8.82	1.73	.20	.90	7.45	.84	.36	8.87	1.00	.39
IQ Scatter (NDEV)	28.79	17.45	.61	.61	80.87	2.81	.10	57.89	2.01	.11
IQ V-P Difference (actual)	.91	.83	.91	.44	1.77	1.94	.16	.32	.36	.78
IQ V-P Difference (mean deviation)	.55	.36	.66	.58	.58	1.07	.30	.81	1.48	.22
Discrepancy Levels	3.61	7.97	2.21	.09	6.09	1.69	.20	11.19	3.10	.03

Table A3. Means, standard deviations, correlation coefficients, and significance levels between subtest scores for Wide Range Achievement Test performance

Subtests	Subtests											
	Word Recognition				Spelling				Arithmetic			
	Mean	SD	Coe	SL	Mean	SD	Coe	SL	Mean	SD	Coe	SL
LD and LA<26th												
Word Recognition	2.60	1.77			2.45	1.59	.849	.001	2.86	1.59	.679	.001
Spelling			.849	.001							.679	.001
Arithmetic			.849	.001			.679	.001				
LD and LA<17th												
Word Recognition	2.53	1.70			2.38	1.45	.833	.001	2.87	1.56	.638	.001
Spelling			.833	.001							.638	.001
Arithmetic			.638	.001			.638	.001				
LD												
Word Recognition	2.49	1.81			2.53	1.66	.789	.001	2.86	1.61	.552	.001
Spelling			.789	.001							.554	.001
Arithmetic			.552	.001			.554	.001			.554	.001
LA<26th												
Word Recognition	2.71	1.72			2.53	1.66	.919	.001	2.86	1.61	.853	.001
Spelling			.919	.001							.823	.001
Arithmetic			.853	.001			.823	.001				
LA<17th												
Word Recognition	2.57	1.58			2.38	1.38	.923	.001	2.87	1.54	.821	.001
Spelling			.923	.001							.796	.001
Arithmetic			.821	.001			.796	.001				

Table A4. Correlations between dependent variables

	Achievement Range (Scatter)	IQ Range (Scatter)	IQ Range (Mean Deviation)	Achievement Range (Mean Deviation)	IQ Scatter (NDEV)	IQ Diff (Actual)	IQ Diff (Mean Deviation)	Discrepancy Levels
	coe/*p*	coe/*p*	coe/*p*	coe/*p*	coe/*p*	coe/*p*	coe/*p*	coe/*p*
Achievement Range (scatter)	0	.03/.30	.07/.11	.82/.00	.06/.15	.03/.26	.00/.47	.13/.01
IQ Range (scatter)	.03/.30	0	.25/.00	.07/.09	.32/.00	.04/.20	.10/.03	.19/.00
IQ Range (mean deviation)	.07/.11	.25/.00	0	.03/.26	.93/.00	.05/.20	.09/.05	.06/.15
Achievement Range (mean deviation)	.82/.00	.08/.10	.03/.26	0	.02/.34	.04/.21	.01/.42	.09/.06
IQ Scatter (actual)	.06/.15	.32/.00	.93/.00	.02/.34	0	.06/.12	.03/.27	.04/.22
IQ Diff (actual)	.03/.26	.04/.20	.05/.20	.04/.21	.05/.12	0	.19/.00	.09/.05
IQ Diff (mean deviation)	.00/.47	.10/.03	.09/.05	.01/.42	.03/.27	.19/.00	0	.10/.03
Discrepancy Levels	.13/.01	.19/.00	.06/.15	.09/.05	.04/.22	.09/.05	.10/.03	0

Table A5. Correlations between covariates

	Referral Reason coe/p	Sex coe/p	SES coe/p	District Type coe/p	Age Range coe/p	IQ Levels coe/p
Referral reason	0	.03/.27	.05/.21	.01/.45	.03/.31	.03/.30
Sex	.03/.27	0	.09/.06	.02/.36	.04/.24	.02/.36
SES	.05/.21	.09/.06	0	.04/.23	.00/.48	.15/.01
District type	.01/.45	.02/.36	.04/.23	0	.11/.02	.15/.00
Age range	.03/.31	.04/.24	.00/.48	.11/.02	0	.11/.01
IQ levels	.03/.30	.02/.36	.16/.01	.15/.00	.11/.01	0

Table A6. Univariate analyses with covariates: LD and LA<26th groups

	Group by Grade Range Effect				Group Effect			Grade Range Effect			Covariate Effect	
	Within Cells MS	Between Group MS	F	p	Between Group MS	F	p	Between Group MS	F	p	T-Value	p
Covariate: Primary Referral Reason												
IQ Scatter (NDEV)	22.40	24.37	.92	.43	150.07	5.68	.02	59.21	2.24	.08	.46	.65
Discrepancy levels	3.59	5.85	1.63	.18	14.52	4.04	.05	13.36	3.72	.01	1.24	.22
Covariate: Sex												
IQ Scatter (NDEV)	26.28	20.17	.77	.51	140.64	5.35	.02	55.19	2.10	.10	-0.42	.68
Discrepancy levels	3.60	5.29	1.47	.22	15.48	4.30	.04	13.68	3.80	.01	-1.31	.19
Covariate: IQ Level												
IQ Scatter (NDEV)	26.20	20.83	.79	.50	140.64	5.35	.02	57.61	2.20	.09	.65	.52
Discrepancy levels	1.86	6.71	3.61	.01	9.43	5.07	.03	31.30	16.82	0	19.25	0

Table A7. Univariate analyses with covariates: LD and LA<17th Groups

	Group by Grade Range Effect				Group Effect			Grade Range Effect			Covariate Effect	
	Within Cells MS	Between Group MS	F	p	Between Group MS	F	p	Between Group MS	F	p	T-Value	p
Covariate: Primary Referral Reason												
Discrepancy levels	3.59	8.27	2.30	.07	5.35	1.48	.22	11.27	3.14	.03	1.53	.13
Covariate: Sex												
Discrepancy levels	3.62	7.62	2.11	.10	5.36	1.48	.22	11.13	3.08	.03	-0.76	.50
Covariate: IQ Level												
Discrepancy levels	1.86	7.11	3.83	.01	.004	.002	.97	28.92	15.57	0	18.10	0

REFERENCES

REFERENCES

Ackerman, P., Peters, J., & Dyckman, R. Children with specific learning disabilities: WISC profiles. Journal of Learning Disabilities, 1981, 4, 1950-1966.

Algozzine, B., Forgnone, C., Mercer, C. D., & Trifiletti, J. J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives. Learning Disability Quarterly, 1979, 2, 25-31.

Algozzine, B., & Ysseldyke, J. E. Classification decisions in learning disabilities. Educational and Psychological Research, 1982, 2, 117-129. (a)

Algozzine, B., & Ysseldyke, J. E. Learning disabilities as a subset of school failure: The oversophistication of a concept (Research Report No. 69). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982. (b)

Anderson, M., Kaufman, A. S., & Kaufman, N. L. The use of the WISC-R with a learning disabled population: Some diagnostic implications. Psychology in the Schools, 1976, 13, 381-386.

Bateman, B. An educator's view of a diagnostic approach to learning disorders. In J. Hellmuth (Ed.), Learning Disorders (Vol. 1). Seattle: Special Child Publications, 1965.

Bender, L. Specific reading disability as a maturational lag. Bulletin of the Orton Society, 1963, 13, 25-44.

Bernard, R. A description of a population of school verified learning disabled children in the state of Michigan across certain select variables. Ph.D. dissertation, Michigan State University, 1978.

Birch, H. G., & Belmont, L. Auditory-visual integration in brain damaged and normal children. Develop. Med. Child Neurol., 1965, 7-135.

Bond, G. L., & Tinker, M. A. Reading difficulties: Their diagnosis and correction. New York: Appleton-Century-Crofts, 1957.

Brenton, B. W., & Gilmore, D. An operational definition of learning disabilities (cognitive domain) using WISC full scale IQ and Peabody Individual Achievement Test scores. _Psychology in the Schools_, 1976, _13_(4), 427-432.

Burns, E. Relation between ITPA average deviation and Stanford-Binet intelligence scores. _Psychology in the Schools_, 1976, _13(4)_, 385-387.

Chusid, J. G. _Correlative neuroanatomy and functional neurology_ (17th ed.). Los Altos, CA: Lange Medical Publications, 1979.

Clarizio, H., & Bernard, R. Recategorized WISC-R scores of learning disabled children and differential diagnosis. _Psychology in the Schools_, 1981, _18_, 5-12.

Clements, S. D. _Minimal brain dysfunction in children_. Washington, D.C.: U.S. Department of Health, Education, & Welfare, 1966.

Cowan, W. M. The development of the brain. In Scientific American (Ed.), _The brain_. San Francisco: W. H. Freeman, 1979.

Deno, S. L., & Mirkin, P. K. _Data based program modification: A manual_. Reston, VA: Council for Exceptional Children, 1977.

Divoky, D. Education's latest victim: The "LD kid." _Learning_, 1974, _3_, 20-25.

Dunn, I. M. _Peabody Picture Vocabulary Test_. Circle Pines: American Guidance Service, 1965.

Epps, S., McGue, M., & Ysseldyke, J. E. Interjudge agreement in classifying students as learning disabled. _Psychology in the Schools_, 1982, _19_, 209-220.

Epps, S., Ysseldyke, J. E., & Algozzine, B. _An analysis of the conceptual framework underlying definitions of learning disabilities_ (Research Report No. 98). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982. (a)

Epps, S., Ysseldyke, J. E., & Algozzine, B. _Public policy implications of different definitions of learning disabilities_ (Research Report No. 99). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982. (b)

Epps, S., Ysseldyke, J. E., & McGue, M. Differentiating LD and non-LD students: "I know one when I see one." _Learning Disability Quarterly_, 1982, _2_, 3-13.

Erickson, M. T. The Z-score discrepancy method for identifying reading-disabled children. Journal of Learning Disabilities, 1975, 8, 308-312.

Frame, R. E., Clarizio, H. F., Porter, A. C., & Vinsonhaler, J. R. Interclinician agreement and bias in school psychologists' diagnostic and treatment recommendations for a learning disabled child. Psychology in the Schools, 1982, 19, 319-327.

Gallagher, J. J. Children with developmental imbalances: A psychoeducational definition. In W. M. Cruickshank (Ed.), The teacher of brain-injured children. Syracuse, NY: Syracuse University Press, 1966.

Geschwind, N. The development of the brain and the evolution of language. In C. I. Stuart (Ed.), Monograph series on language and linguistics (Vol. 17). Washington, D.C.: Georgetown University Press, 1964.

Gutkin, T. B. WISC-R scatter indices: Useful information for differential diagnosis? Journal of School Psychology, 1979, 17, 368-371.

Harris, A. How to increase reading ability (4th ed.). New York: McKay, 1970.

Hobbs, N. The future of children. San Francisco: Jossey-Bass, 1975.

Hull, C. H., & Nie, N. H. SPSS update 7.9: New procedures and facilities for releases. New York: McGraw-Hill, 1981.

Jastak, J. F., & Jastak, S. R. The wide range achievement test. Wilmington: Guidance Associates, 1965.

Kaufman, A. A new approach to the interpretation of test scatter on the WISC-R. Journal of Learning Disabilities, 1976, 9, 160-168. (A)

Kaufman, A. Verbal-Performance IQ discrepancies on the WISC-R. Journal of Consulting and Clinical Psychology, 1976, 44, 739-744. (b)

Lerner, J. W. Children with learning disabilities (2nd ed.). Boston: Houghton-Mifflin, 1976.

Lerner, J. W. Learning disabilities: Theories, diagnosis and teaching strategies (3rd ed.). Boston: Houghton-Mifflin, 1981.

Lidz, C. W. Improving assessment of school children. San Francisco: Jossey-Bass, 1981.

Matarazzo, J. D. Wechsler's measurement and appraisal of adult intelligence (5th ed.). Baltimore: Williams and Williams, 1972.

Maurer, A. Whatever happened to witches? Journal of School Psychology, 1972, 19, 107-110.

McKinney, J. D. The search for subtypes of specific learning disability. Journal of Learning Disabilities, 1984, 17(1), 43-50.

Mercer, C. D. Children and adolescents with learning disabilities. Columbus, OH: Charles E. Merrill, 1979.

Mercer, C. D., Forgnone, C., & Wolking, W. D. Definitions of learning disabilities used in the United States. Journal of Learning Disabilities, 1975, 9, 376-386.

Monroe, M. Children who cannot read. Chicago: University of Chicago Press, 1932.

Monroe, M. Monroe Diagnostic Reading Test. Chicago: Stoelting, 1928.

Myklebust, H. R. Learning disabilities: Definition and overview. In H. R. Myklebust (Ed.), Progress in learning disabilities (Vol. 1). New York: Grune and Stratton, 1968.

Norman, C. A., & Zigmond, N. Characteristics of children labeled and served as learning disabled in school systems affiliated with child service demonstration centers. Journal of Learning Disabilities, 1980, 13, 542-547.

Piotrowski, R. Abnormality of subtest score differences on the WISC-R. Journal of Consulting and Clinical Psychology, 1978, 46, 569-570.

Reynolds, C. R. The fallacy of "two years below grade level for age" as a diagnostic criterion for reading disorders. Journal of School Psychology, 1981, 19(4), 350-358.

Reynolds, M. C., & Wang, M. C. Restructuring "special" school programs: A position paper. Paper presented at the national invitational Conference on Public Policy and the Special Education Task of the 1980s." Racine, WI, September 1981.

Schere, R. H., Richardson, E., & Bialer, I. Toward operationalizing a psychoeducational definition of learning disabilities. Journal of Abnormal Child Psychology, 1980, 8, 5-20.

Schooler, D., Beebe, M., & Koepke, T. Factor analysis of WISC-R scores for children identified learning disabled, educable mentally impaired, and emotionally impaired. Psychology in the Schools, 1978, 15, 478-485.

Shepard, L. An evaluation of the regression discrepancy method for identifying children with learning disabilities. Journal of Special Education, 1980, 14(1), 79-89.

Shepard, L. A., Smith, M. L., & Vojir, C. P. Characteristics of pupils identified as learning disabled. American Educational Research Journal, 1983, 20(3), 309-331.

Shinn, M. R., Ysseldyke, J. E., Deno, S., & Tindal, G. A comparison of psychometric and functional differences between students labeled learning disabled and low achieving (Research Report No. 71). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

Tabachnick, B. Test scatter on the WISC-R. Journal of Learning Disabilities, 1979, 12, 626-628.

Taylor, H. G., Satz, P., & Friel, J. Developmental dyslexia in relation to other childhood reading disorders: Significance and clinical utility. Reading Research Quarterly, 1979, 15(1), 84-101.

Thompson, R. The diagnostic utility of WISC-R measures with children referred to a developmental evaluation center. Journal of Consulting and Clinical Psychology, 1980, 48, 440-447.

Thurlow, M. L., & Ysseldyke, J. E. Current assessment and decision-making practices in model programs for the learning disabled. Learning Disability Quarterly, 1979, 2, 15-24.

Tucker, J. A. Ethnic proportions in classes for the learning disabled: Issues in nonbiased assessment. The Journal of Special Education, 1980, 14, 93-105.

Tucker, J. A., Stevens, L. J., & Ysseldyke, J. E. Learning disabilities: The experts speak out (Research Report No. 77). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

U.S. Office of Education. Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. Federal Register, 1977, 42, 65082-65085.

Vaughan, R. W., & Hodges, L. A statistical survey into a definition of learning disabilities: A search for acceptance. Journal of Learning Disabilities, 1973, 6, 658-664.

Veres, E. An evaluation of the use of WISC-R Subtest scores for the identification of mildly handicapped students. Ph.D. dissertation, Michigan State University, 1982.

Warner, M. M., Schumaker, J. B., Alley, G. R., & Deshler, D. D. Learning disabled adolescents in the public schools: Are they different from other low achievers? Exceptional Education Quarterly, 1980, 1, 27-35.

Wechsler, D. Manual for the Wechsler Intelligence Scale for Children--Revised. New York: The Psychological Corporation, 1974.

Ysseldyke, J. E., & Algozzine, B. Perspectives on assessment of learning disabled students. Learning Disability Quarterly, 1979, 2, 3-13.

Ysseldyke, J. E., Algozzine, B., Shinn, M. R., & McGue, M. Similarities and differences between low achievers and students classified learning disabled. Journal of Special Education, 1982, 16, 73-85.

Zigmond, N. A prototype of comprehensive services for secondary students with learning disabilities: A preliminary report. Learning Disabilities Quarterly, 1978, 1, 39-49.