This is to certify that the

dissertation entitled

An evaluation of the use of WISC-R subtest scores for the identification of mildly handicapped children.

presented by

Valerie E. Veres

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling, Educational Psychology & Special Education

Harvey F Clarizio
Major professor

Date 11-22-82

AN EVALUATION OF THE USE OF WISC-R SUBTEST SCORES FOR THE IDENTIFICATION OF MILDLY HANDICAPPED CHILDREN

By

Valerie Etta Veres

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

1982

ABSTRACT

AN EVALUATION OF THE USE OF WISC-R SUBTEST SCORES FOR THE IDENTIFICATION OF MILDLY HANDICAPPED CHILDREN

By

Valerie Etta Veres

The WISC-R subtest scores of 703 children from 45 school districts throughout the State of Michigan were analyzed for characteristic diagnostic group patterns. The children, all of whom had been referred for psychological testing, were categorized as learning disabled (234), educable mentally impaired (77), emotionally impaired (64), and physically or otherwise handicapped (38). The remaining 290 children were neither diagnosed as educationally handicapped nor placed in any special education program. These not-placed children served as a logically appropriate normal control group for the major analyses.

Historically, a number of diagnostic rules and guidelines have been derived from observations of supposed group related differences in WISC-R profiles. In the area of learning disabilities, Verbal - Performance IQ discrepancies, abnormal subtest scatter, and particular patterns of recategorized scores have been suggested as diagnostic indicators. Both Verbal - Performance IQ discrepancies and a specific Information/

Similarities subtest pattern have been proposed as useful signs of emotional impairment. Throughout the research supporting these formulations there has been a general absence of appropriate control comparisons as well as a paucity of evidence regarding the diagnostic utility of the suggested decision rules.

In the present investigation, diagnostic utility was the primary concern. Any diagnostic rule that erred in its classification of 15 or more percent of cases to which it was applied was rejected as not useful. Despite the occasional observance of a statistically significant ($p < .05$) group level difference, none of the diagnostic rules met the less than 15 percent error criterion. In fact, no rule using linear combinations of WISC-R information to classify cases into diagnostic groups was successful for even 70 percent of cases.

Results were discussed in terms of their implications for public policy, clinical practice, and educational research.

A set of additional minor analyses focused on the validity of shortened versions of the WISC-R for particular subgroups.

Statistical techniques employed in this study included multiple regression, discriminant function and classification, analysis of variance, and a form of multivariate profile distance analysis.

ACKNOWLEDGEMENTS

The members of my dissertation committee have each contributed from their separate areas of expertise to the completion of this work and to my graduate training in general. I hope they realize that I appreciate their guidance more than I will be able to express in these few words.

Harvey Clarizio, my dissertation chairman, provided direction in my academic program and much encouragement in this research. He seemed to know when to push and when to wait.

Bill Mehrens has been patient enough to explain the mysteries of measurement. I have appreciated his ability to generate relevant practical examples.

Geb Blom has been important to both my graduate education and my personal growth. His consistent and generous encouragement laid the foundation for my professional confidence. I hope, someday, to approach his level of clinical insight.

Helen Benedict has been friend, confidant, and teacher. In each of these roles she has helped me to synthesize seemingly disparate phenomena. I know that I will continue to learn from her over the years.

Among the friends who have been important to me during my time in graduate school, there are a few whose support deserves special mention.

Bill Given has made these years much easier by providing the institutional support I needed to continue my education.

Thanking Judy Zivick for the care and patience she showed while typing so many papers from my handwritten drafts is not sufficient. I also want her to know that I have appreciated her friendship and encouragement.

Vicki Powell has been like a sister to me. The list of things that I would thank her for would take more pages than the rest of this document. Through life's changes, past and future, our friendship continues to grow.

Were it not for Craig Hauenstein's faith in me, I might not have started graduate school. For many years, Craig provided an environment that made it possible to balance the many competing demands on my time. He sustained me during the hardest times I have ever experienced.

More than anyone else, my husband, Jack Condon, has made the completion of this dissertation possible. His willingness to read and discuss drafts and his thoughtful and insightful comments have helped to clarify much of the text. His statistical and analytic expertise is evident throughout this work. He has taught me to appreciate some of the more esoteric aspects of psychological research. Although, at times, I produced more tears, excuses, and other forms of resistance than actual work, I always knew that he believed that I would finish. In a way, I know that it is not necessary to thank Jack, because whatever we do, we do for each other, with love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Assessment of Educational Disabilities

One of the primary aims of the Great Society programs of the 1960's was a systematic assault on all forms of educational disadvantage. Gradually, the concerns that engendered these programs led to the establishment of particular classes of children who were to be given educational opportunities tailored to their specific deficits.

Although the laws and regulations that established these educational categories provided some guidelines for their application, the job of interpreting the rules in specific cases fell to educators, and in most instances, to school psychologists.

For some types of educational handicap, diagnosis is reasonably obvious. The identification of children who have physical handicaps such as blindness and deafness, or of those with an extremely low overall level of intellectual functioning is a relatively straightforward task. However, for other categories of educational problems, such as learning disability and emotional impairment, the diagnostic criteria have been, to say the least, more vague.

The learning disabled classification evolved from early notions about unspecified neurological dysfunction, through several ideas about reading or language deficits, to its current conception which includes children whose educational failures are thought to arise from dysfunction in some basic psychological process.

The emotionally impaired classification is intended for children whose educational difficulties seem to be caused by an inadequate psychological adjustment that may have additional effects on mood and peer relationships.

The task facing the clinician who must decide whether or not a child fits into one of these two categories is, by no means, well-defined. In fact, even the most casual perusal of the relevant educational journals will bear out the difficulties associated with the assessment of learning disabilities, and, to a lesser extent, that of emotional impairment.

Uneven Intellectual Abilities

Early work on the identification of learning disability was based on the implicit assumption that children who are having difficulty succeeding in school, which can be traced neither to a physical handicap nor to mental retardation, must be deficient in some component of cognitive ability. This assumption, coupled with the notion that intelligence tests assess the structure of cognitive processing, led to the search for characteristic patterns in the intelligence test profiles of these children. The widespread use of the Wechsler Intelligence

Scales (WISC/WISC-R) and the availability of their subtest scores led naturally to hypotheses regarding connections between particular patterns of scores and learning disabilities. The literature on this topic is reviewed in the next chapter. For now, it is sufficient to note that the early work in the area was somewhat encouraging in that patterns that seemed to characterize the subtest scores of learning disabled children were identified.

Uneven function hypotheses have also appeared in the emotional impairment literature. The rationale behind these formulations is less clear than in the learning disabilities area. To some extent they were guided by the expectation that psychosocial deficits would have a more profound effect on verbal than on non-verbal abilities. In this area too, the early research results seemed promising.

The Mythical Conception of Normality

Until recently, it was assumed that the "normal" child had a uniform distribution of cognitive abilities. After all, the *mean* score on each of the WISC/WISC-R subtests is, in fact, its *norm*. It was not until the mid- 1970's that the extent of variation in the profiles of normal childen was realized. Up to that time, most research proceeded with neither control groups nor any other form of base rate information. It became evident, to some, that the rather variant subtest patterns that had been thought to be indicative of educational disabilities were present in the profiles of many children who were experiencing no educational difficulties.

The implications of the more recent data for educational diagnosis have yet to be specified completely. Many practitioners still feel justified in using previously established variant patterns as indicators of specific educational handicaps.

Synopsis of the Present Investigation

The diagnostic utility of the patterns that were once considered to be reliable signs of educational handicaps is far from established. In fact, the bulk of the available empirical evidence now suggests that profiles that are characteristic of either learning disability or emotional impairment will not be found. The range of normal within-person variation of specific cognitive abilities may even be as great as the average between-person variation in these specific abilities. If this is the case, then the search for an abnormally variant pattern is, by definition, futile.

The primary concern of the present investigation was the analysis of the practical value of WISC-R subtests for differential diagnosis. While there are some findings of significant profile differences in comparisons of educationally handicapped children with normal samples, there is little evidence that the differences observed are of any diagnostic import. The analyses in the present study put questions about diagnostic utility at the forefront. For each of the supposed characteristic patterns, the probabilities of false decisions were estimated. Perhaps a series of studies like the present one, using different samples of children, will help to end any misuse of WISC-R subtests for differential diagnosis.

CHAPTER 2

LITERATURE REVIEW

Overview

The preceeding chapter contained a history of attempts to use WISC and WISC-R response patterns to differentiate children with specific educational handicaps. The literature associated with these endeavors is reviewed in this chapter. The sections of the chapter contain reviews of research and theoretical speculation in the areas of: (1) characteristic Verbal-Performance IQ discrepancies among learning disabled children, (2) subtest scatter in the profiles of learning disabled children, (3) attempts to recategorize subtest scores to detect learning disabilities, (4) the discrimination of learning disabilities using all of the WISC-R subtests, (5) age and Full-Scale IQ differences among learning disabled children, (6) Verbal - Performance discrepancies among emotionally impaired children, (7) more specific patterns suggested as being characteristic of emotional impairment, (8) the discrimination of various educational disabilities using all of the WISC-R subtests, and (9) the search for a valid short-form version of the WISC-R. The final section of the chapter contains a statement of the direction of the present investigation.

Learning Disabilities: Verbal-Performance Discrepancy

One of the early attempts to link particular WISC patterns with the learning disabled classification was based on the observation of differences between the Verbal and Performance IQ scores of learning disabled children. A reliable finding of higher Performance IQ among the learning disabled in the absence of a similar pattern among the non-learning disabled could have provided a useful diagnostic indicator.

Studies examining the significance of the discrepancy between the Verbal and Performance subscale scores on the Wechsler intelligence tests (WISC and WISC-R) have appeared for a number of years. More than a decade ago, Ackerman, Peters, and Dykman (1971) wrote that their research suggested "discordance between WISC Verbal and Performance IQ's (15 points or greater) was somewhat more frequent in the [LD] sample, and more commonly in favor of Performance IQ" (p. 48). In 1976, Anderson, Kaufman, and Kaufman reported a statistically significant ($p < .05$) V-P difference (regardless of sign) for children diagnosed as learning disabled. However, they qualified their interpretation of this finding by acknowledging that, "The average LD child in this study did *not* have an unusually large discrepancy, despite the statistical significance that was obtained" (p. 385). The Smith, Coleman, Dokecki, and Davis (1977a) study supported the Ackerman, *et al.* findings. They reported that in a large group ($N = 208$) of learning disabled children, Performance IQ was significantly higher than Verbal IQ. Smith, *et al.*, who examined both "high" and "low" IQ learning disabled children, concluded "with a high degree of confidence that in similar samples of children, Verbal IQ's are less than Performance IQ's regardless of the relative level of intellectual functioning" (p. 355).

One of the problems that characterized the early research on V-P differences was the universal absence of base-rate data from the normal population. Kaufman (1976a) reported a Verbal-Performance difference of nine or more points in about half (48 percent) of the cases in the WISC-R standardization sample. Moreover, 34 percent of those in the sample had V-P differences of 12 or more points, and 25 percent had differences of 15 points or greater. Kaufman also noted that neither a verbal greater than performance nor a performance greater than verbal pattern predominated within any age group.

Kaufman's study is important in that it provided an estimate of the prevalence of V-P discrepancies in the normal population. It was not unreasonable to expect future studies of exceptional subgroups to take these data into account.

Despite the obvious implications of Kaufman's report, Smith, *et al.* (1977a, 1977b) published later studies that included no reference to either a normal control group or to Kaufman's reported V-P discrepancies in the standardization sample.

The bases for Kaufman's conclusions were the standard deviations of the Verbal and Performance IQ subscales. Piotrowski (1978) suggested the need for a more quantitatively sound method of assessing the "abnormality" of a difference. He stressed that the estimate chosen should give more information than does the standard error of measurement. The procedure that he chose was adapted from the work of Silverstein (1973), who pointed out that the standard error of measurement indicates how large a sample difference must be so that it cannot be attributed to chance, while the abnormality measure that he proposed reveals how large a difference must be for there to be little likelihood of it occurring in a normal population. The application of this measure led Piotrowski to suggest a more conservative approach to the interpretation of Verbal-Preformance differences than is usually employed. The abnormality index for estimating the difference between these two scores indicated that at the .05 level a separation of 18 points was needed for a difference to be considered "abnormal", while at the .01 level, a 24 point difference was required. Using these criteria, only 4 percent of the standardization sample had differences that were significantly abnormal at the .01 level.

The combination of Kaufman's (1976a) and Piotrowski's (1978) observations made it apparent that studies of the distinctiveness of the learning disabled category would. have to acknowledge the amount of Verbal-Performance dispersion typically found in a normal sample. In fact, a number of studies that recognized this problem followed.

Schooler, Beebe, and Koepke (1978) found that learning disabled groups demonstrated few V-P differences when they were compared to other groups. Gutkin (1979) reported that, "Even though the current investigation found statistically significant differences on the V-P, FSIQ, and PIQ scatter indices, there was a substantial overlap in the distributions for the normal and special education children" (p. 371).

In a similar vein, Thompson (1980) found that "The VIQ - PIQ mean discrepancy scores did not differ significantly among the various clinical groups or between any of the clinical groups and the standardization sample" (p. 445).

Despite the paucity of support for the use of Verbal-Performance differences for differentiating a learning disabled group from a normal population, researchers and practitioners in the field continue to look for a means to use this type of data to define clinical groups. Apparently, appreciation of Kaufman's and Piotrowski's admonitions does not extend to all individuals who work and publish in the area of learning disabilities (cf. Smith, *et al.*, 1977a, 1977b; Strichart and Love, 1979). In fact, Kaufman (1981) continued to suggest the necessity for research in this area. While he admitted that, "We have learned that normal children do not have flat WISC-R profiles, and that virtually all exceptional samples do not possess the stereotypical high V-P IQ discrepancies" (p. 525), he also stated that, "Research remains to be conducted to show that an abnormal V-P discrepancy of subscore scatter index is highly predictive of learning disabilities or other exceptionalities" (p. 524).

In summary, while there seems to be clear evidence that the observed Verbal-Performance IQ differences among learning disabled children are mirrored by similar discrepancies in non-learning disabled groups, it is also evident that the publication of studies that lack appropriate normal controls has continued.

Learning Disabilities: Subtest Scatter

The focus on Verbal-Performance IQ discrepancies is an example of a specific pattern thought to characterize the profiles of learning disabled children. A more general hypothesis, that has received quite a bit of attention, is that the WISC/WISC-R subtest scores of learning disabled children are abnormally variant. In other words, it has been suggested that the subtest profiles of learning disabled children are more scattered than would be expected in a non-learning disabled population. If this hypothesis were accurate, then abnormal subtest scatter might provide a guide for the diagnosis of learning disabilities.

Kaufman (1976b) used two indices of subtest scatter in an analysis of the WISC-R standardization sample. These were range (defined as highest minus lowest subtest scaled score) and deviations (NDEV, defined as the number of subtest scaled scores deviating by three or more points from the mean of the child's scaled scores). Kaufman found that on the average children in the standardization sample had 7 $\pm$ 2 points separating their highest and lowest subtest scaled scores (i.e., mean range = 7 $\pm$ 2). For the other index of scatter, deviations, he reported that the average child in the standardization sample had 1.7 $\pm$ .02 subtest scores that were three or more points different from that child's own subtest mean (i.e., mean NDEV = 1.7 $\pm$ .02). Those children who had higher overall IQ's exhibited more range than did lower IQ children.

In spite of the availability of Kaufman's base rate data, Tabachnick (1979), in her comparison of learning disabled and normal samples, reported greater range for the learning disabled group (mean = 7.7) than for her normal group (mean = 7.0). Although this result apparently met the requirements of statistical significance for the size of Tabachnick's sample, her learning disabled result is well within boundaries that Kaufman found in the standardization sample. Tabachnick reported no between-groups difference in deviations. She did qualify her results by stressing that, "A recommendation for diagnosis on the basis of subtest scatter alone, however, is not forthcoming. The overlap in subtest scatter between learning disabled and normal children is substantial and some learning disabled children in our sample are characterized by exceptionally *low* scatter" (p. 62).

Gutkin (1979), in a study that included several special education groups (i.e., learning disabled, emotionally impaired, brain-impaired, and educable mentally retarded) as well as a normal group, found significantly more subtest range within both the Verbal and Performance subscales and across the test as a whole for the four special education groups when compared with the normal controls. However, Gutkin found it difficult to interpret this result, as approximately 40 percent of the normal children in his study either equalled or exceeded the average score of the special education children on the ranges of all three (sub)scales.

Thompson (1980) came to much the same conclusion as Gutkin when he stated that, "The general scarcity of significant differences among clinically meaningful and diverse groups of children on various measures of subtest scatter, and the similarity of these subtest scatter measures to that reported for the normative sample, along with the magnitude of subtest scatter in the normative sample, would again suggest caution with regard to the empirical basis for diagnostic inferences based on WISC-R scatter" (p. 446). Other researchers (Ryckman, 1981; Henry and Wittman, 1981; Gutkin, 1979; Clarizio and Bernard, 1981) have also reported considerable overlap in the distributions of WISC-R subscale scatter of learning disabled and normal samples.

At this point, it would be premature to suggest an unequivocal summary of the data on the relationship between WISC-R subtest scatter and learning disabilities. On one hand, there is evidence from a few studies that learning disabled children have significantly more scatter in their subtest profiles than do those children in the control samples. Nevertheless, a large proportion of profiles of normal children show more scatter than that of the average learning disabled child. It would seem that there could be some aspect of learning disability that contributes to the slight differences in sample scatter, and that might manifest itself in a more reliable and diagnostically useful specific subtest pattern. The next section of this chapter reviews some of the work directed at isolating such a pattern.

Learning Disabilities: Bannatyne/Kaufman Recategorizations

Despite the lack of reliable indication that the scatter in the WISC-R profiles of learning disabled children differs in a diagnostically meaningful way from that observed among normal children, there has been a persistent search for a *specific* WISC-R profile that is characteristic of learning disabled children. Much of this research stems from Bannatyne's work with a population he labeled genetic dyslexics. Bannatyne (1968) regrouped the WISC subtest scores into three factors rather than the two suggested by Wechsler (1948). He postulated that more could be learned about the genetic dyslexics by recategorizing the WISC subtest scores into Spatial (SP), Conceptual (CP), and Sequential (SQ) composites. His Spatial category included the Picture Completion, Block Design, and Object Assembly subtests. The Conceptual score was comprised of Similarities, Vocabulary, and Comprehension. The third category, Sequential, contained Picture Arrangement, Coding, and Digit Span. Bannatyne supposed that "true" (i.e., genetic) dyslexics would show a pattern of SP > CP > SQ.

Rugel's (1974a, 1974b) factor analytic studies of the WISC supported Bannatyne's contention that there was a third factor to be considered. In addition, Rugel extended the use of the three factor model to describe disabled readers as well as genetic dyslexics. He refined Bannatyne's sequential category by substituting the Arithmetic subtest for Picture Arrangement. However, Rugel warned that, while there was support for this regrouping of the WISC subtests for theoretical purposes, the practical implications of his findings were, to some extent, problematic. Indeed, he cautioned that, "Not all disabled readers will be highest in spatial subtests and lowest in memory [i.e., Sequential] subtests. Some may be lowest in spatial subtests whereas others may show no difference at all" (Rugel, 1974a, p. 584). In part due to the lack of success in specifying a profile characteristic of disabled readers, Rugel argued for a more broad definition of academic difficulty. He suggested the existence of a subclass of readers, more encompassing than the genetic dyslexic category, who experienced inordinate difficulty in mastering reading.

Bannatyne (1974) accepted Rugel's reformulation of the Sequential category. He also posited the existence of another factor in the WISC subtest which he thought was indicative of general school related difficulties. He called this dimension Acquired Knowledge which he defined as the mean of the Information, Arithmetic, and Vocabulary subtests.

By the early 1970's the practice of regrouping the WISC subtest scores was becoming popular because it seemed to provide a reasonable response to a growing concern for the legal rights of handicapped children. The National Advisory Committee on Handicapped Children (1968) developed a definition of "specific learning disability" which was incorporated in the Education for All Handicapped Children Act (Public Law 94-142) in 1975. That definition included:

> Those children who have a disorder in one or more of the basic psychological processes involved in understanding or using language, spoken or written, which disorder may manifest itself in imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. Such disorders include such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include children who have learning problems which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, or of environmental, cultural, or economic disadvantage. (*Federal Register*, 1976, p. 52404)

One of the problems that practitioners had in interpreting this legislation was its failure to provide defensible quantitative guidelines for the identification of eligible children. As a consequence of this, the WISC recategorization schemes became quite appealing because they offered "objective" criteria for identifying learning disability; this, in spite of the warnings against practical application of the tentative results of this line of research.

In the mid- 1970's, when the WISC was revised, Kaufman (1975) confirmed the presence of a third factor in the standardization sample of the newly developed WISC-R. This factor, which Kaufman labeled Attention-Concentration, corresponded to the Sequential category that had been discussed by Bannatyne and Rugel.

The confirmation of a third factor in the WISC/WISC-R subtest scores gave some amount of impetus to the practice of recategorizing the subtests (Kaufman, 1976b, 1979). However, it is important to note that a connection between the three-factor model and the identification of learning disabilities was not a necessary corollary of the mere existence of the third factor. Hence, the pressure to arrive at a quantitative definition of learning disability, coupled with the evidence supporting the weak (see Swerdlick, 1978) third factor, led to a number of attempts to discover whether a *particular* factor structure characterized the WISC-R profiles of learning disabled children (Smith, *et al.*, 1977a, 1977b, 1978; Schooler, *et al.*, 1978; Gutkin, 1979; Dean, 1980; Vance, Wallbrown, and Blaha, 1978; Clarizio and Bernard, 1981; Reynolds, 1981; Reynolds and Gutkin, 1981; Ryckman, 1981).

Smith, *et al.*, (1977a, 1977b) reported that high and low IQ learning disabled children had similar WISC-R factor structures. They concluded that this similarity was evidence of an underlying characteristic learning disabled pattern. This conclusion, in the absence of a non-learning disabled comparison group, was, to say the least, unwarranted. Their later replication, which suffered from the same low-level methodological flaw, offered no more compelling evidence in support of their conclusion.

The Smith, *et al.* research program typifies those studies that might be used to support the existence of a distinctive pattern among the recategorized subtest scores of learning disabled children. The common thread running through these studies is their universal lack of non-impaired comparison samples (Vance, *et al.*, 1978, 1979; Gutkin, 1979; Dean, 1980; Stevenson, 1980).

In the few studies of recategorized scores that did have appropriate controls there was no convincing support for the use of these composites for differential diagnosis (Schooler, *et al.*, 1978; Ryckman, 1981; Clarizio and Bernard, 1981).

It is surprising that at least some researchers (and journal editors) in this area seem reluctant to allow the force of the empirical evidence to defeat their cherished hypotheses. Indeed, the first sentence of the abstract of Reynolds's most recent article on the topic indicated a presumption that the case for the use of recategorized scores has been proven: "Bannatyne's recategorization of Wechsler subtests into the process oriented categories of Spatial, Conceptual, and Sequential has been shown to be useful in differentiating groups of normal from reading disabled children and may have utility for understanding the cognitive functioning of individual children" (Reynolds, 1981, p. 468). If Reynolds's statement is indicative of the prevailing view of any subgroup of educators, then more demonstrations of the rate of diagnostic failure that accompanies the use of the recategorized scores are necessary.

Perhaps Miller's (1980) admonition regarding "preconceived conclusions" guiding research on WISC-R subtest patterns is worthy of more attention. After numerous attempts to verify the diagnostic utility of *particular* WISC-R profiles, it might be time to take a step back and ask the more general question of whether there is *any* information in the factorial composition of the subtests that could be used to diagnose learning disabilities. Certainly, no single study will be sufficient to answer this question. In fact, one purpose of the present effort is to provide a model for the identification of possibly useful diagnostic indicators in similar data sets. When a number of such studies using diverse samples have been accomplished, and if the cumulative results indicate that the use of WISC-R data for the diagnosis of learning disabilities results in an unacceptably high rate of error, then, perhaps we will be able to admit that the horse we have been beating never was alive.

Discrimination of Learning Disabilities Using WISC-R Subtests

Although there is not much credible support in the literature for the Bannatyne/Kaufman recategorizations, there may be *some other* factorial arrangement of the WISC-R subtests that would separate learning disabled children from the non-impaired population. Some of the factor analytic studies, cited in earlier sections of this chapter, have utilized the subtest data to arrive at a solution for the learning disabled samples (Rugel, 1974a, 1974b; Schooler, *et al.*, 1978). Schooler and his associates even went so far as to compare the factor structure of the learning disabled group with that of the normal group. They concluded that the factor structures of the groups in their study were quite similar. However, the technique that they used to compare the solutions essentially involved "eyeballing" the factor loading matrices of the groups in search of striking differences. Such an approach is too imprecise to allow potentially important differences to emerge. An analysis routine that uses most of the information in the subtests while attempting to maximize between-group differences would be more appropriate for the question at hand. Discriminant function analysis meets these requirements insofar as it would identify the *best linear composites* of the subtests that provide separation between the learning disabled children and their non-disabled peers. This was the strategey chosen to address this question in the present research.

Learning Disabilities: Subgroups

Although there is no commonly accepted or legally defined lower boundary IQ for the learning disabled classification, there seems to be a general consensus that children with overall depressed intellectual functioning are not properly classified as learning disabled (Mercer, Forgnone, And Wolking, 1976). Rather, the learning disabled designation is reserved for children of normal intelligence who exhibit specific scholastic failures (Sattler, 1982).

Perhaps the profiles of higher IQ children who have been classified as learning disabled are different from those of children who have been called learning disabled who have lower IQ's. This possibility was explored in the research of Smith *et al.* (1977a) who reported no such differences. The present research contained a similar comparison. In addition, analyses were conducted to ascertain whether the Acquired Knowledge, Verbal, or Performance subscales or the Full Scale IQ's of the learning disabled and normal children were associated with either grade level or the interaction of grade level with diagnostic group.

Emotional Impairment: Verbal-Performance Discrepancy

Efforts to utilize WISC/WISC-R data to identify emotionally impaired children have paralleled the work in the area of learning disabilities. Most of the research and speculation on this topic has centered around the relationship between Verbal-Performance IQ discrepancies and emotional impairment.

A number of researchers have shown that the Performance IQ's of groups of emotionally impaired children significantly exceeded their Verbal IQ's. The samples selected for these studies have varied within a range generally accepted as emotionally impaired. Andrew (1974) reported V-P differences in favor of Performance IQ among delinquent children on probation. Saccuzzo and Lewandowski (1976) found that the Performance IQ's of a sample of juvenile delinquents, whose Full Scale IQ's were between 80 and 89, were, on the average, higher than their Verbal IQ's, regardless of the children's race or sex. Lewandowski, Saccuzzo, and Lewandowski (1977) replicated this finding with a sample of delinquent children with Full Scale IQ's between 70 and 79. Dean (1977b) reported this same, Performance greater than Verbal, pattern among a sample of children with conduct disorders.

Although there have been studies that showed no V-P difference among the emotionally impaired (e.g., Bortner and Birch, 1969), the weight of the empirical evidence *would seem to favor* the existence of a Performance IQ greater than Verbal IQ pattern that is characteristic of emotionally impaired children. However, such a conclusion is unfounded, as none of the studies cited above contained a normal control group. This omission is especially disturbing in the light of Kaufman's (1976a) report of frequent large Verbal-Performance discrepancies in the WISC-R standardization sample.

In the present study, the examination of the utility of Verbal-Performance discrepancies for the diagnosis of emotional impairment included a normal control sample.

The use of a design that contains a control group is not a mere methodological nicety. Quite simply, there is no other way to ascertain the probabilities associated with false diagnoses.

In summary, the question of whether the Performance greater than Verbal pattern, observed in many studies of emotionally impaired children, distinguishes them from children who are not emotionally impaired remains unanswered. While Kaufman's analysis of the WISC-R standardization sample would suggest that these differences will not be particularly useful for diagnosis, there have, as yet, been no direct comparisons of emotionally impaired and normal children.

Emotional Impairment: High Similarities, Low Information Pattern

Dean's (1977b) review of the research of Wechsler and Jaros (1965) and Saccuzzo and Lewandowski (1976) led him to search for peculiar signs that characterized the WISC-R profiles of emotionally impaired children. He reported that, "In some 50% of the cases, the Information subtest was the lowest or second lowest Verbal subtest, whereas the Similarities subtest was the first or second highest Verbal subtest in approximately the same number of cases" (p. 489). Dean concluded that these signs would be useful aids to the clinician faced with problems in the diagnosis of emotional impairment.

Perhaps Dean's low Information - high Similarities pattern may be of some clinical use. However, the utility of the pattern is hard to judge, as Dean's study contained no normal control group. In the present study, the rate of occurrence of the Information/Similarities pattern in the emotionally impaired sample was compared to its prevalence in the normal sample, thus, providing the first methodologically proper estimate of its clinical utility.

WISC-R Subtests and Diagnostic Categories

It would seem that hypotheses regarding relationships between uneven intellectual functioning and various forms of educational impairment are quite abundant. What seem to be scarce, however, are empirical attempts to confirm these hypotheses that contain appropriate control comparisons.

The data that were available for secondary analysis in the present study included WISC-R subtest scores of children who fell into five educational categories (i.e., learning disabled, emotionally impaired, educable mentally impaired, otherwise handicapped, and a normal control group consisting of children who had been referred for testing but who had not been placed in any of the handicapped categories). The existence of this data set is fortunate in that it allowed analyses of the subtest scores for group related profile differences. The approaches taken, discriminant function analysis, and a form of profile distance analysis introduced by Cronbach and Gleser (1953), were chosen to allow any existing characteristic group profile differences to emerge.

Best Predictors of Full Scale Intelligence

Several short form versions of the WISC and WISC-R have been offered (Yudin, 1966; Satz and Mogel, 1967; Kaufman, 1976c). The validity of these shortened IQ tests rests solely on their ability to predict full scale intelligence with the least amount of multicolinearity or redundancy. The final three analyses of the present study identified the subtests that were the best predictors of WISC-R Full Scale IQ for the normal sample, the entire learning disabled sample, and those children in the learning disabled sample who were in the eighth grade or higher.

Direction of Present Study

As documented throughout this chapter, the literature on the Wechsler Intelligence Scales is replete with attempts to show that various arrays of subtest scores are indicative of specific educational handicaps. There have been some instances in which the subtest score arrangements of comparison groups have differed significantly. Furthermore, statistically significant differences in specific abilities have, at times, been identified within certain groups. Yet, *in no case* has any convincing evidence emerged that suggests that the distribution of a child's subtest scores can be used reliably to diagnose anything other than that child's overall level of intellectual ability. Therefore, the major hypothesis guiding the present investigation was that:

> *Whether or not significant group level differences in WISC-R profiles are found, such differences will be of no diagnostic utility because any decision rules that they generate will result in unacceptably high percentages of false classifications.*

The specific versions of this hypothesis apply to Verbal-Performance IQ discrepancies, subtest scatter indices, recategorized scores, and the general form of profiles; and the relationships of these characteristics to each of the various educational categories.

CHAPTER 3

METHOD

Source of Data

The data used in this study were collected during the 1975-1976 school year by the Michigan Department of Education, Special Education Services. Many of the essential features of this data set have been described by Bernard (1978).

Sample

Forty-five of the 645 school districts in Michigan agreed to participate in the project. Each of these school districts provided data on all children seen by the Educational Placement and Planning Committee. The data form for each child was completed by a school psychologist and contained: the reason for the child's referral, WISC-R or other IQ test scores, grade placement, child's sex, child's age, any relevant special educational classification (i.e., diagnosis), as well as a number of scores on achievement tests.

Of the 1449 children for whom data were available, 323 were reevaluations. These children were excluded from the sample because it was not possible to determine from the data presented whether they had ever been classified in other categories or how long they had been in their present categories before testing. Another reason to exclude these children was to avoid any bias in classification due to previous placement.

Only 711 of the remaining children had been given the WISC-R. Eight of these children came from homes where English was not the primary language. The remaining 703 children constituted the sample for the analyses reported in this study.

Seventy-four percent of the children in the sample were male. They ranged from five to eighteen years of age, with a mean of 10.7 years. Their grade placements ranged from first to twelfth, with a median of fourth and a rather prominant mode at the first grade.

The majority (562) of the children had been referred because of academic problems. Behavior problems (126) and physical handicaps (10) accounted for the rest of the known reasons for referral. Reason for referral was not indicated for five children.

The children were assigned by the Educational Placement and Planning Committees to five diagnostic categories: learning disabled (234), educable mentally impaired (77), emotionally impaired (64), physically or otherwise handicapped (38), and those children who were not eligible for placement (290). The use of the not placed children as a normal control group is reasonable in that they represent the class of children

For the purpose of the present investigation, any rule that leads to a correct decision 85 percent of the time will be considered useful. That is, if more than 15 percent of the cases analyzed are placed in incorrect categories, the decision rule is not useful. Any discomfort generated by this arbitrary 85 percent requirement should be mitigated by the fact that the exact correct decision rates are provided with each analysis. Thus, anyone who wishes to apply a different acceptance rule may do so easily.

Hypotheses

The application of the term "useful" in this section follows from the discussion in the previous section. That is, a decision rule is useful if it is correct for 85 percent of the cases to which it is applied.

With this in mind, the hypotheses for the present study were:

1. Simple comparisons of Verbal and Performance IQ scores will not lead to a useful decision rule governing the assessment of learning disabiliies.
2. WISC-R scatter indices will not provide decision rules that are useful in discriminating learning disabled children from normal children.
3. The recategorizations of WISC-R subtests that have been suggested in the literature will lead to no useful learning disabled versus normal discrimination.

4. No linear combination of WISC-R subtests will yield a useful learning disability diagnosis rule.
5. WISC-R subtests will not be useful at any grade level in separating learning disabled children into either age or Full Scale IQ groups.
6. Simple comparisons of Verbal and Performance IQ scores will not lead to a useful decision rule governing the assessment of emotional impairment.
7. The high-Similarities, low-Information pattern of signs suggested by Dean will not yield a useful emotional impairment diagnosis rule.
8. No linear combination of WISC-R subtest scores will provide a decision rule that is useful in separating any of the groups studied from the other groups.

CHAPTER 4

RESULTS

Overview

The primary focus of this research was to analyze the value of WISC-R patterns for identifying children as either normal, learning disabled, educable mentally impaired, or emotionally impaired.

In the first part of this chapter, the utility of each of the specific models proposed for the identification of learning disabilities is examined. Of central importance to the questions to be addressed is the distinction between statistical and practical significance. While a particular theoretical formulation may provide statistically reliable differentiation between large groups of individuals, such a finding may be of little practical value to those charged with making educational judgments if the decision rule that it generates results in an unacceptably high number of placements of children in incorrect diagnostic categories.

Following the analysis of particular hypothesized WISC-R patterns among the learning disabled is a more general exploration of the data to determine whether the WISC-R subscale scores provide *any* basis for a meaningful distinction between learning disabled and normal children.

Another section of the chapter focuses on children who have been categorized as learning disabled. The WISC-R profiles of this group are examined for differences that are related to either age or full-scale IQ.

In the next section, the emotionally impaired children are compared to the normal children to determine whether either Verbal-Performance discrepancies or Dean's (1977b) suggestions about the extremity of the Information and Similarities subscale scores provide any valid separation between these groups.

In the next set of analyses, the WISC-R profiles of all five groups of children are examined for characteristic patterns. The analysis techniques in this section include discriminant function analysis and a form of profile distance analysis.

The final section of the chapter contains an attempt to identify a small set of subtest scores that best predict full-scale intelligence for each of three groups -- the normal sample, the learning disabled sample as a whole, and those learning disabled children who are in the eighth grade or higher.

<u>Learning Disabilities</u>: <u>Verbal-Performance Discrepancy</u>

As discussed in a previous chapter, researchers have postulated a difference between the WISC-R Verbal and Performance scales for learning disabled children. As Kaufman (1979) has noted, a difference of 12 points between these scores is required for significance at the .05 level. Table 1 contains a crosstabulation of normal and learning disabled children by whether or not they had a verbal-performance difference of 12 or more points, regardless of the direction of that difference. The chi-square statistic for this table ($\chi^2 = 11.14$, $df = 1$, $p < .001$) highlights the fact that a disproportionate number of learning disabled children have a significant VP difference. However, a more detailed examination of this theoretically interesting result reveals that 30 percent of the normal children also show the "characteristic" learning disabled pattern and that 53 percent of the learning disabled group would have to be be classified as normal if a 12-point VP discrepancy were the sole criterion for the classification of learning disability. In other words, a rule based on a 12-point VP discrepancy yields 30 percent false positives and 53 percent false negatives when applied to these data. Indeed, the phi coeficient of .17 associated with

Table 1. Verbal Performance Discrepancies for Learning Disabled and Normal (Not Placed) Children - (Absolute Difference)

	Norm(not placed)	Learning Disabled
No Significant Difference	156 (70%)	93 (53%)
Twelve Point Difference	68 (30%)	83 (47%)

Table 1 means that these two variables (diagnostic group and dichotomized absolute VP difference) have a mere three percent of their variance in common. The point biserial correlation between group (Normal, LD) and the full-range absolute VP difference, a more generous measure of the association between these variables, is only .19. The four percent shared variance that this represents is hardly a basis for sound decisions.

When the direction of the V-P difference is considered, the statistical analysis again confirms theoretical expectations ($X^2 = 15.40$, $df = 1$, $p < .001$). In Table 2, the learning disabled children are over-represented in the group with a significant (12-point) Performance greater than Verbal discrepancy. However, closer inspection of this table does not lead to any convincing decision rule about the

Table 2. Verbal Performance Discrepancies for Learning Disabled and Normal (Not Placed) Children - (Directional)

	Norm(not placed)	Learning Disabled
$P \geq (V+12)$	53 (24%)	53 (30%)
No Significant Difference	156 (70%)	93 (53%)
$V \geq (P+12)$	15 (7%)	30 (17%)

classification of learning disabilities. Indeed, the learning disabled group is also over-represented at the theoretically unexpected (V>P) end.

One important theoretical prediction was not supported. Collapsing the bottom two rows in Table 2, to test whether learning disabled children were over-represented among those with Performance at least twelve points higher than Verbal resulted in $\chi^2 = 1.79$, $df = 1$, n.s. Thus, using a 12-point performance greater than Verbal rule yields 24 percent false positives and 70 percent false negatives.

These Verbal-Performance discrepancy analyses offer broad support for hypothesis 1. The twelve point difference in either direction rule resulted in incorrect classification for 40 percent of cases. The Performance twelve or more points greater than Verbal rule was correct only 56 percent of the time. Neither of these met the 85 percent correct requirement.

Learning Disabilities: Subtest Scatter

Another proposition presented earlier concerned the amount of scatter in the subtest scores of the learning disabled sample when compared to other groups. In addition to Kaufman's Range and NDEV measures, the variance of each child's subtest scores was used as a scatter index. This, within person variance, measure was computed as the simple variance of the child's subtest scores about the mean of that child's scores (i.e., the second moment of the distribution of the child's subtest scores). Table 3 contains the group means for the Range (highest minus lowest subtest score), number of deviations from an individual's own mean (NDEV), and the within person variance across the

Table 3. Group Means for Three Measures of Scatter

	Range	NDEV	Within-person Full-IQ Variance
Normal	6.78 a	1.56 a	5.04 a
Learning disabled	7.53 b	2.15 b	6.27 b
Educable Mentally Impaired	6.36 a	1.36 a	4.62 a
Emotionally Impaired	7.31 ab	1.80 ab	5.49 ab
Physically or Otherwise Handicapped	6.86 ab	1.67 ab	5.30 ab

Within each column of this table, means that do not share a common subscript differ from each other, $p < .05$.

full set of subtest scores. Analyses of variance on these data revealed significant group related differences on Range (F [4,533] = 4.523, $p <$ .002), number of deviations from mean (F [4,533] = 6.04, $p < .0001$), and within person subtest variance (F [4,533] = 5.44, $p < .0003$).

The most casual inspection of Table 3 suggests that across all three indices of scatter, the learning disabled group exhibited the most scatter. Newman-Keuls tests -- indicated by subscripts in the table -- support this observation. It is important to note that these three indices of scatter are not mathematically independent. However, in these data, it is safe to conclude that the WISC-R subtests of the learning disabled children are somewhat more variant than those of the other groups.

Table 4. Summary of Use of Scatter Measures to Classify Children As Either Normal or Learning Disabled

Measure of Scatter	Correct Classif. Norm.	Correct Classif. LD	Incorrect* Classif. Norm.	Incorrect* Classif. LD	Percent False Positive	Percent False Negative	Percent Incorrect Overall
Range	148	84	77	94	34	53	42
NDEV	115	117	110	61	49	34	42
Within-Pers. Variance	146	94	79	84	35	47	40

*The children in these columns are labeled by actual rather than "incorrect" group. For example, 77 normal children were closer to the LD mean than to the normal mean on Range.

Might this group level difference in scatter be useful in classifying learning disabled children? To answer this question, each of the three scatter scores (Range, NDEV, and Variance) of each of the normal and learning disabled children in the sample was checked to see if it was closer to its respective normal or learning disabled mean. The results of this analysis may be found in Table 4, where, among other things, false positives and false negatives are noted. In summary of this analysis, the practitioner who can afford to be wrong 40 percent or more of the time may use one of these scatter indices for classification. One who desires more precision will have to look elsewhere. The right-hand column in Table 4 may be used to evaluate hypothesis 2. This hypothesis was supported; in no case was a decision rule based on one of the scatter indices successful in more than 60 percent of applications.

Learning Disabilities: Bannatyne/Kaufman Recategorizations

While several patterns of WISC-R subscale scores have been thought to distinguish children with learning disabilities, Kaufman's (1975) recategorizations have received the most extensive acceptance among practitioners. Kaufman's basic formulation involves the computation of three scores: (1) Spatial (SP), the mean of the Picture Completion, Block Design, and Object Assembly subtests; (2) Verbal Comprehension (VC), the mean of Comprehension, Similarities, and Vocabulary; and (3) Attention Concentration (AC), the mean of Arithmetic, Coding, and Digit Span. The data available for this study do not include the optional Digit Span subtest scores. In this situation, the recommended common practice it to compute AC as the mean of Arithmetic and Coding (Kaufman, 1979).

Kaufman and others (Bannatyne, 1974; Smith *et al.*, 1977a, 1977b) have suggested that the recategorized scores of learning disabled children typically are arrayed in a pattern that distinguishes them from their normal counterparts. The occurrence of this particular pattern (SP>VC>AC) is the focus of a later analysis. In its most simple form, the differential pattern hypothesis implies the presence of a group by recategorized score interaction in an analysis of variance. Table 5 contains the results of a 2 (Normal versus LD) by 3 (SP, VC, AC) analysis of variance with repeated measures on the three-level factor. While the significant main effect for test may be of some interest, it has no bearing on the question that motivated the analysis. The absence of the predicted group by test interaction presents a problem for advocates of the differential pattern hypothesis.

Table 5. Analysis of Variance - Diagnostic Group by Recategorized Test Scores

Source of Variation	df	Mean Square	F Ratio	p<
Group (G)	1	30.31	3.02	n.s.
Subjects w/in G (error)	500	10.03		
Test (T)	2	96.81	32.87	.001
G X T	2	5.92	2.01	n.s.
Subjects by T w/in G (error)	1000	2.95		

Means

	SP	VC	AC	Group Marginals
Norm(not placed)	9.73	8.81	9.10	9.21
Learning Disabled	9.42	8.76	8.61	8.93
Test Marginals	9.59	8.79	8.89	

Grand Mean = 9.09

Another, more practically interesting way of addressing the question of the relationship of the recategorized scores to the classification of learning disabled children is to use those scores in a discriminant function analysis of group membership (Normal, LD). This analysis was performed using the three recategorized scores after subtracting from each of them the child's own mean across the three scores. This forces profile scatter rather than profile elevation to be used for discrimination. The elevation of a profile is, simply, the mean score of the profile. The scatter of a profile is its variance.

The function derived from this analysis was not significant. Fifty-six percent of the cases used in this analysis were from the normal group. The discriminant function correctly classified 58 percent of the cases. That represents a two percent improvement over chance using the best possible linear function of the recategorized scores. This failure to provide correct classification for at least 85 percent of the cases supports hypothesis 3.

The next attempt to gauge the utility of the SP-VC-AC recategorization was less statistically sophisticated but may be more practically informative. SP, VC, and AC were computed for each of the normal and learning disabled children. The three pairwise differences between these scores were evaluated. Two scores were considered to be different if they differed by three or more points. This is consistent with the clinical use of this type of data (Kaufman, 1979). Of the 27 relationships among the scores, eight are excluded logically. Table 6 contains the remaining 19 patterns along with the number and percent of children in each group whose scores exhibited each pattern. Three things are noteworthy in Table 6. Most normal and learning disabled children show no significant difference in the recategorized scores (row 1). No children in either group exhibit the perfect case of the Bannatyne learning disability pattern (row 10). For the most part, equal proportions of each group are represented in each of the rows of the table. In brief, the children in the normal and learning disabled groups seem to have roughly similar SP-VC-AC patterns. There is no evidence in this table that would lead to anything close to 85 percent correct

Table 6. Children with Particular SP-VC-AC Pattern (3 Point Difference)

				Norm (not placed)		Learning Disabled	
				N	Percent	N	Percent
1.	SP=VC	VC=AC	SP=AC	152	(54)	110	(50)
2.	SP=VC	VC=AC	SP>AC	13	(5)	7	(3)
3.	SP=VC	VC=AC	SP<AC	6	(2)	2	(1)
4.	SP=VC	VC>AC	SP=AC	8	(3)	14	(6)
5.	SP=VC	VC>AC	SP>AC	16	(6)	13	(6)
6.	SP=VC	VC<AC	SP=AC	13	(5)	9	(4)
7.	SP=VC	VC<AC	SP<AC	10	(4)	7	(3)
8.	SP>VC	VC=AC	SP=AC	16	(6)	12	(5)
9.	SP>VC	VC=AC	SP>AC	16	(6)	12	(5)
10.	SP>VC	VC>AC	SP>AC	0	(0)	0	(0)
11.	SP>VC	VC<AC	SP=AC	17	(6)	14	(6)
12.	SP>VC	VC<AC	SP>AC	0	(0)	1	(0)
13.	SP>VC	VC<AC	SP<AC	1	(0)	0	(0)
14.	SP<VC	VC=AC	SP=AC	3	(1)	7	(3)
15.	SP<VC	VC=AC	SP<AC	5	(2)	1	(0)
16.	SP<VC	VC>AC	SP=AC	4	(1)	13	(6)
17.	SP<VC	VC>AC	SP>AC	0	(0)	0	(0)
18.	SP<VC	VC>AC	SP<AC	0	(0)	0	(0)
19.	SP<VC	VC<AC	SP<AC	0	(0)	0	(0)

Table 7. Children with Particular SP-VC-AC Pattern (Any Difference)

				Norm (not placed)		Learning Disabled	
				N	Percent	N	Percent
1.	SP=VC	VC=AC	SP=AC	0	(0)	0	(0)
2.	SP=VC	VC=AC	SP>AC	0	(0)	0	(0)
3.	SP=VC	VC=AC	SP<AC	0	(0)	0	(0)
4.	SP=VC	VC>AC	SP=AC	0	(0)	0	(0)
5.	SP=VC	VC>AC	SP>AC	4	(1)	4	(2)
6.	SP=VC	VC<AC	SP=AC	0	(0)	0	(0)
7.	SP=VC	VC<AC	SP<AC	11	(4)	3	(1)
8.	SP>VC	VC=AC	SP=AC	0	(0)	0	(0)
9.	SP>VC	VC=AC	SP>AC	7	(2)	5	(2)
10.	SP>VC	VC>AC	SP>AC	55	(20)	54	(24)
11.	SP>VC	VC<AC	SP=AC	7	(2)	5	(2)
12.	SP>VC	VC<AC	SP>AC	59	(21)	35	(16)
13.	SP>VC	VC<AC	SP<AC	49	(18)	40	(18)
14.	SP<VC	VC=AC	SP=AC	0	(0)	0	(0)
15.	SP<VC	VC=AC	SP<AC	3	(1)	0	(0)
16.	SP<VC	VC>AC	SP=AC	3	(1)	3	(1)
17.	SP<VC	VC>AC	SP>AC	39	(14)	38	(17)
18.	SP<VC	VC>AC	SP<AC	20	(7)	20	(9)
19.	SP<VC	VC<AC	SP<AC	23	(8)	15	(7)

classifications.

Knowledgeable researchers in this area might debate the choice of the three-point difference in SP-VC-AC scores that was used to construct Table 6. In fact, several versions of this table were constructed with 4, 3, 2, 1.5, and 1-point differences required for inequality. These tables are not reproduced here (see Appendix) as their only interesting feature is the regularity with which the group proportions mirror each other in the different rows of the tables. However, one extreme version of this table is worthy of inspection. Table 7 was constructed by allowing *any* difference between two scores to count as an inequality. In Table 7, row 10, the classic Bannatyne pattern accounts for 24 percent of the learning disabled children. However 20 percent of the normal children also exhibited this pattern. Attention to a learning disabled sample alone could easily lead a researcher to a false conclusion about these data.

Kaufman's more recent suggestion of a depressed Arithmetic, Coding, Information, and Digit Span composite among the learning disabled was evaluated by computing the mean of the first three of these scores (ACI) for each of the normal and learning disabled children in the sample (Digit Span scores not available) and the mean of the remaining seven scores (REM7). Table 8 contains the counts of children in each of these groups whose ACI scores were at least three points lower than the mean of the remaining scores. The X^2 (chi-square) for these data was 2.28, df = 1, n.s. The point biserial correlation between group (Normal, LD) and REM7 minus ACI was .10, $p < .05$, showing a moderate tendency on the part of the learning disabled children to have depressed ACI scores. However

Table 8. Relationship of ACI Scores to Mean of Remaining Subtest Scores
(Table entries are *N*s)

	Norm (not placed)	Learning Disabled
*ACI not lower	201	149
*ACI lower	24	29

*3-Point difference required

as the data in Table 8 show, this relationship is of questionable practical value, as the rule that it generates leads to an incorrect classification for 43 percent of the cases (11 percent false positives and 84 percent false negatives). Therefore, under the 85 percent correct criterion established for this study, the low ACI rule is not useful. Hence, there is more support for hypothesis 3.

<u>Discrimination of Learning Disabilities Using WISC-R Subtests</u>

At this point, it is worth asking whether the WISC-R subtest scores are of *any* use in discriminating between normal and learning disabled children. To address this question, a discriminant function analysis was performed using the ten subtest scores, each deviated from the individual's own mean across the subtests.

The function obtained was significant ($X^2 = 18.035$, $df = 9$, $p < .05$). However, the function itself is not of interest at this point. What is critical is its ability to classify individuals appropriately. It is important to note that 56 percent of the children in this analysis

Table 9. Classification of Cases Using Function Derived from Ten Mean-deviated Subtest Scores

	Predicted Group	
Actual Group	Norm(not placed)	Learning Disabled
Norm(not placed)	178 *	47
Learning Disabled	109	69 *

* Indicates correct classification.

were from the normal group; hence, by chance any prediction will be correct 56 percent of the time. Table 9 contains the classification that resulted from using this function. Seventy-nine percent of the normals were correctly classified as such. Only 39 percent of the learning disabled were so classified. The overall correct classification was 61.3 percent. Therefore, hypothesis 4 was supported. The off-diagonal elements in Table 9 are unacceptably high. *Using all of the information available in the mean-deviated subtest scores 21 percent of the normal children were classified as learning disabled and 61 percent of the learning disabled children were classified as normal.*

The implications of these data for the use of WISC-R subtest scores in the diagnosis of learning disabilities are discussed in the next chapter.

Learning Disabilities: Subgroups

A discriminant function analysis was performed to explore the possibility that learning disabled children with higher IQ's have WISC-R subtest profiles that are different from those of learning disabled children with lower IQ's. For this analysis, Higher IQ was defined as having Full Scale, Verbal, and Performance IQ's all equal to or greater than 90. Sixty-four percent of the learning disabled children in these data fell into the Low IQ group. Mean-deviated subtest scores were used in this analysis. The discriminant function was not significant (X^2 = 14.1, *df* = 9, n.s.). The classification results are presented in Table 10. Sixty-seven percent of the cases in this analysis were classified correctly. As 64 percent represents chance classification, the results in Table 10 do not show any practically meaningful discrimination between the Low- and High-IQ learning disabled groups. This provides support for hypothesis 5.

Table 10. Classification of High- and Low-IQ Learning Disabled Children Using Mean-deviated Subtest Scores

	Predicted Group	
Actual Group	Low IQ	High IQ
Low IQ	98 *	14
High IQ	44	20 *

* Correct classification.

Table 11. Numbers of High- and Low-IQ Learning Disabled Children with Significant (12-Point) V-P Discrepancy

	High IQ	Low IQ
Significant VP discrepancy	23	60
No Discrepancy	41	52

$X^2 = 4.4$, $df = 1$, $p < .05$

Table 12. Analysis of Variance of Acquired Knowledge Scores by Group and Grade Level

Source of Variation	df	Mean Square	F Ratio	p<
Group (G)	1	6.032	1.413	n.s.
Grade Level (GL)	1	35.768	8.378	.004
G x GL	1	.113	.027	n.s.
Error	459	4.270		

Means

Group

Grade Level	Norm (not placed)	Learning Disabled	Grade Level Marginals
7 and below	8.57	8.35	8.47
8 and above	7.83	7.53	7.71
Group Marginals	8.45	8.24	

Grand Mean = 8.35

Table 13. Analysis of Variance of Verbal Scale IQ Scores by Group and Grade Level

Source of Variation	df	Mean Square	F Ratio	p<
Group (G)	1	809.357	5.071	.025
Grade Level (GL)	1	1959.097	12.270	.001
G x GL	1	301.862	1.891	n.s.
Error	509	159.617		

Means

	Group		
Grade Level	Norm (not placed)	Learning Disabled	Grade Level Marginals
7 and below	91.79	88.35	91.22
8 and above	89.95	82.35	88.71
Group Marginals	90.98	85.74	

Grand Mean = 90.11

For the next analysis, those learning disabled children with a significant (i.e., 12-point) Verbal-Performance discrepancy were identified. Table 11 contains a breakdown of these children by High versus Low IQ. As is evident, more of the Low IQ learning disabled children had significant V-P discrepancies.

The hypothesis that older learning disabled children are likely to have depressed Acquired Knowledge scores was evaluated using a two-way analysis of variance. Acquired Knowledge was computed as the mean of the Information, Arithmetic, and Vocabulary subtests. The sample was divided

Table 14. Analysis of Variance of Performance Scale IQ Scores by Group and Grade Level

Source of Variation	df	Mean Square	F Ratio	p<
Group (G)	1	819.357	4.907	.027
Grade Level (GL)	1	225.502	1.350	n.s.
G x GL	1	1.644	.010	n.s.
Error	509	166.990		

Means

Group

Grade Level	Norm (not placed)	Learning Disabled	Grade Level Marginals
7 and below	96.13	94.21	95.80
8 and above	93.53	93.21	93.27
Group Marginals	94.98	93.21	

Grand Mean = 94.69

into age groups of seventh grade or lower and eighth grade or higher. The factors in this analysis were group (LD versus normal) by grade level (as described). Table 12 contains these results. While it is true that older learning disabled children had lower Acquired Knowledge scores, the same is true for the older normal children. This is indicated by the significant grade level main effect in Table 12. Neither the group main effect nor the group by grade level interaction was significant.

Table 15. Analysis of Variance of Full Scale IQ Scores by Group and Grade Level

Source of Variation	df	Mean Square	F Ratio	p<
Group (G)	1	782.612	5.608	.018
Grade Level (GL)	1	903.523	6.475	.011
G x GL	1	60.162	.431	n.s.
Error	509	139.548		

Means

	Group		
Grade Level	Norm(not placed)	Learning Disabled	Grade Level Marginals
7 and below	93.18	90.42	92.71
8 and above	90.99	86.38	90.24
Group Marginals	92.21	88.66	

Grand Mean = 91.62

The Full Scale, Verbal, and Performance IQ scores of the older and younger normal and learning disabled children were subjected to the same two-way analysis of variance. The results of these three analyses are presented in Tables 13, 14, and 15.

Across these analyses, the significant Group effect indicates that on Verbal IQ, Performance IQ, and Full Scale IQ, the scores of the normal children were slightly higher than those of the learning disabled children. This is not an unexpected result. It should be noted that all of the group means were within the normal range. For both Verbal IQ and

Full Scale IQ the scores of the younger children were higher than those of the older children. In none of these three analyses was the two way -- Group by Grade Level -- interaction present. Hence, it cannot be concluded that depressed IQ scores are specifically associated with older learning disabled children.

Emotional Impairment: Verbal Performance Discrepancy

As noted earlier, Dean and others have hypothesized that emotionally impaired children are more likely than their non-impaired counterparts to exhibit a high discrepancy between their Verbal and Performance IQ scores. According to Kaufman (1979), a difference of 12 points is required for these two scores to be considered different at the .05

Table 16. Verbal-Performance Discrepancies for Emotionally Impaired and Normal Children

	Group	
	Norm (not placed)	Emotionally Impaired
$P \geq (V + 12)$	53 (23.7%)	10 (20.4%)
No significant VP difference	156 (69.6%)	33 (67.3%)
$V \geq (P + 12)$	15 (6.7%)	6 (12.2%)

level. Table 16 contains a breakdown of the normal and emotionally impaired (EI) samples by extent of Verbal-Performance discrepancy. The χ^2 (chi-square) of 1.8239 (*df* = 2, n.s.) associated with this table does not support this expectation.

Relatively few children from either the EI or the normal group had elevated Verbal scores. In addition, there was no marked tendency for either group to be over-represented among those with elevated Performance scores. The rule of any twelve point difference was correct for only 63 percent of these children. The Performance at least twelve points greater than Verbal rule was successful only 66 percent of the time. Thus, hypothesis 6 was supported.

<u>Emotional Impairment</u>: <u>High Similarities, Low Information Pattern</u>

As discussed in an earlier section, Dean (1976) has noted two features or signs that occur in the verbal subtest profiles of emotionally impaired (EI) children. Dean reported that, within this group, Information typically is one of the two lowest verbal subtest scores and Similarities is one of the two highest.

Dean's suggestion has at least two implications. First, if his hypothesis is correct, one would expect that a large percentage of EI children would share this pattern. Second, in order for this pattern to be "characteristic" of the subtest profiles of EI children, it would have to be demonstrated that the pattern occurs more frequently among the emotionally impaired than it does in some appropriate comparison group.

Dean's (1976) analysis does not provide much guidance to the researcher who attempts to evaluate these patterns. Exactly what is meant by "highest" or "lowest"? Should the conventional three-point rule be used to decide whether a score is different from the others? Or, does the completely flat profile conform to Dean's pattern? That is, when all scores are the same, are they not each "highest" and "lowest"? Thus, to provide a fair test of Dean's hypothesis, two analyses were accomplished.

In the first analysis, a sign was present for a child if not more than one other subtest was more extreme than the subtest of interest. That is, Information was low if not more than one Verbal subtest was lower than Information, and Similarities was high if not more than one Verbal subtest was higher than Similarities. While this counts the

Table 17. Dean (1976) Signs for Emotionally Impaired and Normal Children (Dean's Criteria)

	Norm (not placed)	Emotionally Impaired
Both Signs	76 (31%)	19 (36%)
Information Sign	63 (26%)	14 (26%)
Similarities Sign	36 (15%)	6 (11%)
Neither Sign	71 (29%)	14 (26%)

$\chi^2 = .78$, $df = 3$, n.s.

absurd completely flat profile as an occurrence of Dean's pattern, a careful reading of Dean's (1976) paper made it clear that this was the rule that he employed. For this analysis, then, each child could have one of four outcomes -- both signs might be present, only the Information sign might be present, only the Similarities sign might be present, or neither sign might be present. Table 17 contains the results of this analysis.

It is evident in these data that the proposed signs do not provide any meaningful distinction between these two groups.

Perhaps the data in Table 17 do not provide a fair test of Dean's proposal. It could be the case that the liberal criterion used for the occurrence of the signs resulted in an artificially high number of signs among the normal sample. Maybe a more strict criterion would result in greater attenuation of the signs for the normal children than for the emotionally impaired children.

The second analysis of Dean's proposal employed a more strict requirement for the occurrence of the signs. In this analysis, a sign was present in a child's profile if not more than one Verbal subtest score was more extreme than the score of interest, and the score of interest was at least three points different from the mean of the remaining four Verbal subtest scores. That is, the Information sign was present if not more than one Verbal score was lower than Information, and Information was at least three points lower than the mean of Comprehension, Arithmetic, Similarities, and Vocabulary; the Similarities sign was present if not more than one other Verbal test was higher than Similarities, and the Similarities score was at least three points higher

Table 18. Dean (1976) Signs for Emotionally Impaired and Normal Children (Strict Criteria)

	Norm (not placed)	Emotionally Impaired
Both Signs	3 (1%)	0 (0%)
Information Sign	26 (11%)	3 (6%)
Similarities Sign	10 (4%)	5 (9%)
Neither Sign	207 (84%)	45 (85%)

$\chi^2 = 4.24$, $df = 3$, n.s.

than the mean of Information, Arithmetic, Vocabulary, and Comprehension. Table 18 contains the results of this analysis.

Although the chi-square test on this table is inappropriate due to the low cell frequencies, the results of the analysis are clear. The two groups do not differ.

In summary, neither of these analyses supported Dean's hypothesis regarding the Verbal subtest profiles of emotionally impaired children. In no case did a rule emerge from the data that would provide correct classification for even 50 percent of these children. Thus, hypothesis 7 was supported.

WISC-R Subtests and Diagnostic Categories

The cases available for this study included data from children who have been classified as learning disabled (LD), educable mentally impaired (EMI), and physically or otherwise handicapped (Other), as well as a normal sample -- those children who were referred for testing who were not placed in any of the other categories. As might be expected, the EMI children are characterized by an overall depression in their subtest scores. Other than that, it would be useful to determine whether any of these groups can be discriminated using WISC-R subtest scores. To eliminate differences in test elevation (mean), mean-deviated subtest scores (i.e., subtest scores minus the child's own subtest mean) will be used in these analyses.

Table 19. Classification of Five Groups Using Mean-deviated Subtest Scores

	Predicted Group				
Actual Group	Normal	LD	EMI	EI	Phys. H'cap.
Normal	176 *	49	0	0	0
LD	103	75 *	0	0	0
EMI	44	5	0 *	0	0
EI	37	11	0	1 *	0
Phys. H'cap.	31	3	0	2	0*

* Correct classification

47 % of all cases classified correctly.

A discriminant function using the mean-deviated subtest scores of the children from each group for whom all subtest data were present yielded one significant function ($X^2 = 58.34$, $df = 36$, $p < .02$). Once again, while the particular form of this function may be of some theoretical interest, the primary concern in the present research is the success of classification of children using this function. Table 19 contains the classification results. As 42 percent of the children in this analysis were from the normal group (i.e., the largest group), that percentage represents chance level classification. In other words, in the absence of any subtest information, the differences in group n's would allow for 42 percent correct classification simply by predicting "normal" for all cases.

The 47 percent correct classification in this analysis fell far short of the 85 percent requirement. No useful decision rule was generated. Hypothesis 8 was supported. It might be argued that the over-representation of normal and learning disabled children in this analysis invalidates these classification results. However, for two reasons, this is not true. First, the proportions of these categories of children in this analysis represent accurately those proportions in the clinical practice of school psychologists throughout Michigan. Second, if the subtest profile of any group in this analysis were distinguishable from those of the others, the statistical algorithm chosen for the analysis would have been sensitive to that fact, classification on that group's centroid would have been more precise, and the results in Table 19 would have been different. As these results stand, the five percent improvement on chance and the 53 percent error rate do not provide a reasonable basis for differential diagnosis using WISC-R subtest scores.

Another interesting way to assess profile distances between individuals was suggested by Cronbach and Gleser (1953). They proposed three measures of the distance between individuals on a set of scores. The first of these, *D*, is defined as the square root of the average squared difference between the individuals across the set of scores. That is, the first score for individual 1 is subtracted from the first score for individual 2; the result is squared and added to the squared differences between these individuals for the succeeding scores; this sum is then divided by the number of differences (scores) contributing to it; the square root of this quotient is *D*. Cronbach and Gleser's second distance measure, *D*', is a variant of *D* in which the individuals' own means across all scores are subtracted from each of the scores before *D* is computed as above. Hence, *D*' is *D* with elevation removed but scatter remaining. The third measure proposed by Cronbach and Gleser, *D*'', is a variant of *D* and *D*' from which both elevation and scatter have been removed, leaving only shape.

The particular concerns of this research -- i.e., group related differences in WISC-R subtest *scatter* -- led to the choice of *D*' as the appropriate measure of the distance between individuals. *D*' was computed for every pair of individuals, regardless of group identification, using only those children with no missing subtest scores.

For computational purposes:

$$D' = \{\sum_{i=1}^{k}[(X_{ij} - \bar{X}_j) - (X_{im} - \bar{X}_m)]^2/k\}^{1/2}$$

where: j and m are any two individuals

x_1, x_2, ..., x_k are the subtest scores,

k = 10 (subtests)

Table 20 contains average distances (*D*') between individuals within each group and the average distances between each of the individuals in each of the groups and individuals in each of the other groups. For

Table 20. Average Within- and Between-group Profile Distances (*D*')

	Normal	LD	EMI	EI	Phys. H'cap.
Normal	2.77 (25200)				
LD	2.93 (40050)	3.06 (15753)			
EMI	2.78 (11025)	2.96 (8722)	2.77 (1176)		
EI	2.85 (11025)	2.99 (8722)	2.84 (2401)	2.89 (1176)	
Phys. H'cap.	2.79 (8100)	2.97 (6408)	2.79 (1764)	2.87 (1764)	2.82 (630)
Group N	225	178	49	49	36

Parenthesized values are *N*s contributing to each comparison.

example, the first diagonal element in this table is the average distance, as measured by *D'*, between individuals in the normal group. The element below that is the average distance between individuals in the normal group and those in the learning disabled group. The subtraction of the individual means does not change the metric of these distance measures. The *D'* values in Table 20 are in the original metric of the WISC-R subtest scores.

There are two reasons why statistical tests on elements of Table 20 are somewhat meaningless. First, each individual contributed a number of times to the average difference in each of the cells. Second, the number of comparisons contributing to each mean -- shown in parentheses under each of the elements of the table -- is so large, and ,therefore, the standard error so small, that virtually any difference is statistically significant. There are, however, some noteworthy observations to be gleaned from this analysis. For example, the average distance between children in the learning disabled sample is greater than (or, at least, equal to) the average distance between learning disabled children and children in the normal sample. In other words, we can conclude that, *in terms of profile distances, these learning disabled children are no more similar to each other than they are to the normal children.* One summary statement that can be made about Table 20 is that the within-group profile distances are at least as great as the between-group profile distances. In short, there seems to be no reliable information in the mean-deviated individual subtest scores that can be used to distinguish between groups.

FIGURE 1. MEAN SUBTEST SCORES FOR DIAGNOSTIC GROUPS

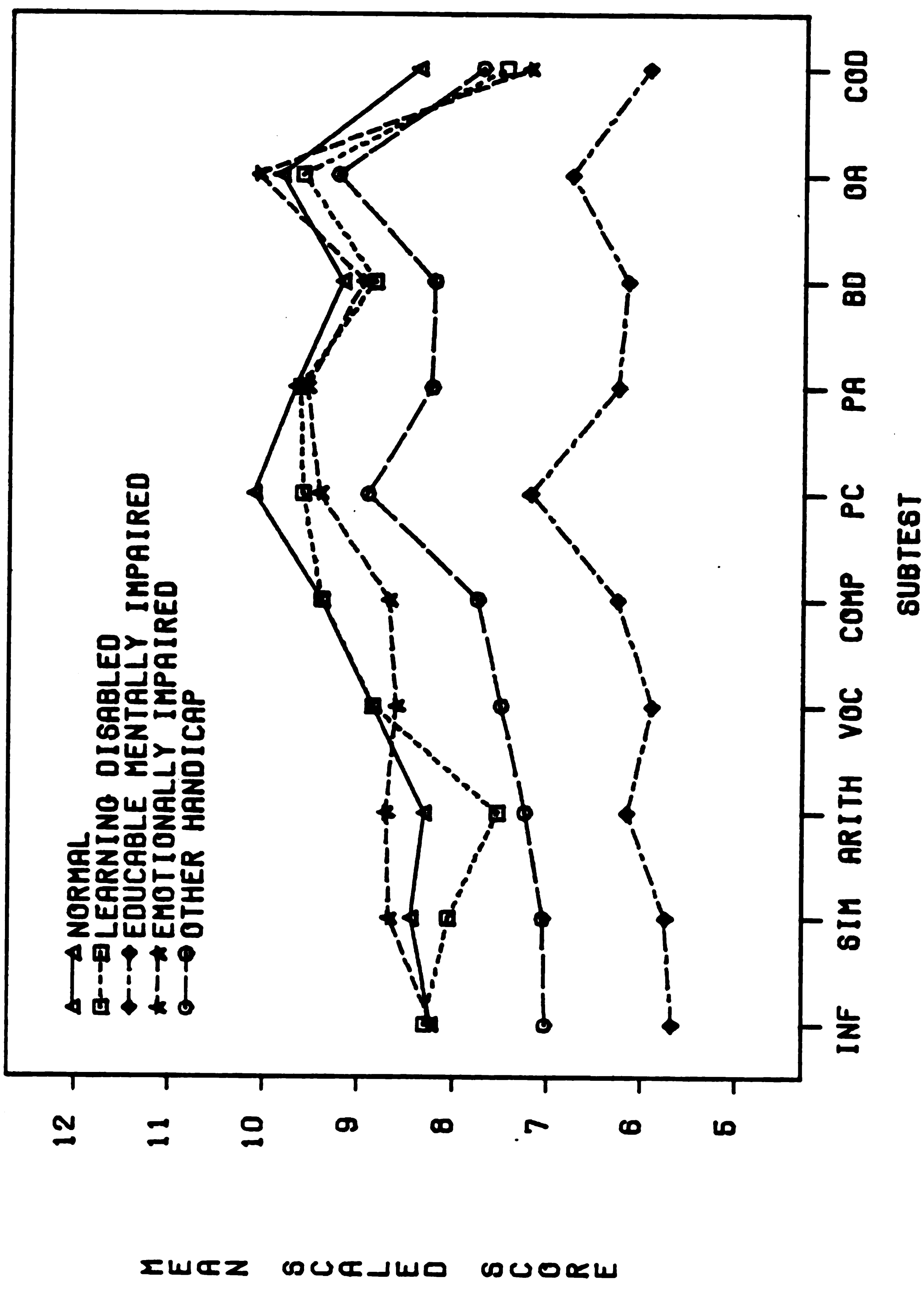


FIGURE 2. MEAN MEAN-DEVIATED SUBTEST SCORES FOR DIAGNOSTIC GROUPS

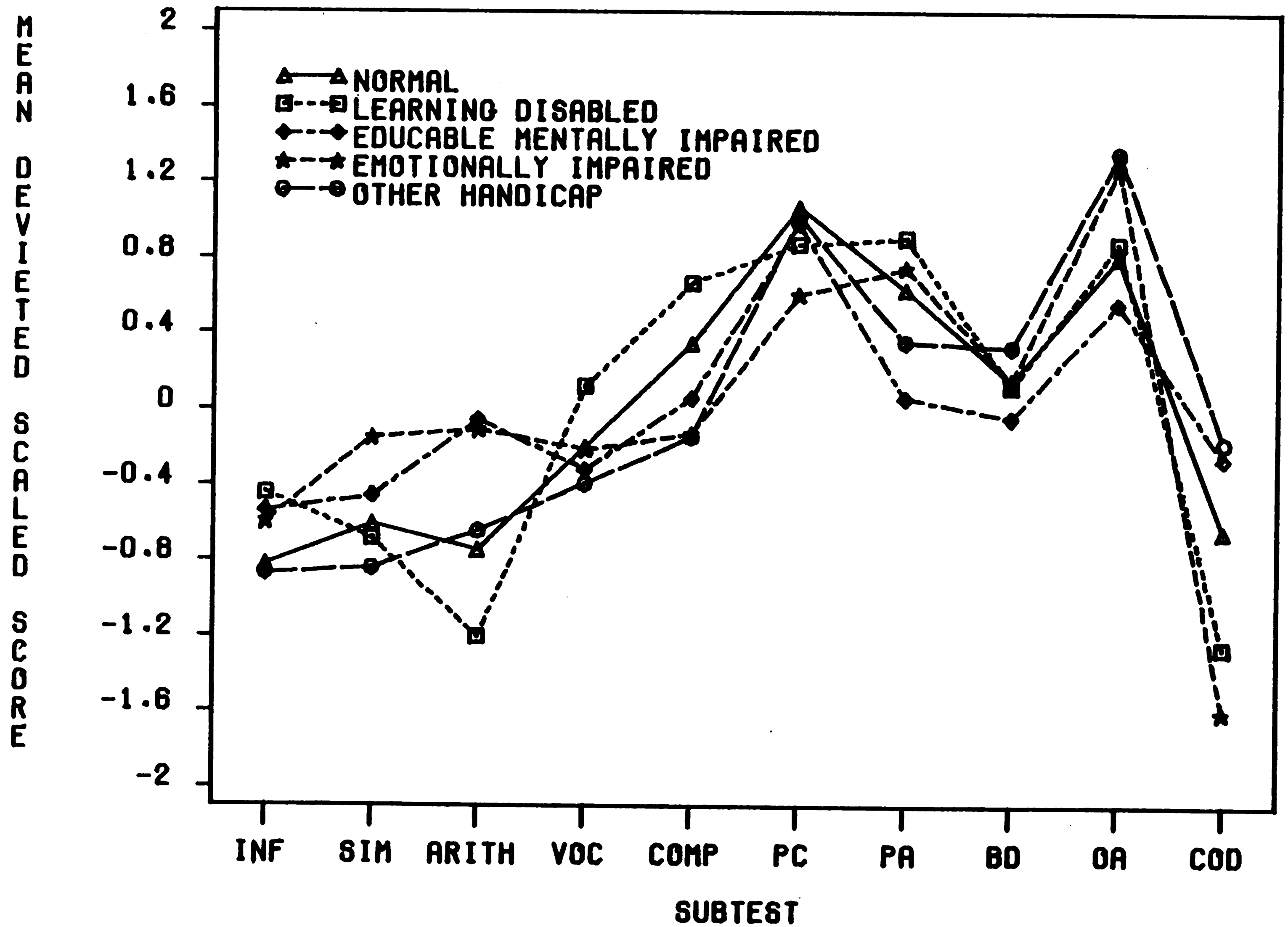
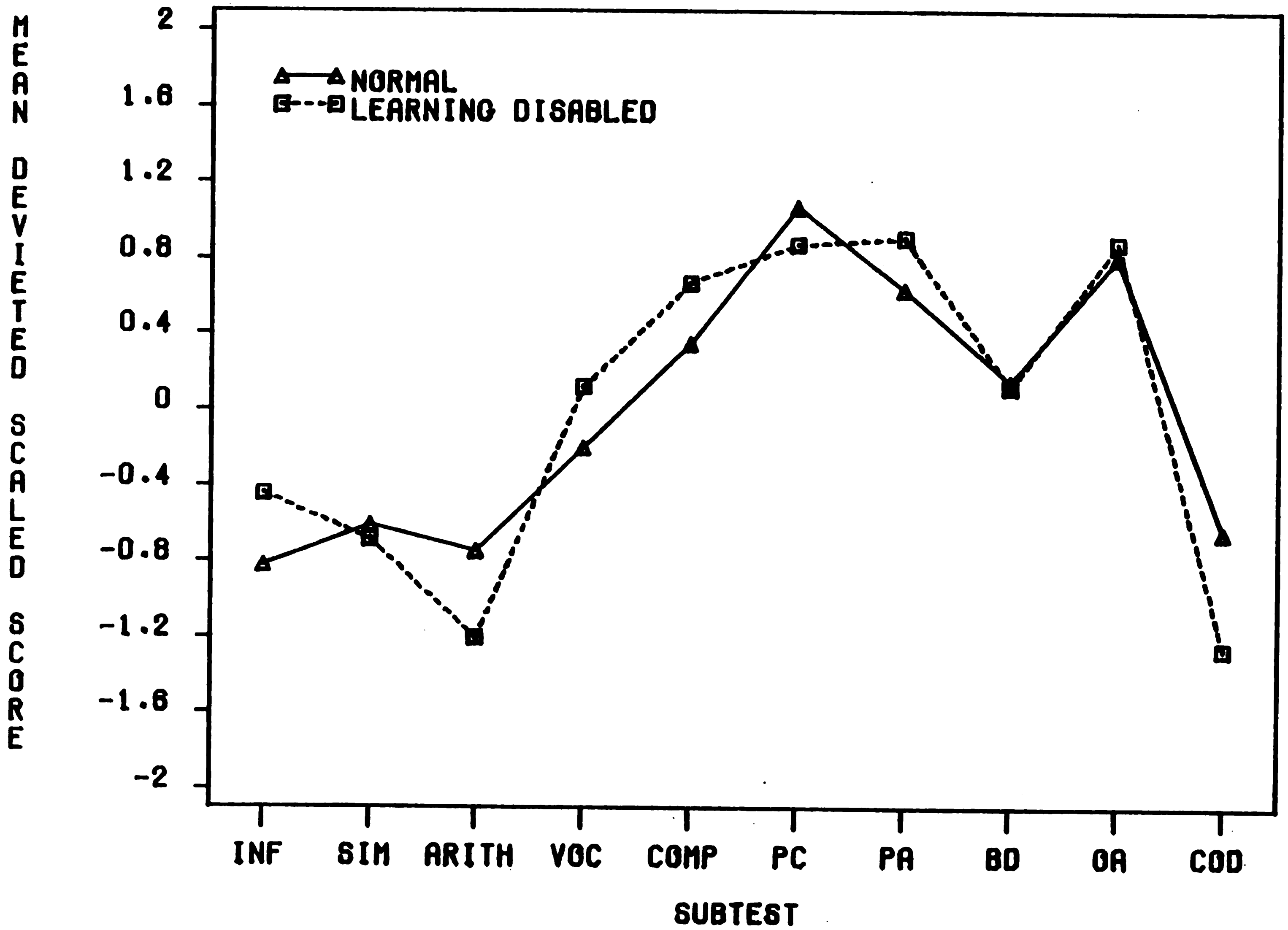


FIGURE 3. MEAN MEAN-DEVIATED SUBTEST SCORES FOR NORMAL AND LEARNING DISABLED GROUPS



Figures 1 and 2 contain graphic representations of the results of the preceeding analyses. Figure 1 is a graph of the average subtest scores for each of the five diagnostic groups. Figure 2 shows the average mean-deviated subtest scores for each group. That is, each child's mean subtest score was subtracted from each subtest score before the group means were computed. Figure 3 is a less cluttered version of Figure 2, showing the average mean-deviated scores of only the normal and learning disabled groups. The similarity among these curves is striking.

Best Predictors of Full Scale Intelligence

The question of what subset of WISC-R would be of most use in predicting Full Scale IQ in the absence of a complete profile was addressed for three groups -- the normal sample, the entire learning disabled sample, and those LD children in grade eight or higher. Summaries of step-wise multiple regressions for these three groups, with Full Scale IQ as the criterion, are reported in Tables 21, 22, and 23. In these tables, at each of the steps, the partial correlations between the criterion and those subtest scores not already in the regression equation are reported. Given the high degree of multicolinearity among the predictors, these partials help to indicate the arbitrary nature of the statistical decision to enter one variable ahead of another at some of the steps. The analysis for the eighth grade and higher learning disabled children, reported in Table 23, should be interpreted cautiously. The availability of only 23 cases for this renders these results extremely unreliable.

Table 21. Summary of Regression Using Subtest Scores to Predict Full Scale IQ (Normal Sample)

Step	Variable Entered	F to Enter	Sig. <	R	R^2	Overall F	Sig. <	Partial Corrleations between Variables Not in Equation and Criterion Sim	OA	Ari	Cod	Comp	BD	Inf	PC	PA	Voc	Order of Partial
0	-	-	-	-	-	-	-	.65	.56	.55	.38	.68	.61	.68	.59	.61	.79	0
1	Voc	363.2	.001	.788	.621	363.2	.001	.34	.40	.36	.29	.38	.46	.37	.47	.49		1
2	PA	71.2	.001	.844	.713	274.7	.001	.31	.34	.31	.26	.34	.38	.36	.44			2
3	PC	52.0	.001	.876	.768	242.8	.001	.30	.24	.33	.30	.35	.32	.37				3
4	Inf	35.2	.001	.895	.800	219.2	.001	.24	.26	.30	.29	.32	.34					4
5	BD	28.3	.001	.907	.823	202.9	.001	.19	.16	.26	.30	.38						5
6	Comp	37.2	.001	.921	.849	203.3	.001	.09	.22	.27	.32							6
7	Cod	24.9	.001	.930	.865	197.0	.001	.13	.24	.26								7
8	Ari	15.5	.001	.935	.874	185.9	.001	.10	.25									8
9	OA	14.7	.001	.939	.882	177.4	.001	.10										9
10	Sim	2.1	.152	.940	.883	160.7	.001											

Table 22. Summary of Regression Using Subtest Scores to Predict Full Scale IQ (Learning Disabled Sample)

Step	Variable Entered	F to Enter	Sig. <	R	R^2	Overall F	Sig. <	Partial Corrleations between Variables Not in Equation and Criterion: Comp	Cod	Inf	PA	OA	Ari	Voc	PC	BD	Sim	Order of Partial
0	-	-	-	-	-	-	-	.61	.46	.58	.62	.58	.55	.70	.61	.63	.70	0
1	Sim	170.9	.001	.704	.496	170.9	.001	.36	.45	.33	.51	.56	.35	.44	.53	.61		1
2	BD	101.0	.001	.826	.681	185.0	.001	.35	.37	.41	.42	.33	.34	.44	.46			2
3	PC	47.2	.001	.866	.750	172.1	.001	.38	.35	.40	.33	.29	.36	.46				3
4	Voc	46.1	.001	.896	.803	174.4	.001	.21	.36	.26	.35	.33	.41					4
5	Ari	33.7	.001	.914	.836	173.0	.001	.23	.34	.20	.36	.37						5
6	OA	26.6	.001	.926	.856	170.2	.001	.26	.28	.25	.34							6
7	PA	22.2	.001	.935	.875	167.4	.001	.29	.24	.32								7
8	Inf	18.8	.001	.942	.887	164.3	.001	.28	.32									8
9	Cod	19.3	.001	.948	.899	164.1	.001	.30										9
10	Comp	16.2	.001	.953	.908	162.9	.001											

Table 23. Summary of Regression Using Subtest Scores to Predict Full Scale IQ (Older Learning Disabled Sample)*

Step	Variable Entered	F to Enter	Sig. <	R	R^2	Overall F	Sig. <	Ari	Sim	Comp	Inf	Cod	BD	PA	Voc	PC	OA	Order of Partial
								Partial Corrleations between Variables Not in Equation and Criterion										
0	-	-	-	-	-	-	-	.27	.31	.49	.37	.23	.53	.38	.38	.53	.56	0
1	OA	9.6	.005	.560	.314	9.6	.005	.40	.37	.50	.51	.12	.35	.44	.52	.55		1
2	PC	8.8	.008	.724	.524	11.0	.001	.33	.40	.51	.38	.06	.50	.45	.52			2
3	Voc	7.1	.015	.808	.653	11.9	.001	.46	.20	.32	.10	.12	.50	.51				3
4	PA	6.5	.020	.863	.745	13.2	.001	.47	.12	.37	.05	.15	.60					4
5	BD	9.3	.007	.914	.836	17.3	.001	.34	.22	.16	.27	.45						5
6	Cod	4.0	.063	.932	.868	17.6	.001	.21	.30	.25	.39							6
7	Inf	2.6	.127	.942	.888	17.0	.001	.21	.20	.28								7
8	Comp	1.2	.298	.947	.897	15.2	.001	.08	.11									8
9	Sim	.1	.706	.948	.898	12.7	.001	.07										9
10	Ari	.1	.815	.948	.898	10.6	.001											

*Interpret with caution as only 23 cases contributed to this regression.

For the normal sample, the best four indicators of Full Scale IQ were Vocabulary, Picture Arrangement, Picture Completion, and either Information or Comprehension (partial r = .37 and .35, respectively before the fourth step). These four predictors account for about 80 percent of the variance in Full Scale scores for these normal children.

For the entire learning disabled group, the best four predictors of Full Scale IQ were Similarities, Block Design, Picture Completion, and Vocabulary, which together accounted for about 80 percent of criterion variance. Sattler (1982) suggested various short form versions of the WISC-R. His best four subtest version included the Information, Vocabulary, Comprehension, and Block Design subtests. In the standardization sample, this composite had a .947 correlation with Full Scale IQ. For the learning disabled sample of the present study, Sattler's composite had a multiple R of .873 with Full Scale IQ. This should be compared with the .896 multiple R for the best four predictors in this analysis. The difference between the two multiple correlations from the present study is hardly worth noting. Thus, Sattler's composite is about as good a predictor of Full Scale IQ for these learning disabled children as can be found in these data.

The fact that there were so few (23) older learning disabled children precludes any reasonable interpretation of the data in Table 23.

CHAPTER 5

DISCUSSION

Overview

In this chapter, utility conclusions are drawn from the results of the analyses of the WISC-R subtest scores of the educational subgroups. Then, the implications of the present data for a short form version of the WISC-R for learning disabled children are discussed.

The conclusions section is followed by separate recommendations sections for clinical practice, public policy, and research.

Conclusions

The learning disabled children were more likely statistically than the normal children to exhibit a twelve or more point Verbal-Performance difference. However, contrary to theoretical expectations, the learning disabled children were not more likely to have Performance IQ twelve or more points higher than Verbal IQ. In fact, classifications of children using the Verbal-Performance discrepancy of at least twelve points in either direction, and Performance at least twelve points greater than Verbal rules were correct for only 60 percent and 56 percent of cases, respectively. In neither case, did the decision rule meet the practical definition of useful established for the present study. For practical applications, these results are barely better than chance classification.

The scatter index results were no more promising than those of the V-P difference analyses. In no case was a diagnostic rule based on scatter accurate more than 60 percent of the time.

The SP, VC, and AC recategorized scores placed children in the correct groups 58 percent of the time. The ACI rule was successful in classifying 57 percent of the children. Neither rule met the 85 percent correct utility requirement.

No linear combination of WISC-R subtests could correctly classify learning disabled and normal children more than 62 percent of the time. Hence, no useful learning disability decision rule was supported in these data.

No reliable statistical differences resulted either from comparisons of older and younger learning disabled and normal children or from comparisons of high and low Full Scale IQ learning disabled and normal children. Therefore no useful decision rules concerning these groups could be generated.

The 63 percent and 66 percent correct classifications of emotionally impaired and normal children using either the any twelve point V-P difference or the Performance at least twelve points greater than Verbal rules are not considered useful.

The successful classification rate using Dean's sign was 63 percent. However, this is somewhat misleading as it includes as correct the 41 percent of normal children who had at least one of the emotional impairment signs. In any case, the sign rule was not useful.

In the discriminant analysis that included all five diagnostic groups, only 47 percent of the children were classified correctly. Thus, using all possible linear combinations of the subtest scores did not provide a useful diagnostic rule.

What summary conclusion can be drawn from these analyses? Quite simply, by the rather lenient 85 percent correct criterion, none of the diagnostic rules tested in this study was useful. In fact, doubling the allowable error rate to 30 percent leads to the same conclusion. It would seem that, to the extent that the sample of children used in this study is representative of the population of those referred for educational diagnoses, the use of either Verbal-Performance discrepancies, the WISC-R subtest scores, or their various suggested recategorizations for classification is, to say the least, misguided.

The selection of a short form of the WISC-R to use with learning disabled children requires a more extensive sample than was available for the present study. However, there was little evidence in these data to suggest that the best predictors of Full Scale IQ for the learning disabled are appreciably different than those of normal children.

Clinical Practice Recommendations

Any school psychologist who uses WISC-R data to diagnose educational handicaps should be advised that there is little or no useful information in any linear combination of the subtest scores to aid in such diagnosis. To many who have read and understood Kaufman's (1976a) paper on the extent of WISC-R variation in the standardization sample, this will come as no surprise. Others, who, like Reynolds (1981), still believe that WISC-R data contain valuable diagnostic information, should take this admonition to heart. The use of these data for differential diagnosis is, from a practical standpoint, not much better than a random process. Cherished hypotheses notwithstanding, there is simply no credible support either in the educational psychology literature or, more particularly, in the present study for this practice.

These limitations apply equally to the recently available WISC-R diagnostic computer programs. In fact, given the total absence of evidence of discriminating information in the WISC-R subscales, it is difficult to imagine how an ethically defensible diagnostic program could be constructed.

Any advances in diagnostic rules will have to await changes in the definitions of categories of educational handicap and research aimed at their specification.

Public Policy Recommendations

It is quite evident that the current usages of such categories as learning disabled and emotionally impaired are very broad. Until these educational handicaps are broken down into more homogeneous and more clearly specified classifications neither clinical practice nor research can advance very far from their present state of confusion. It is not reasonable to expect more precise educational diagnoses until the form of such precision is legally mandated. Without precise diagnostic data, research on the utility of various formulations will be fruitless.

Educational researchers and school psychologists should lend their expertise to policy makers in systematic attempts to define appropriate categories. Of course, this will have to be an iterative process requiring rounds of discussion leading to tentative definitions followed by empirical research on the sensitivity and specificity of those definitions.

Research Recommendations

In addition to the policy oriented research discussed in the previous section, two kinds of research seem necessary.

First, because old habits die slowly and because proper science demands replication, the analyses of the present study should be applied to at least three other large samples of WISC-R scores. It is likely that such data sets are available from school districts around the country. There is no reason to expect that the results of such studies will differ in any practically meaningful way from those of the present study.

A second, more interesting, line of research would address the problem of diagnostic utility more directly. It may be possible to arrive at a simple measure that captures the costs associated with poor decisions and the benefits of correct ones. Preliminary work on this topic would have to include such factors as educational gains resulting from appropriate placements, the dollar cost per student of various special programs, the later academic and vocational successes and failures of both correctly placed and misplaced students, and the impact of special education placements on the education process in the regular classroom. Of course, this list is by no means exhaustive. More advanced work in the area would have to tackle such elusive problems as the psychological and social costs of educational failure as well as the stigma associated with misdiagnosis.

From a societal point of view, it would seem foolish for research to continue to focus solely on the location of statistically significant group-level differences. Until the diagnostic questions are answered, coherent theories of educational handicaps are not likely to emerge and those that do emerge will not be testable.

REFERENCES

Ackerman, P., Peters, J., Dyckman, R. Children with specific learning disabilities: WISC profiles. *Journal of Learning Disabilities*, 1981, *4*, 150-166.

Anderson, M., Kaufman, A. S., and 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.

Andrew, J. M. Delinquency, the Wechsler P > V sign, and the I-level system. *Journal of Clinical Psychology*, 1974, *30*, 331-335.

Bannatyne, A. Diagnosing learning disabilities and writing remedial prescriptions. *Journal of Learning Disabilities*, 1968, *1*, 242-249.

Bannatyne, A. *Language, reading, and learning disabilities.* Springfield, Illinois: Charles C. Thomas, 1974.

Bortner, M., and Birch, H. G. Patterns of intellectual ability in emotionally impaired and brain-damaged children. *Journal of Special Education*, 1969, *6*, 351-369.

Clarizio, H., and Bernard, R. Recategorized WISC-R scores of learning disabled children and differential diagnosis. *Psychology in the Schools*, 1981, *18*, 5-12.

Cronbach, L. J., and Gleser, G. C. Assessing similarity between profiles. *Psychological Bulletin*, 1953, *50*, 456-473.

Dean, R. Patterns of emotional disturbance on the WISC-R. *Journal of Clinical Psychology*, 1977, *33*, 486-490.

Dean, R. Distinguising learning disabled and emotionally disturbed children on the WISC-R. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 381-382.

The Fedreal Register, Assistance to states for education of handicapped children. 1976, *41*,.
52404-52407

Gutkin, T. B. Bannatyne patterns of caucasian and Mexican-American learning disabled children. *Psychology in the Schools*, 1979a, *16*, 178-183.

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

Hamm, H., and Evans, J. G. WISC-R subtest patterns of severely emotionally disturbed students. *Psychology in the Schools*, 1978, *15*, 188-190.

Henry, S. and Wittman, R. Diagnostic implications for Bannatyne's recategorized WISC-R scores for identifying learning disabled children. *Journal of Learning Disabilities*, 1981, *14*, 517-520.

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

Kaufman, A. Factor analysis of the WISC-R at eleven age levels between 6 1/2 and 16 1/2 years. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 135-147.

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

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

Kaufman, A. *Intelligent testing with the WISC-R.* New York: John Wiley and Sons, 1979.

Kaufman, A. The WISC-R and learning disabilities assessment: state of the art. *Journal of Learning Disabilities*, 1981, *14*, 520-526.

Lewandowski, N., Saccuzzo, D., and Lewandowski, D. The WISC-R as a measure of personality types. *Journal of Clinical Psychology*, 1977, *33*, 285-291.

Meehl, P. E., and Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 1955, *52*, 194-216.

Mercer, C., Forgnone, C., and Wolking, W. Definitions of learning disabilities used in the United States. *Journal of Learning Disabilities*, 1976, *9*, 376-386.

Miller, M. On the attempt to find WISC-R profiles for learning and reading disabilities (A response to Vance, Wallbrown, and Blaha). *Journal of Learning Disabilities*, 1980, *13*, 338-340.

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

Reynolds, C. A note on determining significant discrepancies among category scores on Bannatyne's regrouping of WISC-R subtests. *Journal of Learning Disabilities*, 1981, *14*, 468-469.

Reynolds, C., and Gutkin, T. B. Statistics for the interpretation of Bannatyne recategorization of WPPSI subtests. *Journal of Learning Disabilities*, 1981, *14*, 464-467.

Ryckman, D. Searching for a WISC-R profile for learning disabled children: an inappropriate task? *Journal of Learning Disabilities*, 1981, *14*, 508-511.

Rugel, R. WISC subtest scores of disabled readers: a review with respect to Bannatyne's recategorization *Journal of Learning Disabilities*, 1974a, *7*, 48-55.

Rugel, R. The factor structure of the WISC in two populations of disabled readers. *Journal of Learning Disabilities*, 1974b, *7*, 581-585.

Saccuzzo, D. and Lewandowski, D. The WISC-R as a diagnostic tool. *Journal of Clinical Psychology*, 1976, *32*, 115-124.

Sattler, J. *Assessment of childrens' intelligence and special abilities.* Boston: Allyn and Bacon, 1982.

Satz, P. and Mogel, S. An abbreviation of the WISC for clinical use. *Journal of Consulting Psychology*,,,.
1967 31 108

Silverstein, A. Reliability and abnormality of differences between DTVP subtest scores. *Psychology in the Schools*, 1973, *10*, 204-206.

Stevenson, L. WISC-R analysis: implications for diagnosis and intervention. *Journal of Learning Disabilities*, 1980, *13*, 60-63.

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

Smith, M. Stability of WISC-R subtest profiles for learning disabled children. *Psychology in the Schools*, 1978, *15*, 4-7.

Smith, M. D., Coleman, J. M., Dokecki, P. R., and Davis, E. E. Recategorization of WISC-R scores of learning disabled children. *Journal of Learning Disabilities*, 1977a, *10*, 48-54.

Smith, M. D., Coleman, J. M., Dokecki, P. R., and Davis, E. E. Intellectual characteristics of school labeled learning disabled children. *Exceptional Children*, 1977b, *43*, 352-359.

Strichart, S and Love, E. WISC-R performance of children referred to a university center for learning disabilities. *Psychology in the Schools*, 1979, *16*, 183-188.

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

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.

Thompson, R. The diagnostic utility of Bannatyne's recategorized WISC-R scores with children referred to a developmental evaluation center *Psychology in the Schools*, 1981, *18*, 43-47.

Vance, H., Wallbrown, F., and Blaha, J. Determining WISC-R profiles for reading disabled children. *Journal of Learning Disabilities*, 1978, *11*, 657-661.

Wechsler, D. *Wechsler Intelligence Scale for Children (Manual)*. New York: The Psychological Corporation, 1949.

Wechsler, D. and Jaros, E. Schizophrenic patterns on the WISC. *Journal of Clinical Psychology*, 1965,..

21 288-291

Yudin, L. An abbreviated form of the WISC for use with emotionally disturbed children. *Journal of Consulting Psychology*, 1966, *30*, 272-275.

APPENDIX

Table A1. Children with Particular SP-VC-AC Pattern
(4 Point Difference)

				Norm (not placed)		Learning Disabled	
				N	Percent	N	Percent
1.	SP=VC	VC=AC	SP=AC	211	(75)	153	(69)
2.	SP=VC	VC=AC	SP>AC	15	(5)	7	(3)
3.	SP=VC	VC=AC	SP<AC	2	(1)	0	(0)
4.	SP=VC	VC>AC	SP=AC	6	(2)	14	(6)
5.	SP=VC	VC>AC	SP>AC	2	(1)	6	(3)
6.	SP=VC	VC<AC	SP=AC	9	(3)	8	(4)
7.	SP=VC	VC<AC	SP<AC	2	(1)	3	(1)
8.	SP>VC	VC=AC	SP=AC	11	(4)	7	(3)
9.	SP>VC	VC=AC	SP>AC	10	(4)	7	(3)
10.	SP>VC	VC>AC	SP>AC	0	(0)	0	(0)
11.	SP>VC	VC<AC	SP=AC	8	(3)	9	(4)
12.	SP>VC	VC<AC	SP>AC	0	(0)	0	(0)
13.	SP>VC	VC<AC	SP<AC	0	(0)	0	(0)
14.	SP<VC	VC=AC	SP=AC	2	(1)	4	(2)
15.	SP<VC	VC=AC	SP<AC	1	(0)	0	(0)
16.	SP<VC	VC>AC	SP=AC	1	(0)	4	(2)
17.	SP<VC	VC>AC	SP>AC	0	(0)	0	(0)
18.	SP<VC	VC>AC	SP<AC	0	(0)	0	(0)
19.	SP<VC	VC<AC	SP<AC	0	(0)	0	(0)

Table A2. Children with Particular SP-VC-AC Pattern (2 Point Difference)

				Norm (not placed)		Learning Disabled	
				N	Percent	N	Percent
1.	SP=VC	VC=AC	SP=AC	78	(28)	65	(29)
2.	SP=VC	VC=AC	SP>AC	11	(4)	16	(7)
3.	SP=VC	VC=AC	SP<AC	9	(3)	6	(3)
4.	SP=VC	VC>AC	SP=AC	12	(4)	2	(1)
5.	SP=VC	VC>AC	SP>AC	27	(10)	22	(10)
6.	SP=VC	VC<AC	SP=AC	10	(4)	4	(2)
7.	SP=VC	VC<AC	SP<AC	24	(9)	9	(4)
8.	SP>VC	VC=AC	SP=AC	12	(4)	6	(3)
9.	SP>VC	VC=AC	SP>AC	33	(12)	17	(8)
10.	SP>VC	VC>AC	SP>AC	2	(1)	4	(2)
11.	SP>VC	VC<AC	SP=AC	23	(8)	28	(13)
12.	SP>VC	VC<AC	SP>AC	8	(3)	7	(3)
13.	SP>VC	VC<AC	SP<AC	5	(2)	3	(1)
14.	SP<VC	VC=AC	SP=AC	5	(2)	1	(0)
15.	SP<VC	VC=AC	SP<AC	8	(3)	4	(2)
16.	SP<VC	VC>AC	SP=AC	11	(4)	19	(9)
17.	SP<VC	VC>AC	SP>AC	2	(1)	6	(3)
18.	SP<VC	VC>AC	SP<AC	0	(0)	3	(1)
19.	SP<VC	VC<AC	SP<AC	0	(0)	0	(0)

Table A3. Children with Particular SP-VC-AC Pattern (1.5 Point Difference)

				Norm (not placed)		Learning Disabled	
				N	Percent	N	Percent
1.	SP=VC	VC=AC	SP=AC	54	(19)	38	(17)
2.	SP=VC	VC=AC	SP>AC	6	(2)	10	(5)
3.	SP=VC	VC=AC	SP<AC	4	(1)	3	(1)
4.	SP=VC	VC>AC	SP=AC	5	(2)	4	(2)
5.	SP=VC	VC>AC	SP>AC	36	(13)	28	(13)
6.	SP=VC	VC<AC	SP=AC	6	(2)	9	(4)
7.	SP=VC	VC<AC	SP<AC	25	(9)	14	(6)
8.	SP>VC	VC=AC	SP=AC	6	(2)	6	(3)
9.	SP>VC	VC=AC	SP>AC	37	(13)	20	(9)
10.	SP>VC	VC>AC	SP>AC	10	(4)	6	(3)
11.	SP>VC	VC<AC	SP=AC	32	(11)	28	(13)
12.	SP>VC	VC<AC	SP>AC	13	(5)	10	(5)
13.	SP>VC	VC<AC	SP<AC	10	(4)	6	(3)
14.	SP<VC	VC=AC	SP=AC	2	(1)	2	(1)
15.	SP<VC	VC=AC	SP<AC	11	(4)	4	(2)
16.	SP<VC	VC>AC	SP=AC	12	(4)	17	(8)
17.	SP<VC	VC>AC	SP>AC	5	(2)	11	(5)
18.	SP<VC	VC>AC	SP<AC	3	(1)	6	(3)
19.	SP<VC	VC<AC	SP<AC	3	(1)	0	(0)

Table A4. Children with Particular SP-VC-AC Pattern
(1 Point Difference)

				Norm (not placed)		Learning Disabled	
				N	Percent	N	Percent
1.	SP=VC	VC=AC	SP=AC	21	(7)	15	(7)
2.	SP=VC	VC=AC	SP>AC	1	(0)	4	(2)
3.	SP=VC	VC=AC	SP<AC	1	(0)	3	(1)
4.	SP=VC	VC>AC	SP=AC	3	(1)	3	(1)
5.	SP=VC	VC>AC	SP>AC	27	(10)	26	(12)
6.	SP=VC	VC<AC	SP=AC	3	(1)	2	(1)
7.	SP=VC	VC<AC	SP<AC	25	(9)	13	(6)
8.	SP>VC	VC=AC	SP=AC	5	(2)	0	(0)
9.	SP>VC	VC=AC	SP>AC	39	(14)	23	(10)
10.	SP>VC	VC>AC	SP>AC	25	(9)	19	(9)
11.	SP>VC	VC<AC	SP=AC	33	(12)	33	(15)
12.	SP>VC	VC<AC	SP>AC	23	(8)	16	(7)
13.	SP>VC	VC<AC	SP<AC	19	(7)	13	(6)
14.	SP<VC	VC=AC	SP=AC	2	(1)	1	(0)
15.	SP<VC	VC=AC	SP<AC	9	(3)	7	(3)
16.	SP<VC	VC>AC	SP=AC	13	(5)	17	(8)
17.	SP<VC	VC>AC	SP>AC	15	(5)	15	(7)
18.	SP<VC	VC>AC	SP<AC	5	(2)	7	(3)
19.	SP<VC	VC<AC	SP<AC	11	(4)	5	(2)