Order Number 8726812

# A comparison of the performances of re-evaluated and newly referred learning disabled students and newly referred non-learning disabled students on the Wechsler Intelligence Scales for Children-Revised and the Woodcock-Johnson Tests of Cognitive Ability

Conroy, David S., Ph.D.

The University of Arizona, 1987

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

**PLEASE NOTE:**

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __√__.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy __√__
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages __√__
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) ________ lacking when material received, and not available from school or author.
12. Page(s) ________ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received _____
16. Other ____________________

# A Comparison of the Performances of Re-Evaluated and Newly Referred Learning Disabled Students and Newly Referred Non-Learning Disabled Students on the Wechsler Intelligence Scales for Children-Revised and the Woodcock-Johnson Tests of Cognitive Ability

by

David Conroy

---

A Dissertation Submitted to the Faculty of the

DIVISION OF EDUCATIONAL FOUNDATIONS AND ADMINISTRATION

In Partial Fulfillment of the Requirements
For the degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN EDUCATIONAL PSYCHOLOGY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 8 7

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by David S. Conroy

entitled A Comparison of the Performances of Re-Evaluated and Newly Referred Learning Disabled Students and Newly Referred Non-Learning Disabled Students on the Wechsler Intelligence Scales for Children-Revised and the Woodcock-Johnson Tests of Cognitive Ability

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

| | |
|---|---|
| Darrell Sabers | 4-27-87 Date |
| John E. Obrzut | 4-27-87 Date |
| Candace S. Bos | 4/27/87 Date |
| Anthony K. Van Reusen | 4/27/87 Date |
| Shitala P. Mishra | 4/27/87 Date |

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

| | |
|---|---|
| Darrell Sabers | 7-10-87 |
| Dissertation Director | Date |

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Donald S. Conway

ACKNOWLEDGEMENTS

I would like to thank Dr. Darrell Sabers for serving as chairman of my committee. Your guidance of this project is gratefully appreciated. I would also like to thank Dr. Candy Bos, Dr. Shitala Mishra, Dr. John Obrzut, and Dr. Tony Van Ruesen for your time, expertise and the many valuable contributions you made. Thanks to the Doctors O'Reilly for allowing me to abuse your computer and especially for keeping the U of C attitude alive and well.

To my family, Mom and Dad, Donna, Janice, Michael and Peggy: I've learned more about more important things from you than I ever have in school. I'm very fortunate and I know it. Thank you all.

TABLE OF CONTENTS

TABLE OF CONTENTS--continued

LIST OF TABLES

LIST OF TABLES--continued

ABSTRACT

There has been much controversy concerning the comparability of the Wechsler Intelligence Scales for Children-Revised (WISC-R) and the Woodcock-Johnson Tests of Cognitive Ability (WJTCA). Previous research has raised the issue of a mean score discrepancy between the tests when used with the learning disabled. This study analyzed and compared performances on these two tests by re-evaluated and newly referred LD students and newly referred non-LD students. In addition, subtypes of LD students were formed on the basis of achievement test scores. These students' test performances were also analyzed and compared.

The results of this study were consistent with previous research. The Full Scale scores from the two tests were highly correlated in all three groups, but the WISC-R was significantly higher than the WJTCA for each group. Across the identified LD subtypes there was a significant difference between the Full Scale scores from the two tests. However, meaningful patterns of strengths and weaknesses across aspects of cognitive functioning were not uncovered.

These results indicate that the WISC-R and WJTCA result in significantly different estimates of the cognitive ability of LD and referred students. This difference can be attributed to a combination of three possible explanations--the effects of the use of non-random samples, the use of different norm groups when the tests were standardized, and the tests contain different content.

# CHAPTER 1

# INTRODUCTION

PL 94-142, The Education of All Handicapped Children Act (U.S. Office of Education, 1977), has served as one of the main forces shaping the practice of school psychology. This legislation has affected the provision of psychological services in the public schools in a variety of ways. Due process considerations, issues of confidentiality, placement procedures, and least restrictive environment requirements have all affected how school psychologists do their job. Perhaps the greatest effect has been felt in the area of assessment. Because of the requirements of PL 94-142 psycho-educational assessment continues to be the central function of school psychologists. Indeed, while there are a variety of functions in the schools that school psychologists could fulfill, it has been reported that they spend two-thirds or more of their time on individual psycho-educational assessments (Fagan, 1981). The majority of these assessments are conducted to determine if a student experiencing academic difficulties is eligible for special education services.

## PL 94-142 Definition of a Learning Disability

The criteria for determining the existence of a learning disability are detailed in PL 94-142. This law specifies that a learning disability exists if:

1) a child does not achieve commensurate with his or her age and ability levels in oral expression, listening comprehension, written expression, reading skill or comprehension, math calculation or reasoning, given appropriate learning experiences for the child's age and ability level and

2) there is a severe discrepancy between achievement and intellectual ability in one or more of the areas listed above (U.S. Office of Education, 1977).

This definition, with only slight alterations, has been adopted by state departments of education and local school districts. School psychologists and special education personnel have been charged with operationalizing these criteria to determine who is, and who is not, learning disabled.

Operationalizing these criteria requires a method to determine a severe discrepancy between achievement and ability. This is typically accomplished by the administration of a test of intelligence, such as the Wechsler Intelligence Scale for Children-Revised (WISC-R), and tests of achievement, such as the Peabody Individual Achievement Test (PIAT). Results from other assessment instruments, observations, and clinical judgment are also used in deciding whether a severe discrepancy exists between a student's estimated ability and achievement levels.

While straightforward, this procedure is plagued by problems. One difficulty is that the testing is usually conducted by more than one person. The presence of different examiners may differentially affect a student's performance on the various tests. Another major difficulty arises when comparing performance by an individual on two different tests. Comparing performance on tests that have been normed on different samples at different times introduces error into the resulting difference score. A third major problem arises when dealing with these difference scores. PL 94-142 does not define what is meant by a "severe discrepancy." School districts are left to their own resources to decide how large a discrepancy must be present before a student is considered learning disabled. Local control of this issue, as intended by law, results in a great deal of inconsistency in terms of students identified as learning disabled. This inconsistency can be very perplexing to parents when they discover that in District B their child is no longer considered learning disabled, making the child ineligible for the special education services that had been available when they lived within the boundaries of District A.

Epps, Ysseldyke, and Algozzine (1983) applied 14 representative operational definitions of learning disabilities to a sample of 48 school-identified learning disabled (LD) students and 96 non-LD students. These definitions identified 7 to 81 percent of the school-identified LD students and 3 to 65 percent of non-LD students as LD. These researchers concluded that students identified as LD in one setting are likely to be different from those identified as LD in another setting, supporting the contention discussed above that there is indeed much local variance in the identification of LD students.

## Woodcock-Johnson Psycho-Educational Battery

Some of the difficulties that presently hamper the valid assessment of suspected LD students are addressed and possibly rectified by the Woodcock-Johnson Psycho-Educational Battery (WJ; Woodcock & Johnson, 1977). The WJ consists of 27 individually administered subtests designed to test an individual's cognitive ability, scholastic aptitude, academic achievement, and interests. The Battery is divided into three parts. Part I, the Woodcock-Johnson Tests of Cognitive Ability (WJTCA), consists of 12 subtests which assess cognitive ability. As a composite they form the Woodcock-Johnson Full Scale Broad Cognitive Ability (WJ Full Scale BCA) score. These subtests can also be joined in various combinations to derive clusters describing areas of cognitive functioning and specific scholastic aptitudes. Part II, the Woodcock-Johnson Tests of Achievement (WJTA), contains 10 subtests assessing academic achievement which can be grouped into four achievement clusters. Part III consist of five subtests designed to assess an individual's pattern of interest.

Woodcock (1984) states that the Battery was developed "as a response to the need for an instrument based on a psycho-educational assessment model that would provide the necessary information for making educational decisions" (p.342). Since its introduction, the WJ has rapidly gained acceptance by school personnel and has been widely adopted for evaluation of students referred for possible learning disabilities (Bracken, Prasse, & Breen, 1984; Coleman & Harmer, 1985). There are several important characteristics of the WJ that can account for its wide-spread acceptance.

One important characteristic of the WJ is that both the cognitive and achievement measures were normed on the same standardization sample. This eliminates the error inherent in comparing performances across tests normed on different samples at different times. Thus, one could expect the difference score resulting from the WJ to be more meaningful than difference scores resulting from comparisons of scores on a WISC-R and PIAT, for example.

A second important characteristic of the Battery is that the entire Battery can be administered by one person, thus eliminating the potential error involved when more than one examiner is involved. At the same time, however, this characteristic introduces examiner bias as a potential alternative source of error. An examiner assessing a student's achievement would be knowledgeable about the student's ability level, and since Parts I and II of the Battery can be given in either order, the reverse could also be true. This potential for bias can be reduced, but not eliminated, by not scoring the protocols until the entire battery has been administered.

A third factor responsible for the WJ's popularity is that the various subtests are grouped into clusters (weighted linear composites of two or more subtests) that are considered indices of more general functions (McGue, Shinn, & Ysseldyke, 1982). This clustering is an attempt to increase the reliability of scores and eliminate the temptation of over-generalizing from specific subtests (Cummings & Moscatto, 1984). In this way data are presented on a student's performance in four areas of cognitive functioning as well as four scholastic aptitudes. These clusters are based on varying combinations of subtests from Part I of the Battery. The various Aptitude Clusters, as measures of predicted achievement, can be directly compared with the appropriate Achievement clusters, composed of subtests from Part II of the Battery.

A fourth factor contributing to the popularity of the WJ is the availability of the Relative Performance Index (RPI). One use of this index is to compare a student's performance on the Achievement Clusters with his/her performance on the Aptitude Clusters. The RPI provides a comparison between a student's actual achievement and expected achievement. Functioning levels ranging from "very superior" to "severe deficit" can then be attached to RPI or cluster difference scores to give interpretive meaning.

Similar to the use of a WISC-R Full Scale IQ below 70 as one factor in determining the identification of a student as Educable Mentally Handicapped, an RPI score of some agreed upon value could be used to determine the existence of a severe discrepancy between aptitude and achievement. This standardization in determining when a severe discrepancy exists would help reduce the variability that currently exists. Along this line, Reeve, Hall, and Zakreski (1979) report that the five Learning Disabilities Research Institutes agreed to use the WJ as a "marker variable" or common measure to describe students involved in the Institutes' research. Unfortunately this plan was not implemented.

## Issues Concerning the Use of the WJ

Given the advantages that the WJ purports to offer, its level of popularity is not surprising. However, at this point it would be inappropriate to view the Battery as a panacea for professionals involved in psycho-educational decision making. There is concern that the WJTCA underestimates a referral or identified LD student's level of cognitive functioning. Research with these populations has identified differences ranging from 2.5 to 11.9 standard score points with the WJ Full Scale BCA consistently being lower than the WISC-R Full Scale IQ (Reeve, Hall, Zakreski, 1979; Ysseldyke, Shinn, & Epps, 1981; Algozzine, Ysseldyke, & Shinn, 1982). Other researchers, however, have found no significant mean difference between the two scores (Ipsen, McMillan, & Fallen, 1982; McGrew, 1983). This issue is important because if achievement levels are held constant and the WJTCA underestimates an LD or referral student's cognitive functioning level, it makes it more difficult to obtain a severe discrepancy between aptitude and achievement, resulting in inappropriate placement decisions. Students who would otherwise qualify as learning disabled would not be identified.

A different problem arises if achievement levels are underestimated to the same degree as cognitive functioning. While the discrepancy would remain the same the result would be a uniform decrease in scores. Students whose "true" functioning is within the borderline range of intellectual ability with corresponding achievement levels, would suddenly find themselves within the educable mentally handicapped

(EMH) range. Although measures of adaptive behavior are necessary before a student can be identified as EMH, a concurrent underestimation of intellectual ability and academic achievement greatly increases the potential for misidentifying a student.

Hypotheses ranging from norming errors on the WJ (Reeve, et al., 1979) to an emphasis on academic achievement (Ysseldyke, et al., 1981) have been suggested to explain LD or referral students' poorer performance on the WJTCA than on the WISC-R. Woodcock (1980,1984) has responded to these suggestions by critically examining the research and concluding that many of the studies are flawed because of instrument, score, or sample bias. Observed differences between the tests may be artifacts resulting from these flaws rather than from supposed norming errors or an inappropriate achievement emphasis. He also noted that because the WJTCA correlates higher with achievement measures than does the WISC-R, differences should be expected between the two tests for selected samples. Sabers (personal communication) has also suggested that these differences are to be expected, but he has looked to the tests' standardization samples as the explanatory mechanism.

One of the difficulties with research in this area is that the research populations usually consist of school identified LD students or referral groups. The variability that exists in the identification of LD students was previously discussed, but within a research context this variability makes it difficult to generalize beyond any specific study or to make comparisons across studies. Indeed, even if learning disabilities were consistently defined across settings, there is increasing recognition that the LD population is a heterogeneous group with varying patterns of performance (Rourke, 1985). One of the pressing needs in the field of learning disabilities is to develop methods for differentiating subgroups or subtypes within the larger LD population (Adelman & Taylor, 1985). McGrew (1983) has suggested that the differences that have been found between the WJ Full Scale BCA score and the WISC-R Full Scale IQ for LD and referral students may be an accurate reflection of certain of these students' performances. It may be that for certain types of students the WISC-R and WJTCA do not function as equivalent measures of cognitive ability. Research conducted to date has not addressed this issue.

In summary, research conducted to date has not settled the issue of the mean score discrepancy between the WISC-R Full Scale IQ and the WJ Full Scale BCA score for LD or referral groups. Various explanations have been given regarding the observed differences but none are completely satisfactory. The practitioner is left not knowing if the WJ is a valuable new assessment instrument ideally suited for assessment of learning disabilities or if the instrument is inappropriate for use in this context.

Research needs to be conducted that will be able to address these issues. Previous research, which is reviewed in Chapter 2, has been plagued with instrument, score, or sample bias (Woodcock, 1980, 1984). Many of these studies have included relatively small samples with a restricted grade range. Those working within the field of learning disabilities have become increasingly aware of the need to differentiate specific subtypes of learning disabled students from the larger heterogeneous LD population. Research examining the WISC-R and WJ needs to be conducted from within this framework.

## Statement of Purpose

The purpose of this study was to analyze and compare scores on the Woodcock-Johnson Psycho-Educational Battery and the Wechsler Intelligence Scale for Children-Revised for three samples of subjects: those referred for psycho-educational evaluation and subsequently identified as LD, those evaluated but not identified as LD, and those who were previously identified as LD and evaluated as part of a three year re-evaluation. In addition, subtypes of the identified LD students were determined and their WISC-R and WJTCA scores were analyzed and compared.

## Research Questions

The following research questions were investigated:

1) Is there a significant difference between the distribution of the WISC-R Full Scale IQs and WJ Full Scale BCA scores for samples of newly referred and identified LD students, newly referred but not placed students, and re-evaluated LD students?

Previous researchers suggest that the WISC-R and WJTCA may not provide similar estimates of the cognitive abilities of referral and LD students (Cummings & Moscatto, 1984).

2) What is the correlation between the WISC-R Full Scale IQ scores and WJ Full Scale BCA scores for these three groups?

Lack of correlation between the scores suggests that the tests are assessing different constructs for these groups. However, both of the tests are intended to be measures of cognitive ability and, as such, are used extensively in psycho-educational decision making.

3) Based on patterns of WJTCA achievement test scores, what subtypes of LD students can be identified in the newly referred LD students and re-evaluated students?

A review of previous research suggests that specific subtypes of LD students can be identified on the basis of patterns of achievement test scores (Rourke, 1985).

4) Is there a significant difference between the distribution of the WISC-R Full Scale IQs and WJ Full Scale BCA scores for the LD subtypes identified in #3 above?

It has been hypothesized that the observed WISC-R/WJ Full Scale score differences are due to differences that exist for certain LD students (McGrew, 1983).

5) What is the correlation between the WISC-R Full Scale IQs and WJ Full Scale BCA scores for the LD subtypes identified in #3 above?

Lack of correlation between the scores would suggest that the tests are assessing different constructs within the LD subtypes.

6) Do the LD subtypes identified in #3 above exhibit different cognitive styles as demonstrated by different patterns of performances across the Cognitive Clusters of the WJTCA?

Extensive research in the LD subtyping area has attempted to identify particular patterns of cognitive abilities corresponding to specific LD subtypes (Rourke, 1985). The WJTCA Cognitive Clusters may prove to be a powerful means of determining patterns of cognitive abilities.

## Significance of Study

This study provides data regarding the comparability of the WISC-R and WJTCA. Previous research has raised the issue of a mean score discrepancy between these tests for referral or LD groups, with the WJ Full Scale BCA providing a lower estimate of cognitive ability than the WISC-R Full Scale IQ. If the WJTCA provides a lower ability estimate than the WISC-R, use of the WJTCA will result in fewer students being identified as learning disabled since a lower ability estimate makes it more difficult to obtain a severe discrepancy between ability and achievement. Research conducted to date has not been able to clarify this issue. This study provides important information about the comparability of the WISC-R and WJTCA.

Research in the area of learning disabilities is being increasingly focused on identifying methods to cull relatively pure subtypes from the larger heterogeneous LD population. Previous research has also focused on particular patterns of cognitive abilities that correspond to LD subtypes. This study provides valuable data about the use of the WJ for these purposes. Additionally, examining the WISC-R/WJTCA mean score discrepancy issue from within a subtype framework helps to clarify this controversy.

# CHAPTER 2

## REVIEW OF THE LITERATURE

In this chapter several different areas of educational and psychological literature will be reviewed in order to provide appropriate background information regarding the research questions discussed at the end of Chapter 1. The first broad area of review will examine the mean score discrepancy issue between the WISC-R and WJTCA, which is the focus of research questions 1 and 2. Topics within this area that will be explored include: a) research comparing the WISC-R and WJTCA b) Woodcock's response to these studies c) the hypothesis that the mean score discrepancy is due to differences in the WISC-R/WJTCA norm groups d) research examining the effect of norm groups on test standardization will be discussed and e) research on two recently introduced intelligence tests will be briefly reviewed as examples of this effect.

The second broad area of review will examine subtypes of learning disabilities, which is the focus of research questions 3-6. Topics in this area to be examined include a) methods of subtyping and b) results of studies conducted within a neuropsychological, social and WISC-R framework.

## Issues Concerning the WISC-R and WJTCA Mean Score Discrepancy

### Research Examining WISC-R/WJTCA Differences

The comparability of the WJ Full Scale score and WISC-R Full Scale IQ for LD and referral groups has been the focus of many studies since the findings of a mean score discrepancy were first reported by Reeve, Hall, and Zakreski in 1979. They report that while the tests had a .79 correlation (.89 with 5 outliers removed) the average WJ Full Scale BCA score was 12.9 points lower than the average WISC-R Full Scale IQ (85.7 vs 98.6) for a sample of 51 LD students. These researchers also raised the issue of sex differences on the WJTCA. The correlation between the Full Scale scores was much lower for girls than for boys (.84 vs .36 with one female outlier removed, .60 with three female outliers removed), and there were differences in the patterns of strengths and weaknesses on the Cognitive Factor Clusters. However, because the small number of females included in their sample (n=8) precluded statistical significance, these results must be viewed with caution. These researchers concluded that the strong correlation between the tests indicates that the tests tap similar areas of cognitive abilities for LD students. They suggested two reasons to account for the observed Full Scale score differences. One is that the WJTCA taps aspects of intelligence that are more difficult for LD children than for their non-LD counterparts. As a second possibility, they suggested that the norms for the WJTCA may be in error.

This issue was next investigated by Ysseldyke and his colleagues. In their first study, Ysseldyke, Shinn, McGue and Epps

(1981) examined the performances of LD and low-achieving students on the WISC-R and WJTCA. The 50 LD children had scores that were, on average, 7.6 points lower on the WJ Full Scale BCA than on the WISC-R Full Scale IQ. The 49 low-achieving students had a 4.3 point difference between the tests, with the WISC-R again being higher. Using the same data on the LD students, Ysseldyke, Shinn, and Epps (1981) examined the WISC-R/WJTCA differences in more detail. They obtained a .67 correlation between the Full Scale scores for the tests, again indicating that the two tests are assessing similar constructs in the LD population. In an effort to explain the statistically significant mean score difference, Ysseldyke et al., rejected the suggestion by Reeve et al. that the observed differences can be traced to differences in performance on subtests that comprise the Perceptual Speed Cluster. Instead, Ysseldyke et al. suggested that the WJTCA is more a test of acquired knowledge (product dominant or crystallized intelligence) than the WISC-R. Low achieving students, such as those referred or placed in LD classes, could be expected to score lower on a product dominant test than on one that is not. These authors also stated that they could not rule out the possibility of errors in the norms tables.

McGrew (1983) has also examined the comparability of the WISC-R and WJTCA. Unlike the previous studies, which examined the scores of previously identified LD students, McGrew analyzed the scores of 52 students referred for a psycho-educational evaluation. He obtained a non-significant 1.9 point difference between the Full Scale scores. He also obtained a significant .74 correlation between the measures. An examination of the scores of those students (n=31) who were later identified as learning disabled resulted in a non-significant 1.5 point difference between the Full Scale scores, although for this sample the WISC-R score was lower. A case-by-case examination showed that some students exhibited a marked discrepancy in performance on the two tests. This led McGrew to conclude that for some students a WISC-R/WJTCA mean score discrepancy is a valid and meaningful difference. He suggested that such differences may underlie the observed differences obtained in previous research.

Bracken, Prasse, and Breen (1984) conducted a study comparing the WJTCA and WISC-R with 142 children who were referred for psycho-educational evaluation. Of this sample, 104 were later identified as LD. These researchers obtained a mean score discrepancy of 9.1 points for the LD children and 6.1 points for the non-LD children, with the WISC-R Full Scale score being higher. A .63 correlation was obtained between the Full Scale scores for the LD children and a .72 correlation was obtained for the non-LD students. These researchers concluded that the two tests are probably measuring somewhat different abilities to somewhat different degrees.

Thompson and Brassard (1984) examined the mean score discrepancy issue by analyzing scores for three groups of children: normal, mild to moderate LD, and severe LD students. There were 20, 20, and 15 students in these groups, respectively. Levels of severity were determined based on percentage discrepancy (30-45% for mild LD, greater than 45% for severe LD) between a student's expected and actual grade level performance on the Woodcock Reading Mastery Test divided by expected grade placement. Students in the normal group had a non-significant 1.1 point mean score difference between the tests. In the mild and severe LD groups the mean WISC-R Full Scale IQ was 10.5 points higher than the mean WJTCA Full Scale BCA score, a statistically significant difference. Correlations of .86, .74, and .93 were obtained between the tests for the normal, mild and severe LD groups, respectively. These researchers concluded that the design of their study and the obtained results make it possible to eliminate the suggestion of norming errors in the WJTCA as an explanation for the observed differences. Similar to Ysseldyke et al., Thompson and Brassard believe that the two tests are assessing similar areas of functioning and that the observed differences are due to the achievement content of the WJTCA.

Estabrook (1984) has also examined the comparability of these two tests. Using scores obtained from 152 students referred for evaluation he found a .77 correlation between the Full Scale scores and a 5.9 point mean score discrepancy, with the WISC-R resulting in the higher score. In order to more appropriately address the question of the degree of relationship between the WISC-R and WJTCA he also conducted a canonical correlation analysis of the scores. He states that because the WISC-R and WJTCA are composed of a number of subtests, the multivariate approach of canonical correlation is a more appropriate technique for examining the relationship between the tests than is the correlation between the Full Scale scores for the two tests. Using this approach he identified three significant canonical correlations. The first accounted for 76% of the variance between linear composites of the tests, the second accounted for 38% of the residual variance and the third accounted for 24% of the remaining variance. An analysis of the subtest composition led Estabrook to label the first a general factor, in the tradition of Spearman's g, the second a visual-perceptual factor, and the third a numerical-memory factor. Previous researchers have concluded, based on significant correlations between the Full Scale scores for the two tests, that the WJTCA and WISC-R are assessing similar areas of intellectual functioning in learning disabled children. Based on his canonical analysis, however, Estabrook believes that these two tests are measuring different dimensions of cognitive ability in children experiencing learning difficulties. He suggests that unique factor structures is the most likely explanation for the mean score discrepancy between the WISC-R and WJTCA for children with learning difficulties.

Coleman and Harmer (1985) examined WISC-R and WJTCA scores for 54 children in grades one through four who were referred for evaluation but not identified as learning disabled. They obtained a .77 correlation between the Full Scale scores and a significant mean score discrepancy. In their study the mean WISC-R Full Scale IQ was 5.7 points higher than the mean WJ Full Scale BCA score. They concluded that the mean score discrepancy is not a function of grade level. Rather, they suggested that the observed differences may be due to an emphasis on verbal factors in the WJTCA, an area in which children with learning difficulties may have difficulty.

The performances of other special education populations on these two tests have also been examined. Cummings and Sanville (1983) found a .72 correlation between the Full Scale scores for the tests based on scores from 30 students identified as Educable Mentally Retarded who were undergoing a three year re-evaluation. A 16.7 point mean score difference in scores was obtained, with the WISC-R resulting in higher scores. These researchers have suggested that a selection factor operating when the norm samples for the WJTCA were being obtained may have resulted in the omission of children whose parents had lower levels of education. They believe that this could result in the WJTCA providing a lower estimate of a child's intellectual functioning than the WISC-R.

Arffa, Rider, and Cummings (1984) have conducted one of the few studies comparing the WJTCA with an intelligence test other than the WISC-R. They compared the performances of 60 black preschool children on the Stanford-Binet (3rd ed.) and WJTCA. The children obtained a mean IQ of 90.1 on the Stanford-Binet and 94.0 on the WJ Preschool Cognitive Ability score. A one-way repeated measures ANOVA was conducted on these scores along with the WJTCA Skills and Knowledge Cluster scores. Post hoc analysis using Tukey tests indicated that the Knowledge Cluster was significantly lower than the other scores. The authors did not report on the statistical significance of the difference between the Stanford-Binet IQ and the WJTCA Preschool Cognitive Ability score. A correlation of .45 was obtained between the scores. The researchers concluded that differences in the type (handicapped vs non-handicapped, black preschoolers) and age (school-age vs preschool) of their population versus those in other studies could account for the difference in findings.

Phelps, Rosso, and Falasco (1984, 1985) compared the WISC-R and WJTCA with a behavior disordered population. In the 1984 study the sample consisted of 55 adjudicated male adolescents. Twenty-eight of these subjects were later given a secondary diagnosis of learning disabled. A 2.3 point mean score difference, significant at alpha equal .05 but not at alpha equal .01, was obtained between the Full Scale scores for each test. A .84 correlation was also obtained between these scores. These researchers caution that the WJTCA is heavily verbally weighted and could result in an underestimation of the cognitive abilities of behavior disordered youth. In 1985, using an expanded subject pool (n=100 adjudicated male adolescents), these researchers compared WISC-R and WJTCA scores via a multiple regression analysis. They found that the mean WISC-R Full Scale IQ was 3.5 points higher than the mean WJ Full Scale BCA score. Although they did not analyze this mean score difference, it is reasonable to assume that this is a significant difference. A .85 correlation was obtained for these scores. The regression analysis with the WISC-R subtests scores as independent variables and the WJ Full Scale BCA score as the dependent variable indicated that only Verbal subtests were selected for inclusion. These researchers state that performance on the WISC-R Verbal subtests of Information, Similarities, and Arithmetic, subtests which they believe are product dominant or achievement oriented, provides the best estimate of performance on the WJTCA. Based on these findings and beliefs, these researchers concluded that for behavior disordered subjects the WJTCA is a heavily verbally weighted achievement measure, rather than a measure of cognitive ability.

Woodcock's Response to WISC-R/WJTCA Differences

Woodcock (1980,1984), in defense of the Battery, has critically responded to these studies. He believes that there are four areas in which many of these studies can be criticized. He has cited the Ysseldyke, Shinn, McGue, and Epps (1981) study, the Ysseldyke, Shinn, and Epps (1981) study, the Thompson and Brassard (1984) study, and the Reeve, Hall, and Zakreski (1979) study for instrument bias. By this he means that one of the two tests to be compared (usually the WISC-R) has been used to select the sample. He believes that samples in these studies have "spuriously high" WISC-R scores because students are placed in LD classes based on relatively high WISC-R scores coupled with relatively low achievement scores.

A second critical area is what Woodcock calls score bias, which he defines as the practice of using a particular test for both selection purposes and comparison, without re-administration of the test. He states that this type of bias is an exacerbation of instrument bias. The same studies cited for instrument bias, with the exception of the Ysseldyke, Shinn, McGue, and Epps (1981) study, are also cited by Woodcock as exhibiting score bias. He believes that the practice of using WISC-R scores for selection and then for comparison to WJTCA scores results in the use of WISC-R scores with "the most fortuitous combination of true ability and positive error" (1984, p.343) resulting in higher mean WISC-R scores.

The third area discussed by Woodcock is sample bias. By this he means the practice of selecting a sample on other than a random basis from the general population. He states that the selection of such a sample can produce a biased sample that can skew the statistics used based on the degree of correlation between the tests and the sample selection criteria. The higher the correlation between the test and the selection criteria, the further its mean will be from the general mean than will be the mean of a test with a lower correlation to the selection criteria. Since research on the mean score discrepancy issue is centered on special education populations, this problem seems to be inherent in this type of research.

The fourth area identified by Woodcock is an interpretation/reporting bias. He believes that rather than only trying to explain why WJTCA scores are "too low," as has been the case to date, an objective researcher should consider that differences in scores are valid and to be expected or that WISC-R scores are too high. He has also faulted investigators for exaggerating the differences that have been found.

Woodcock has also responded to the hypotheses proposed by others to explain the observed differences between the WJTCA and WISC-R scores. He has dismissed suggestions of norming errors or difficulties in sample selection as not having any empirical support. The achievement content of the WJTCA as an explanation for the score difference has also been disputed by Woodcock. He believes that many authors have confused causation with correlation. While it is true that the WJTCA correlates higher with achievement tests than does the WISC-R, Woodcock believes that this is because the subtests of the WJTCA were deliberately selected "to measure a heterogeneous mix of cognitive skills directly related to school learning" (1984, p.351).

While dismissing these arguments, Woodcock himself looks to the different types of bias he believes to be present in the research to explain the mean score discrepancy. Because of the higher correlation the WJTCA has with school achievement, Woodcock seems to place an emphasis on sample bias as the explanatory mechanism for the mean score discrepancy between the WJTCA and WISC-R when used with special populations. Recently, Sabers (personal communication, 1986) has proposed an additional explanation for the mean score discrepancy between the WJTCA and the WISC-R. He believes that the discrepancy is to be expected given that the tests were normed on different norm groups. Anastasi (1985) has noted that "test norms represent empirical data obtained at a particular time and place, they are subject to the same influences as the test performance of any group" (p.125). Research conducted with other nationally normed tests has indicated that the use of different norm groups can have a great effect on students' obtained scores.

## Differences Due to Norm Groups

Lennon (1954) examined three measures of mental ability in wide use in the early 1950's: the Terman-McNemar Test of Mental Ability, Otis Quick Scoring Mental Ability Tests, and the Pintner General Ability

Tests. His results indicated that while the tests generally gave similar results, differences of two to four points were common in the 90-110 IQ range, and differences of six points were common for scores above and below this range. He suggested that the use of different standardization populations and different test composition could help explain the observed differences.

The Anchor Test Study (ATS; Bianchini & Loret, 1974; Loret, Seder, Bianchini, & Vale, 1975) was a much larger and more psychometrically sound attempt to equate scores on seven widely used standardized reading tests (an eighth test was later included). Generally, differences were limited to two to five point percentile rank discrepancies between the tests, although differences of five to eight points were also found. Large differences were found when comparing the publisher's norms to the more representative ATS norms. Indeed, one of the motivating factors that led to the ATS was a concern about the adequacy of national norms supplied by the test publishers (Linn, 1975). Linn believes that the unsatisfactory rate of cooperation among school districts invited to participate in the norming process has made the representativeness of all national norms questionable.

Baglin (1981) investigated what he labeled a self-selection bias in the norming process of nationally standardized achievement tests. He examined the representativeness of the norm groups for three new editions of nationally standardized achievement test series (the Metropolitan Achievement Tests, 1978; California Achievement Tests, 1977; Iowa Tests of Basic Skills, 1978). The percentage of originally invited districts that ultimately participated in the norming process ranged from a low of approximately 13% to a high of 32%. The likelihood of a school district agreeing to participate in the process was strongly related to that district's prior involvement with the test publishers either in terms of having used a prior edition of the same test, use of a different test series by the same publisher, or because that test publisher's parent company was the major supplier of instructional materials used in the district. He stated that this self-selection resulted in each of the three tests being normed "on a nonrepresentative, nonrandom segment or cross-section of the population. . . The national representativeness claimed is misleading" (p.104). Although these results were based on these three specific test series, Baglin believes that it is not unreasonable to suspect that it is also true of other recent national norming studies. Each test publisher is obtaining a different, not necessarily representative, norming sample. Such differences could help account for discrepancies between tests purportedly measuring similar constructs.

Whenever a new test of intelligence or achievement is introduced or revised, a spate of studies concerned with construct or predictive validity are reported. A sampling of the research reported on the McCarthy Scales of Children's Abilities (MSCA; McCarthy, 1972) or the more recently introduced Kaufman Assessment Battery for Children (K-ABC; Kaufman, 1983) demonstrated that each test, like the WJTCA, exhibits some discrepancy between itself and some other, older tests purported to measure the same construct.

The McCarthy Scales of Children's Abilities

Kaufman (1982), in his review of the research on the MSCA, reports that while there was a substantial correlation between the MSCA General Cognitive Index (GCI) and the Stanford-Binet IQ (r=.82), for nine samples of normal children, there was some discrepancy between the mean scores. The mean Stanford-Binet IQ was 1 to 10 points higher in five of the samples, the MSCA GCI was 1 to 4 points higher for three of the samples, and in one sample the scores were equal. For samples at the extremes of intelligence (one gifted, two retarded) the MSCA resulted in scores 10 to 20 points lower.

Kaufman identified 10 studies with normal children that compared the MSCA to the WISC-R or WPPSI. In five of the studies the MSCA resulted in scores 6 to 8 points lower than the WISC-R, four reported the MSCA to be 1 or 2 points lower, and one study found no difference between the mean IQs. One of the studies with a 1 point difference between the IQs reported a correlation of .51 while the others reported correlations ranging from .71 to .90. Three studies were reported that compared the tests with mentally retarded students as subjects. Correlations ranged from .72 to .82. The WPPSI was found to give scores 8 points higher than the MSCA, one study found the WISC-R to be on average 6 points higher, and the third study found no mean score difference. With samples of learning disabled students, three studies reported mean score differences of 4 or 5 points while three other studies reported differences of 8 to 16 points. All the studies found that the use of the MSCA resulted in lower scores. The correlations for the six studies ranged from .53 to .90. Despite these mean score differences, Kaufman strongly supports the use of the MSCA with children referred for a possible learning disability, with the caution that the examiner keep in mind that the MSCA results in lower estimates of intellectual functioning when used with these children. He states that the GCI "simply correlates too high with the Binet and Wechsler global IQs and includes too many tasks that resemble subtests on these conventional intelligence tests to think of the McCarthy as anything other than a legitimate alternative for assessing mental ability in a psychoeducational battery" (p.153).

The Kaufman Assessment Battery for Children

Because it has been recently introduced, there is not as extensive a body of independent research on the K-ABC as on the MSCA. The data that are available, however, suggest that there is some discrepancy between the K-ABC and the other tests of intelligence. Kaufman and Kaufman (1983) report in the test manual that learning disabled, non-classified, and behaviorally disordered children's mean WISC-R Full Scale IQs are 2 to 4 points higher than the K-ABC Mental Processing Composite (MPC). Naglieri and Haddad (1984) administered the K-ABC and WISC-R to a sample of 33 LD children. The K-ABC MPC and WISC-R Full Scale IQ had a .79 correlation. However, there was a significant 5.2 point mean score difference between the scores, with the K-ABC resulting in a lower estimate of intellectual functioning. These researchers are apparently not disturbed by the significant mean score discrepancy and believe that the K-ABC is a valuable alternative or supplement to the WISC-R. McCallum, Karnes, and Edwards (1984) compared the K-ABC, WISC-R, and Stanford-Binet with a sample of 41 gifted children. They found that the K-ABC MPC resulted in scores 13.1 points lower than the WISC-R Full Scale IQ and 11.7 points lower than the Stanford-Binet IQ. Correlations of .46 and .58 were obtained between the K-ABC MPC and Stanford-Binet IQ and WISC-R Full Scale IQ, respectively. Naglieri and Anderson (1985) also compared the WISC-R and K-ABC with gifted students. With a sample of 38 gifted students they obtained a correlation of .70 and a significant 8 point mean score discrepancy between the tests with the K-ABC resulting in lower scores. They attributed the mean score difference to a ceiling at the upper ages for high functioning children. Naglieri (1985) has also compared the K-ABC and the MSCA with a sample of 51 children. The K-ABC MPC and MSCA GCI had a correlation of .55 and a non-significant mean score difference (GCI=101.3, MPC=102.4).

## Subtypes of Learning Disabilities

Over the last several years, researchers and practitioners have increasingly recognized that children identified as LD are not a homogeneous group. The trend has been toward differentiating subtypes within this heterogeneous population. Adelman and Taylor (1985) surveyed prominent professionals in the learning disabilities field to determine their opinions about the future of the field. Their responses were classified into five categories. Within the area of practice and applied research, over half the respondents discussed the need for developing methods for differentiating subgroups or subtypes within the larger LD population.

## Methods of Subtyping

Two main approaches have been developed by researchers involved in subtyping research. The first method involves clinical inspection of psychometric protocols, and then grouping students into subtypes on the basis of perceived similarity. For example, Breen (1985, 1986) examined the cognitive performances of LD students who were grouped into subtypes on the basis of their WJTA standard scores (mean=100, standard deviation=15) in reading and arithmetic. A set of rules controlled students' placement into one of three groups: those whose math standard score was at least 10 points higher than their reading score, those whose reading score was at least 10 points higher than their math score, and those with equally low math and reading scores as indicated by standard scores of less than 72. These placement rules were chosen to define subtypes that conformed to previous research which identified unique neuropsychological patterns of functioning for similar subtypes of learning disabled children. Although Breen's goal was not solely to identify subgroups, his use of subgroups was a vital element of his study. The clinical approach he used resulted in distinct and easily identifiable subgroups. However, it has been suggested that issues of reliability and validity are more difficult to resolve when using this approach (Fisk & Rourke, 1983; Fletcher & Satz, 1985). Fletcher and Satz state that the clinical approach to subtyping is "easily influenced by a priori biases and assumptions concerning learning disabilities" (p.43) and that if naturally occurring subtypes of learning disabilities exist then classification studies should be able to identify them. Charges of a priori biases and assumptions concerning learning disabilities are valid unless subtypes are defined on the basis of theoretical considerations and without the researcher having first examined the data in order to devise expedient placement rules. Many researchers have begun to look to multivariate statistical techniques to identify subtypes, although this approach is also not without drawbacks.

Q-type factor analysis and cluster analysis are the most commonly used multivariate statistical methods. Lorr (1983) stated that these methods are most appropriately used to identify discrete categories. These procedures reduce the complexity of a data set by sorting entities into a smaller number of homogeneous groups. Adams (1985) has described Q-type factor analysis as

> the reduction of a correlated set of observations (between persons) to some optimal smaller set of linear composites that preserves the underlying information. . . The analysis loads for factors on the basis of correlations between persons over variables. The resulting factors are linear representations in spatial form, and are the basis for identifying types (p.32).

Correlation coefficients for all pairs of subjects are included in the data matrix. Fleiss and Zubin (1969) have identified several difficulties with using a between-persons correlation as a measure of profile similarity.

Among the theoretical difficulties are that it does not always make sense to talk about correlations between people and to assume that a linear model applies to the data matrix of people. Among the technical difficulties are that the maximum number of subtypes that can be identified is always one less than the number of variables, which may artificially limit the number of subgroups identified if the researcher has reduced the number of variables to a reliable minimum, and that the data matrix is difficult to standardize.

Cluster analysis is described by Adams (1985) as a process where data are analyzed and evaluated on the basis of similarity. The object is "to determine an optimal solution in which clusters contain persons whose attributes are more like one another's than they are like those of persons in other clusters" (p.33). While the Q analysis relies on correlations as the measure of similarity, cluster analysis uses standardized Euclidean distance between the data points as the measure of similarity. The use of distance imparts several properties (a value of 0 is obtained only if two profiles are identical and it is possible to speak of one profile as being twice or three times as far from a standard as is a second profile) which are not shared by Q correlations (Fleiss & Zubin, 1969). In addition, Q correlations are sensitive only to the shape of profiles whereas Euclidean distance measures contain information about relative profile elevation, shape, and scatter (Blashfield, 1980). Fleiss and Zubin have identified standardization of each variable and correlations among the variables as two major difficulties when using distance as a measure of similarity in cluster analysis. Various methods, such as single linkage, complete linkage, and average linkage (Blashfield & Aldenfelder, 1978) have been developed for clustering individuals using the distance measure.

These two approaches have been widely accepted as objective and efficient methods for subtype identification by those working within the learning disabilities subtype field (Blashfield & Aldenfelder, 1978). Before reviewing many of these studies, it is important to note that there is disagreement about the appropriate use of clustering procedures (Fleiss & Zubin, 1969; Fletcher & Satz, 1985; Blashfield, 1980). Major criticisms about the use of these procedures to identify subtypes lie in the fact that they "proceed from no mathematical or statistical model whatever. Aside from some vague ideas about seeking to establish homogeneous groups, there really is no statement of what one is looking for" (Fleiss & Zubin, 1969, p.245). In addition, they state that the mechanical process of clustering is likely to turn up clusters regardless of the data. Once identified, it is not difficult to develop sensible explanations and names for them. Much effort would then be wasted in attempting to validate clusters which in actuality were derived from a homogeneous sample. A means of testing the goodness of fit of a homogeneous distribution of one sort or another to the original sample, given only the data from the sample, is not available within the methodologies of cluster analysis.

Neuropsychological Approach

Fletcher and Satz (1985) have reported on studies from the Florida Longitudinal Project based on data from samples of children from 1970 to 1978. A sample of 236 children was administered the Arithmetic, Reading, and Spelling subtests of the WRAT (Jestak & Jestak, 1965). Nine separate subgroups emerged from a cluster analysis of the data, representing varying patterns of performance across the three subtests. Two of these groups, representing relatively poor, flat performance across all three subtests, accounted for 89 of the subjects. These students were considered disabled learners and so were administered several additional measures examining neuropsychological variables. These scores were then subjected to another cluster analysis. Five subtypes ranging from those low in general verbal abilities to those low in visual-spatial abilities to those showing no impairment at all were identified. Although the uniqueness of all of these "processing deficiency subtypes" has not been adequately demonstrated, the authors' report that similar subtypes were obtained in an independent cross-cultural study (van der Vlugt & Satz, 1985) and with independent samples of children in Florida. Additional studies to demonstrate the external validity of these subtypes are being conducted.

Other researchers have examined particular academic disabilities (e.g., reading, arithmetic, written language) in order to determine if subtypes exist within each of these areas. In a series of studies Doehring and his colleagues (Doehring, 1968; Doehring & Hoshko, 1977; Doehring, Hoshko, & Bryans, 1979; Doehring, Trites, Patel, & Fiedorowicz, 1981) examined reading disabilities in samples of children to determine if they could identify subtypes of reading disability. Based on Q-type factor analysis of 31 reading measures, three distinct subtypes were consistently found. The largest subtype consisted of readers who had great difficulty in oral reading relative to their silent reading skills. The second subtype had difficulty associating printed and spoken words, while members of the third subtype had difficulty silently and orally reading letter sequences relative to their ability to read single letters. However, there was much overlap between the subtypes with over a quarter of the subjects loading high on two or more of the factors that defined the subtypes. In addition, these researchers were not able to establish a relationship between reading disability subtype and particular patterns of performance on language and neuropsychological measures.

Petrauskus and Rourke (1979) examined a sample of 160 seven and eight year old children of whom 133 were reading disabled and 27 were normal readers (based on WRAT percentile scores of less than 25 or greater than 45). The subjects were split into two subsamples of 80 with a similar number of normal readers in each group. They were then administered 20 separate measures tapping various skills. Three distinct subtypes of reading disabled children were identified via a Q-type factor analysis of the data. A fourth subtype, consisting principally of normal readers, also emerged. The largest subtype consisted of children with psycholinguistic deficiencies. Deficits in sequential ability characterized children in the second subtype. Members of the third subtype had difficulties with motor tasks.

The area of arithmetic has also been examined to determine if specific subtypes can be identified. Rourke and his colleagues (Rourke & Finlayson, 1978; Rourke & Strang, 1978; Strang & Rourke, 1985) have conducted a series of studies examining this issue. Three groups of students were selected on the basis of the patterns of their scores on the Reading, Spelling, and Arithmetic subtests of the WRAT. Each group was composed of 15 students with a Full Scale IQ of between 86 and 114. Group 1 consisted of students uniformly deficient in all three areas, Group 2 members' arithmetic subtest scores were higher than their reading and spelling scores, and Group 3 consisted of students whose reading and spelling were higher than their arithmetic scores. The arithmetic scores for all groups were approximately the same. These three groups were then assessed on a variety of neuropsychological measures. Groups 2 and 3 differed significantly on the neuropsychological measures, suggesting that their arithmetic deficiency may be a reflection of very different patterns of central processing abilities and deficiencies. Those children with better developed reading and spelling skills (Group 3) had well-developed auditory and perceptual skills and verbal skills (abilities assumed to be subserved primarily by the left cerebral hemisphere) but deficient visual-perceptual-organizational skills (right hemisphere abilities). Group 2 children, those whose math skills were better developed than their spelling and reading skills, had a pattern on the neuropsychological measures exactly opposite that of Group 3. A review of the types of arithmetic errors made by each group indicated that Group 3 children tended to make more mechanical errors and errors due to deficient math judgment and reasoning.

Subtypes of spelling disability have also been identified. Sweeney & Rourke (1978) examined three groups of spellers (normal spellers, phonetically accurate spelling retardates, and phonetically inaccurate spelling retardates) at two age levels (grades 4 and 8). There were 8 children in each group. A variety of different measures were used to assess neuropsychological variables. The results indicated that the older phonetically inaccurate spellers were deficient in basic receptive linguistic operations, while older phonetically accurate spellers had difficulty in associating the spoken word with an analysis of visual-spatial information. Differences in these areas were not noted for the younger age groups.

A different approach to subtyping LD students involves using the behavior of these students as the classifying variable. Over the years, many studies have examined the interpersonal environments of LD children. Only recently, however, have attempts been made to determine whether distinct subtypes of LD children could be identified on the basis of their socio-emotional functioning.

## Social Approach

McKinney (cited in McKinney & Feagans, 1983) conducted one of the early studies in this area. A cluster analysis of data from the Classroom Behavior Inventory, a teacher rating instrument, and observational data of behavior using the SCAN observational system resulted in four distinct subtypes. The first subtype was marked by deficiencies in task orientation and independence, but by strengths in social adjustments in comparison to other subtypes and average achievers. The second subtype consisted of students who were the most severely impaired across all the achievement and behavioral scales. The third subtype were those low on task orientation but high on extroversion and hostility. The fourth subtype was more academically impaired than either subtypes 1 or 3, but showed no behavioral deficiencies in comparison to average achievers. Although firm conclusions could not be drawn, the results of this study indicate that LD children can be subtyped on the basis of behavior. Although all the students were academically deficient, only some were behaviorally inappropriate and inappropriate in different ways.

Strawser and Weller (1985) have also investigated the feasibility of subtyping LD students on the basis of behavior. It is their belief that the use of adaptive behavior measures, in conjunction with traditional measures of intelligence, academic achievement, and language processing, results in a more accurate description of LD subtypes or severity than can be obtained by using ability/performance discrepancies alone. In an effort to investigate this hypothesis, these researchers gathered ability, achievement, language processing, and adaptive behavior data for 74 LD students who received assistance on a resource basis and 38 students who were in self-contained LD classrooms. They conducted a cluster analysis on these data and identified three subgroups. For each group they then developed a continuum of ability/achievement discrepancy in two areas: number and amount. The number of discrepancies was determined by counting the number of instances in which there was a discrepancy between a subject's measure of ability and achievement in academics or language processing. Frequency counts were obtained for discrepancies of between one and two standard deviations, and greater than two standard deviations. The authors state that a similar procedure was used to determine the amount of discrepancy, but it is not at all clear how this analysis was conducted or what the results mean. These researchers then computed correlation coefficients between subjects' adaptive behavior measures (their own Weller-Strawser Scales of Adaptive Behavior) and number and amount of discrepancies. Students in Group 1, described as mildly learning disabled (n=43), had average intellectual ability, mild to moderate levels of adaptive behavior, and discrepancies of between one and two standard deviations in academic areas only. Group 2 students, described as severely learning disabled (n=37), were also of average intelligence, but had a greater number of discrepancies greater than two standard deviations in both academic and language processing areas. Their adaptive behavior measures were in the moderate to severe range. They described Group 3 (n=32) as probably not learning disabled because of borderline to low average intelligence and few discrepancies in academic or processing areas. For Group 1 the adaptive behavior scores correlated poorly with the number and amount of academic discrepancies and moderately (r=.38) for the process discrepancies. For Group 2 the adaptive behavior scores correlated moderately with the academic discrepancies beyond two standard deviations (r=.30) and highly for the process discrepancies (r=.76). Based on these results, these authors concluded that LD subtypes based on the severity of the learning disability are more appropriately determined by measures of adaptive behavior than by discrepancy in performance in academic or language processing areas.

It is not clear from reading this report, however, how Weller and Strawser reached this conclusion. Their data analysis never addressed their original hypothesis--does the use of adaptive behavior scores afford a more accurate determination of the severity of a learning disability. Their analyses did not compare subgroups formed on the basis of discrepancies between ability and achievement in academic or language processing areas with subgroups formed with the added component of adaptive behavior scores. Instead, they relied on poorly defined, complex statistical procedures of questionable appropriateness. For example, it is not clear how these authors calculated what they called amount of discrepancy. Also, it is not clear what advantage there was in conducting a cluster analysis when they then had to use the same data to determine discrepancies and correlations. The groups had to differ on these measures since it was the same data that determined the groups originally. Their computation of the cluster analysis itself, even if it was an appropriate form of analysis, is questionable. They did not account for how the variables were standardized or for the correlations between the variables, two areas identified by Fleiss and Zubin (1969) as important considerations when conducting a cluster analysis. They also did not report any attempts to validate their resulting subtypes even though it is well known that any clustering procedure will produce clusters, regardless of the nature of the data (Blashfield, 1980). In sum, this study seems far too flawed in its statistical analyses to be able to adequately address the role of adaptive behavior in identifying subtypes of learning disabilities.

Porter and Rourke (1985) examined the personality profiles of LD students using the Personality Inventory for Children, a 600 item true-false battery concerning a child's socio-emotional functioning. The mothers of the 100 identified LD students in this sample completed the battery. A Q-type factor analysis was conducted on the scores from the 15 scales chosen that were assumed to optimally reflect the major personality dimensions assessed by the inventory. Four subtypes were identified, accounting for 80 of the 100 subjects. The first subtype, representing 44% of those assigned to subtypes, had a profile indicating a balanced and well adjusted personality. The second subtype, accounting for 26% of those assigned to subtypes, had a profile marked by low self-esteem, high anxiety, and poor interpersonal functioning. Subtype 3, with 13% of the assigned subjects, had elevated somatic concerns. The fourth subtype, consisting of 17% of those assigned, had a profile marked by hyperactivity, impulsiveness, low frustration tolerance, and resistance to authority. Porter and Rourke concluded that the learning disabled do not constitute a single personality type. Rather, it is a heterogeneous group with diverse patterns of functioning.

## WISC-R and Identification of LD Students

Much research has been conducted using the WISC-R to identify LD students. Much of the early work focused on identifying a WISC-R profile unique to LD children which would be useful in differential diagnosis. Numerous factor analytic studies have also been conducted. Kavale and Forness (1984) conducted a meta-analysis of 94 studies that examined the validity of WISC-R scale profiles and recategorizations. They first examined differences between Verbal IQs and Performance IQs. Traditionally, it has been believed that LD students exhibit a significant discrepancy between these two scores with the Performance IQ being higher. Their analysis indicated that the Performance IQs were 3.46 points higher, well short of the 11 point difference needed for statistical significance. They also estimated that 79% of the normal population is likely to exhibit the same or greater discrepancy.

A subtest pattern and scatter analysis of the WISC-R was also conducted. All the subtest scores fell within one standard deviation of the mean, with the average LD student performing more poorly on the Verbal subtests than the average normal student, and slightly better than the normal group on the Performance subtests. It has been widely believed that LD students do especially poorly on the Arithmetic, Coding, Information, and Digit Span subtests. The findings in this study support the existence of this so-called ACID pattern.

The belief that LD students exhibit greater variability across the subtests was not supported. Examination of scaled score ranges, scatter among the individual subtests and comparisons of individual subtest's scores versus average Full Scale, Verbal, and Performance IQ did not reveal any meaningful patterns. On average, the LD students exhibited less variability than the standardization sample resulting in no patterns of diagnostic value.

Over the years numerous factor analyses of the WISC-R have been conducted. Wechsler's Verbal and Performance factors have been consistently recovered for a variety of groups, ranging from blacks and whites (Reschly, 1978), to boys and girls (Reynolds & Gutkin, 1980), and the behaviorally disordered (Dehorn & Klinge, 1978). Most researchers have also recovered a third factor usually labeled Freedom from Distractibility, typically loaded on by Arithmetic, Digit Span, and Coding. Bannatyne (1968, 1974) has proposed two recategorizations of the WISC-R. The first is composed of Spatial, Conceptual, and Sequential factors, with the second approach consisting of the first three factors plus one labelled Acquired Knowledge. Bannatyne (1968) reported that his LD subjects had a Spatial greater than Conceptual greater than Sequential pattern. This pattern, as well as the Bannatyne II scheme of Spatial greater than Conceptual greater than Sequential greater than Acquired Knowledge, was also characteristic of the LD students in the Kavale and Forness meta-analysis. The differences between the factor scores were not statistically significant. However, other patterns based on various patterns or recategorizations were also supported (Keogh & Hall, 1974; Kaufman, 1975; Blaha & Vance, 1979), but many of the factor scores within each pattern were not significantly different.

On the basis of their meta-analysis examining the validity of WISC-R profiles and recategorizations for the differential diagnosis of learning disabilities, Kavale and Forness (1984) concluded that "regardless of the manner in which WISC-R subtests were grouped and regrouped, no recategorization profile, pattern, or factor cluster emerged as a 'clinically' significant indicator of LD" (p.150). Kaufman

(1981), in his analysis of the use of the WISC-R in assessing learning disabilities, concluded that examining regroupings such as the Bannatyne model does not contribute new knowledge to the field, and seems to have no future in the differential diagnosis of learning disabilities when used with heterogeneous groups of LD children. He believes that the key is to investigate LD populations that are defined rather homogeneously.

Snow, Cohen, and Holliman (1985) used a cluster analysis of the WISC-R to define particular subgroups from a larger LD population. The subtest scores of 106 LD students were factor analyzed, resulting in three distinct factors labelled Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility. These factor scores were then used in a group cluster analysis. This procedure resulted in the identification of six subgroups. WRAT subtest scores in Spelling, Arithmetic, and Reading were also available allowing the researchers to determine patterns of intellectual abilities and academic achievement for the six subgroups. Subgroup 2 had uniformly low achievement scores with Comprehension abilities lower than other intellectual abilities. This subgroup was characterized as a language deficient group. Subgroup 3 was characterized as representing children with an attention deficit. Subgroup 4 had patterns similar to Group 2, but there was not as large a difference between Verbal Comprehension and Perceptual Organization abilities. The other three subgroups did not exhibit clear patterns of specific learning disabilities. Subgroup 1, because it was uniformly low in all areas of achievement and intellectual abilities, seemed to represent a borderline IQ, slow-learner group rather than an LD subtype.

Subgroups 5 and 6 were characterized by patterns of low average to average abilities.

The significance of the Snow et al. study is two-fold. First, the study indicates that subgroups of LD children can be formulated on the basis of WISC-R performance. The ability to define relatively homogeneous subgroups of LD children is recognized as a key ingredient in future learning disabilities research. Second, the results of this study, in conjunction with previous research, illustrate that there is some consistency in the types of subgroups identified (e.g., phonetic language deficiencies but intact visual-spatial skills and vice versa).

## Summary

In summary, research to date has raised the issue of a mean score discrepancy between the WISC-R and WJTCA. However, it has not been able to clarify for the practitioner whether the WJ is a useful tool for the diagnosis of a learning disability or whether it is inappropriate to use in this context. Research examining this issue was reviewed in this chapter. The hypothesis that differences between the tests are expected and due, at least in part, to norm group differences was also examined. This chapter also reviewed research conducted on identifying learning disability subtypes and the neuropsychological, academic, and behavioral differences found to exist between the identified subtypes.

CHAPTER 3

# METHODS

## Subjects

Subjects for this study were selected from a school district in Tucson, Arizona. All students who met the following criteria were included in the study:

1) The student had been referred for a special education evaluation due to academic difficulties within the 1984-85 school year or was currently receiving special education services as an LD student and had undergone a three year re-evaluation during the 1984-85 school year.

2) As a routine part of the evaluation, the student had been administered a complete WISC-R and Parts I and II of the Woodcock-Johnson Psycho-Educational Battery during the 1984-85 school year. The WISC-R and Part 1 of the WJ were administered by a certified school psychologist. The WJTA was administered by a certified school psychologist or special education teacher certified in the area of learning disabilities who was trained in its administration.

This procedure resulted in the identification of 161 students, 114 male and 47 female. They ranged in age from 6-1 to 17-5 years with a mean age of 10-11 years and standard deviation of 2-9 years. Of the total sample, 139 were identified as Anglo, 19 as Hispanic, and 3 as some other ethnicity. There were 72 students who had been receiving special education services because of learning disabilities re-evaluated during the 1984-85 school year (Group 1). All continued to be identified as learning disabled. There were 49 students who were newly referred and subsequently identified as learning disabled (Group 2), and 40 students who were newly referred but not identified as eligible for special education services (Group 3). Table 1 lists the distribution of students within each group by sex and grade.

## Procedures

Data for this study was gathered via a review of files of the 161 students identified by school personnel as having met the above mentioned criteria. The file review was conducted by the investigator. The following information from each file was recorded on a data sheet:

1) Scaled subtest scores and Verbal, Performance, and Full Scale IQs on the WISC-R and raw subtest scores from the WJ, Parts I and II

2) Test administrator

3) Date of test administration

4) Age of the student at the earlier of the test administrations

5) Grade of the student at the earlier of the test administrations

6) Sex of the student

7) Ethnicity of the student

8) Special education services received, if any.

The investigator used the recorded raw score of each subtest of the WJ to derive various cluster scores, standard scores, age and grade

Table 1

Distribution of Newly Referred and Re-Evaluated LD Students and Newly Referred Non-LD Students by Sex and Grade

| | | Sex | | Grade | | | |
|---|---|---|---|---|---|---|---|
| Group | n | male | female | k-3 | 4-6 | 7-9 | 10-12 |
| 1. Re-Eval LD | 72 | 51 | 21 | 7 | 30 | 27 | 8 |
| 2. Ref LD | 49 | 35 | 14 | 24 | 17 | 7 | 1 |
| 3. Ref Non-LD | 40 | 28 | 12 | 25 | 11 | 3 | 1 |
| Total | 161 | 114 | 47 | 56 | 58 | 37 | 10 |

scores, percentiles, Relative Performance Indices, and functioninglevels available on the WJ. These scores were obtained by using the Compuscore Scoring System for the WJ, a computer software program developed by Jay Hauger (1984). Use of this program ensured that the same test information was available for each subject and helped avoid the errors that occur when hand-scoring the WJ.

## Data Analysis

The test scores and the demographic information for each student were transferred to data sheets and re-checked for accuracy. The data for each case were entered in a free-field format on a diskette using the Word Perfect word processing program developed for an AT&T personal computer. A printout of the data entered for each case was obtained and checked against the raw data. In order to address the research questions concerned with the WISC-R/WJTCA mean score discrepancy issue, the following statistical analyses were performed using the Statistical Package for the Social Sciences computer software program (Nie, et al., 1975) available for use with the AT&T personal computer.

1) A Kolmogorov-Smirnoff test (K-S; Siegel, 1956) was conducted to determine if the distributions for the Full Scale scores from each test were normally distributed. The K-S is a non-parametric test which allows the user to test whether the observed data could reasonably have come from a theoretical distribution specified by the user. In this case each distribution was tested against a normal distribution using the parameters (mean and standard deviation) calculated from the sample and against a normal distribution using the test parameters (mean=100,

std. dev.=15). In addition, the distribution of the WISC-R Full Scale IQ scores was compared to the distribution of the WJ Full Scale BCA scores for each group to determine if they were comparable. This analysis addressed research question #1.

2) In order to determine if there was a significant difference between the mean Full Scale scores for each group (research question #1) t-test comparisons for paired observations were performed on the WISC-R and WJ Full Scale scores for each group.

3) Pearson correlations (Hayes, 1973) between the WISC-R IQs (Verbal, Performance, and Full Scale) and the WJ BCA scores were determined for each of the three groups in order to help determine if the tests were measuring similar constructs for these groups (research question #2).

4) In order to address the research questions pertaining to subtypes of LD students (research questions 3-6) it was necessary to determine a procedure to identify subtypes of LD students from the sample of LD students in this study (Groups 1 and 2). As previously discussed, there are advantages and disadvantages to both clinical and statistical procedures used to identify LD subtypes. Those studies that have relied on cluster or Q-type analysis have typically employed a great number of neuropsychological variables, which is not the nature of this study. In these types of analyses it is difficult to account for correlations between the variables used in the subtyping. When dealing with a large number of varied neuropsychological measures this is less of a concern, but it is a much larger problem when dealing with

academic achievement scores which would be expected to be at least moderately correlated. For this study, then, a clinical approach to LD subtyping was deemed most appropriate.

Guided by previous research using the clinical approach with academic achievement scores (Breen, 1986; Rourke & Finlayson, 1978; Rourke & Strang, 1978; Strang & Rourke, 1983; Sweeney & Rourke, 1978), the following placement rules were developed. The identification of subtypes of LD students on the basis of WJTA scores addresses research question #3.

Subtype 1: Students with standard scores greater than or equal to 90 on the Reading, Math, and Written Language Clusters on the WJTA. These students are characterized by relatively high, flat achievement profiles.

Subtype 2: Students with standard scores of 80-89 on the Reading, Math, and Written Language Clusters on the WJTA. These students could be described as having flat achievement profiles of a moderate level.

Subtype 3: Students with standard scores 79 and below on the Reading, Math, and Written Language Clusters on the WJTA. They could be described as having relatively low, flat achievement profiles.

Subtype 4: Students with standard scores on the Reading Cluster of the WJTA that were at least 9 points greater than their standard scores on the Math Cluster of the WJTA.

Subtype 5: Students with standard scores on the Math Cluster of the WJTA that were at least 9 points greater than their standard scores on the Reading Cluster of the WJTA.

Other: Those students who could not be placed in one of the subtypes.

The identified LD students in this study (Groups 1 and 2) were identified as one of these subtypes based on an examination of their WJTA Achievement Cluster scores. After all the students had been placed, those students who could not be placed (Other) were re-examined. If the student was not placed in one of the subtypes because one of the cluster scores violated one of the placement rules by one standard score point (e.g., student A has standard scores of 84 on the Reading Cluster, 82 on the Math Cluster, 79 on the Written Language Cluster) the student was assigned to the subtype to which he or she would have otherwise been assigned. In this way, as many students as possible were assigned to the appropriate subtype rather than lose subjects because of an inconsequential one standard score point difference. In order to determine if the students' WJTA Cluster scores could be effective at classifying LD students into subtypes, a step-wise discriminate function analysis was conducted. The WJTA Cluster scores were used as the discriminating variables. Comparisons of the predicted group memberships derived from the discriminant functions with the actual group memberships based on the established set of clinical rules provided information about how effective the discriminating variables were. A large number of misclassifications would indicate that the established clinical rules were not effective in establishing unique subtypes. The following statistical analyses were conducted using the subtype framework.

5) In order to determine if there was a significant difference between the Full Scale scores, t-test comparisons for paired observations were performed on the WISC-R and WJTCA Full Scale scores for each subtype. This analysis addressed research question #4.

6) Pearson correlations (Hayes, 1973) between the WISC-R IQs (Verbal, Performance, Full Scale) and the WJ BCA scores were determined for each of the identified subtypes. This analysis addressed research question #5.

7) A series of one way ANOVAs were conducted on each Cognitive Cluster score for the subjects across each subtype. Scheffe post hoc comparisons were conducted when indicated. This analysis helped determine if the identified LD subtypes demonstrated different patterns of performance across the Cognitive Clusters of the WJTCA. In addition, a discriminate analysis was conducted to determine if the students could be correctly classified into the appropriate subtypes on the basis of their Cognitive Cluster scores. These analyses addressed research question #6.

## CHAPTER 4

## RESULTS

The purpose of this chapter is to present the results of the analyses described in the previous chapter. Specifically, the following topics are included: 1) K-S tests analyzing the distributions of the Full Scale scores from the WISC-R and WJTCA, 2) t-tests for paired observations examining the mean score difference between the WJTCA and WISC-R Full Scale scores for the re-evaluated LD group, newly referred LD group, and newly referred non-LD group, 3) Pearson correlations between the WISC-R IQ and the WJ BCA scores for each group and for the entire sample, 4) a discriminant analysis examining the utility of the WJTA achievement cluster scores to correctly classify the LD students from Groups 1 and 2 into the LD subtypes identified on the basis of clinical rules detailed in chapter 3, 5) t-tests examining the mean differences for the WJTCA and WISC-R Full Scale scores for each subtype, 6) Pearson correlations between the WISC-R IQ and WJ BCA scores for each subtype, 7) one way ANOVAs examining the mean Cognitive Cluster scores from the WJTCA for each subtype. A discriminant analysis conducted on these data will also be presented.

### WISC-R IQ and WJ BCA Score Distribution

A series of K-S tests were conducted to examine the distribution of the WJTCA and WISC-R Full Scale scores for each group and for the entire sample. In each case, the distributions for both tests' Full Scale scores were tested against a normal distribution having the mean and standard deviation of the particular sample, and against a normal distribution having a mean of 100 and standard deviation of 15 (the published test parameters for the two tests). In addition, the distributions of the two scores for the groups and the entire sample were compared. Table 2 lists the results of these tests.

An examination of this table indicates that for the entire sample, and for each group separately, one cannot reject the hypothesis that both the WISC-R and WJTCA Full Scale scores have come from a population having a normal distribution at alpha equal to .01. The hypothesis that each group and the sample as a whole have come from a population having a normal distribution with a mean of 100 and standard deviation of 15 can be rejected at an alpha level of less than .001. The hypothesis that the distribution of the tests' Full Scale scores have come from a population having the same distribution can be rejected for the entire sample and for Groups 1 and 2 at an alpha level of less than .01. This hypothesis cannot be rejected for the students in Group 3.

In summary, these results indicate that for the entire sample, and for each group separately, the WISC-R and WJTCA can be considered to be normally distributed. Only the distribution of the scores from the two tests for students in Group 3, those referred for an evaluation but not identified as learning disabled, could be considered to have come

Table 2

K-S Test Results for Re-Evaluated and Newly Referred LD Students and Newly Referred Non-LD Students

| Data | Group | n | M | SD | Test Distrib | Test Stat | P Val |
|---|---|---|---|---|---|---|---|
| WISC-R | 1,2,3 | 161 | 93.2 | 12.4 | normal | K-S Z=1.06 | .208 |
| WJ | 1,2,3 | 161 | 87.4 | 12.0 | normal | K-S Z=1.43 | .033 |
| WISC-R | 1,2,3 | 161 | 93.2 | 12.4 | norm,m=100,sd=15 | K-S Z=3.70 | .01 |
| WJ | 1,2,3 | 161 | 87.4 | 12.0 | norm,m=100,sd=15 | K-S Z=5.66 | .01 |
| WISC-R,WJ | 1,2,3 | 161 | | | WISC-R vs WJ | D=.268 | .01 |
| WISC-R | 1 | 72 | 88.7 | 10.8 | normal | K-S Z=.639 | .809 |
| WJ | 1 | 72 | 81.5 | 9.1 | normal | K-S Z=.750 | .627 |
| WISC-R | 1 | 72 | 88.7 | 10.8 | norm,m=100,sd=15 | K-S Z=3.31 | .01 |
| WJ | 1 | 72 | 81.5 | 9.1 | norm,m=100,sd=15 | K-S Z=5.57 | .01 |
| WISC-R,WJ | 1 | 72 | | | WISC-R vs WJ | D=.250 | .01 |
| WISC-R | 2 | 49 | 93.8 | 10.4 | normal | K-S Z=.703 | .707 |
| WJ | 2 | 49 | 88.0 | 9.9 | normal | K-S Z=1.04 | .234 |
| WISC-R | 2 | 49 | 93.8 | 10.4 | norm,m=100,sd=15 | K-S Z=2.27 | .01 |
| WJ | 2 | 49 | 88.0 | 9.9 | norm,m=100,sd=15 | K-S Z=3.08 | .01 |
| WISC-R,WJ | 2 | 49 | | | WISC-R vs WJ | D=.245 | .01 |
| WISC-R | 3 | 40 | 100.9 | 13.5 | normal | K-S Z=.739 | .646 |
| WJ | 3 | 40 | 97.1 | 12.5 | normal | K-S Z=.534 | .938 |
| WISC-R | 3 | 40 | 100.9 | 13.5 | norm,m=100,sd=15 | K-S Z=.613 | .846 |
| WJ | 3 | 40 | 97.1 | 12.5 | norm,m=100,sd=15 | K-S Z=1.08 | .194 |
| WISC-R,WJ | 3 | 40 | | | WISC-R vs WJ | D=.175 | .175 |

Note: Group 1 are re-evaluated LD students, Group 2 are newly referred LD students and Group 3 are newly referred Non-LD students.

from a population with the same parameters as the tests' norms (mean=100, standard deviation=15). A direct comparison between the score distributions of each test for each group and for the entire sample, indicates that again only the scores for students in Group 3 could be considered to have come from the same population.

WISC-R and WJTCA Mean Score Comparison

Three t-test comparisons for paired observations were performed on the WISC-R and WJTCA Full Scale scores for each group. Table 3 lists the results of these tests. For all three groups there was a significant difference between the two scores. Students in Group 1, those previously identified as learning disabled and undergoing a three year re-evaluation, showed both the lowest mean scores and the largest difference between the tests (7.2 standard score points). While these students had mean WISC-R scores just below the average range of cognitive ability, their average WJTCA score was just above the borderline range of cognitive ability. Students who were newly referred and subsequently identified as learning disabled (Group 2) overall had higher scores than Group 1 students and showed a 5.7 standard score point difference between the tests. Students who were referred for an evaluation but not identified as learning disabled (Group 3) had the highest scores and the smallest difference.

In order to gain additional information about the differences between the tests, individual performances were examined. The results of this analysis are presented in Table 4. The two most important

Table 3

Mean Difference Between WISC-R and WJTCA Full Scale Scores for Re-Evaluated and Newly Referred LD Students and Newly Referred Non-LD Students

| Group | n | Test | Mean | SD | Mean Diff | t-val | Sig |
|---|---|---|---|---|---|---|---|
| 1 | 72 | WISC-R | 88.7 | 10.8 | | | |
| | | | | | 7.2 | 8.30 | .01 |
| | | WJTCA | 81.5 | 9.1 | | | |
| 2 | 49 | WISC-R | 93.8 | 10.4 | | | |
| | | | | | 5.8 | 5.52 | .01 |
| | | WJTCA | 88.0 | 9.9 | | | |
| 3 | 40 | WISC-R | 100.9 | 13.5 | | | |
| | | | | | 3.8 | 3.38 | .01 |
| | | WJTCA | 97.1 | 12.5 | | | |

Note. Group 1 are re-evaluated LD students, Group 2 are newly referred LD students and Group 3 are newly referred non-LD students.

Table 4

Distribution of WISC-R/WJTCA Full Scale Score Differences for Re-Evaluated and Newly Referred LD Students and Newly Referred Non-LD Students

| | WISC-R greater then WJTCA | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1-5pts. | | 6-10pts. | | 11-15pts. | | gt 15pts. | | Total[b] | |
| Group[a] | n | % | n | % | n | % | n | % | n | % |
| 1 | 13 | 18 | 10 | 14 | 21 | 29 | 10 | 14 | 54 | 75 |
| 2 | 13 | 27 | 14 | 29 | 9 | 18 | 3 | 6 | 39 | 80 |
| 3 | 14 | 35 | 10 | 25 | 5 | 13 | 1 | 2 | 30 | 75 |

| | WISC-R less than WJTCA | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1-5pts. | | 6-10pts. | | 11-15pts. | | gt 15pts. | | Total | |
| Group | n | % | n | % | n | % | n | % | n | % |
| 1 | 7 | 10 | 2 | 3 | 0 | 0 | 0 | 0 | 9 | 13 |
| 2 | 4 | 8 | 3 | 6 | 1 | 2 | 0 | 0 | 8 | 16 |
| 3 | 4 | 10 | 1 | 2 | 1 | 2 | 1 | 2 | 7 | 16 |

Note. Group 1 are re-evaluated LD students, Group 2 are newly referred LD students, and Group 3 are newly referred non-LD students.

[a]) There were 72 students in Group 1, 49 in Group 2, and 40 in Group 3.

[b]) There were 9 students (13%) in Group 1, 2 students (4%) in Group 2, and 3 students (8%) in Group 3 who had identical WISC-R and WJTCA Full Scale Scores.

observations are that under half the subjects (43%, n=69) had WISC-R and WJTCA scores that were within 5 points of each other and that over three-fourths of the subjects (77%, n=123) had WISC-R scores that were higher than their WJTCA scores. The percentage of subjects in each group with higher WISC-R than WJTCA scores is relatively even with 75% (n=54) of the students in Group 1, 80% (n=39) in Group 2, and 75% (n=30) of the students in Group 3 following this pattern. However, the distribution of the magnitude of the WISC-R/WJTCA difference varies considerably according to group membership. Nearly half (43%, n=31) the students in Group 1 had WISC-R scores that were more than 10 points higher than their WJTCA scores. Only 24% (n=12) of the students in Group 2 and 15% (n=6) of the students in Group 3 had differences of such magnitude. On the other side of the table, of the 14.5% (n=24) of the students with WJTCA scores higher than their WISC-R scores only 3 (1.8% of the entire sample) had a WJTCA score that was more than 10 points higher than their WISC-R score. This compares with 49 students (31% of the sample) with WISC-R scores more than 10 points higher than their WJTCA score.

## Correlation Between WISC-R and WJTCA Scores

Table 5 lists the correlations between the WISC-R Verbal, Performance, and Full Scale IQs and the WJ Full Scale BCA score for the entire sample. The two Full Scale scores have a .81 correlation. The WISC-R Verbal IQ has an equally strong .81 correlation with the WJTCA score while the WISC-R Performance IQ has a smaller .45 correlation with the WJTCA score.

Tables 6 through 8 list the correlation matrices for the WISC-R scores and the WJTCA scores separately for each group. As would be expected given the results for the entire sample, the Full Scale scores from each test are strongly correlated in each group (.72, .74, and .85 for Groups 1, 2, and 3, respectively). The correlations between the various scores within each group are as expected, except the WISC-R Performance IQ for students in Group 2. In general, the WISC-R Performance IQ has the smallest correlation in the other groups also, but none as small as the .12, .21, and .39 correlations between the Performance IQ and WJ BCA score, Verbal IQ, and Full Scale IQ, respectively, evident in Group 2.

## Identification of LD Subtypes

Using the procedures outlined in Chapter 3, five subtypes of learning disabled students were identified from the sample of learning disabled students included in this study. Of the 121 students identified as learning disabled, 113 were classified into one of the five subtypes. Six could not be classified and two were excluded because of missing data. Table 9 lists the subtypes and corresponding achievement scores from the WJTA for the 113 learning disabled students.

Ten students were identified as belonging to Subtype 1. They were characterized as having relatively high, flat achievement profiles as indicated by standard scores above 90 in the areas of reading, math, and written language. Twenty-six of the LD students in this sample had moderate, flat achievement profiles as indicated by standard scores in

Table 5

WISC-R/WJTCA Correlation Matrix for Total Sample

| | WISC-R Scores | | | WJTCA Score |
|---|---|---|---|---|
| | PIQ | VIQ | FSIQ | WJ BCA |
| PIQ | | .48 | .70 | .45 |
| VIQ | | | .91 | .81 |
| FSIQ | | | | .81 |

Note. There were 157 students included in these calculations.

Table 6

WISC-R/WJTCA Correlation Matrix for Group 1

| | WISC-R Scores | | | WJTCA Score |
|---|---|---|---|---|
| | PIQ | VIQ | FSIQ | WJ BCA |
| PIQ | | .53 | .87 | .56 |
| VIQ | | | .87 | .71 |
| FSIQ | | | | .72 |

Note. Group 1 are re-evaluated LD students (n=70).

Table 7

WISC-R/WJTCA Correlation Matrix for Group 2

| | WISC-R Scores | | | WJTCA Score |
|---|---|---|---|---|
| | PIQ | VIQ | FSIQ | WJ BCA |
| PIQ | | .21 | .39 | .12 |
| VIQ | | | .88 | .74 |
| FSIQ | | | | .74 |

Note. Group 2 are newly referred LD students (n=48).

Table 8

WISC-R/WJTCA Correlation Matrix for Group 3

| | WISC-R Scores | | | WJTCA Score |
|---|---|---|---|---|
| | PIQ | VIQ | FSIQ | WJ BCA |
| PIQ | | .70 | .88 | .67 |
| VIQ | | | .95 | .86 |
| FSIQ | | | | .85 |

Note. Group 3 are newly referred non-LD students (n=39).

the 80s in the basic academic areas, and were classified as belonging in Subtype 2. Subtype 3, those with low, flat achievement profiles as indicated by achievement scores in the 70s in the basic academic areas, contained 21 subjects. The 19 students in Subtype 4 had superior reading skills in comparison to their math skills as evidenced by at least a 9 standard score point difference between the achievement scores, while the reverse was true of the 35 students in Subtype 5. Six LD students could not be classified according to the procedures developed in Chapter 3. A review of the scores for these individuals indicated that four of them had profiles similar to students in Subtypes 2 and 3, but one of the achievement scores would not qualify (e.g., reading score = 82, written language score = 83, math score = 77). The other two students had scores similar to Subtype 1, but again one score prevented the student from being placed.

A discriminant analysis indicated that 82.9% (95 of 111 students) could be correctly classified into subtypes on the basis of their WJTA scores. Table 10 lists the classification results from this analysis. A review of this table indicates that over 89% (59 of 66) of the students identified as belonging to Subtypes 1, 4, and 5 were correctly classified. Within each of these subtypes, students were incorrectly classified into just one other subtype. Of the students in Subtype 2, 80.8 % (21 of 26) were correctly classified, but misclassification within this subtype was much more diverse than in the other subtypes. One was identified as having an achievement profile

Table 9

Mean Achievement Test Scores for Each LD Subtype

| Subtype | n | Reading | | Math | | Written Lang | |
|---|---|---|---|---|---|---|---|
| | | M | SD | M | SD | M | SD |
| 1 | 10 | 93.0 | 3.9 | 94.8 | 6.8 | 92.3 | 4.1 |
| 2 | 26 | 84.4 | 2.5 | 84.7 | 3.9 | 84.1 | 4.3 |
| 3 | 21 | 71.7 | 7.6 | 71.5 | 8.4 | 74.9 | 5.1 |
| 4 | 20 | 90.4 | 8.3 | 78.1 | 6.5 | 86.4 | 9.1 |
| 5 | 36 | 78.0 | 6.8 | 92.5 | 7.3 | 81.1 | 8.3 |

Note. Subtype 1 is composed of students with high, flat achievement, Subtype 2 had students with moderate, flat achievement, Subtype 3 had students with low, flat achievement, Subtype 4 had students with higher reading than math, Subtype 5 had students with higher math than reading.

Table 10

Classification Results of Discriminant Analysis Using Achievement Scores to Predict Subtype Membership

| Actual Subtype | n | Predicted Subtype Membership | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 1 | 10 | 9 | 1 | 0 | 0 | 0 |
| 2 | 26 | 1 | 21 | 0 | 2 | 2 |
| 3 | 21 | 0 | 5 | 14 | 2 | 0 |
| 4 | 20 | 0 | 0 | 2 | 18 | 0 |
| 5 | 36 | 4 | 0 | 0 | 0 | 32 |

Note. Subtype 1 is composed of students with high, flat achievement, Subtype 2 had students with moderate, flat achievement, Subtype 3 had students with low, flat achievement, Subtype 4 had students with higher reading than math, Subtype 5 had students with higher math than reading.

more similar to Subtype 1, while four (two each) were identified as having profiles more similar to Subtypes 4 and 5. The greatest number of classification errors occurred with students in Subtype 3, although even here 66.7% (14 of 21) of the students were correctly classified. Nearly a quarter of the students in Subtype 3 (5 of 21) were predicted to belong to Subtype 2.

<u>Correlations Between WISC-R and WJTCA Scores Across Subtypes</u>

Tables 11-15 list the correlations between the WISC-R Verbal, Performance, and Full Scale IQ scores and the WJ Full Scale BCA score for each subtype. An examination of this table indicates that there is a wide variation in how the scores relate to each other across the subtypes. Correlations of .70 and above between the Full Scale scores from the tests are noted for Subtypes 1, 3, and 5. A moderately strong .57 correlation also exists between these scores for students in Subtype 2. The Full Scale scores for students in Subtype 2 exhibited a non-significant .42 correlation.

Also of interest is the relationship between the WISC-R Verbal and Performance IQs and the WJ Full Scale BCA score across the subtypes. For each subtype there is a much stronger relationship between the Verbal IQ and the WJ BCA score than there is between the Performance IQ and the WJ BCA score. Except for students in Subtype 4, there was a significantly positive relationship between the Verbal IQs and WJ BCA scores, with correlations ranging from .57 to .84. The Performance IQs, however, had correlations ranging from a high of .60 and .70 for

Table 11

WISC-R/WJTCA Correlation Matrix for Subtype 1

| | WISC-R Scores | | | WJTCA Score |
|---|---|---|---|---|
| | PIQ | VIQ | FSIQ | WJ BCA |
| PIQ | | .85 | .96 | .70 |
| VIQ | | | .96 | .84 |
| FSIQ | | | | .81 |

Note. Students in Subtype 1 had high, flat achievement.

Table 12

WISC-R/WJTCA Correlation Matrix for Subtype 2

| | WISC-R Scores | | | WJTCA Score |
|---|---|---|---|---|
| | PIQ | VIQ | FSIQ | WJ BCA |
| PIQ | | .42 | .80 | .39 |
| VIQ | | | .87 | .57 |
| FSIQ | | | | .57 |

Note. Students in Subtype 2 had moderate, flat achievement.

Table 13

WISC-R/WJTCA Correlation Matrix for Subtype 3

| | WISC-R Scores | | | WJTCA Score |
|---|---|---|---|---|
| | PIQ | VIQ | FSIQ | WJ BCA |
| PIQ | | .43 | .86 | .60 |
| VIQ | | | .83 | .73 |
| FSIQ | | | | .78 |

Note. Students in Subtype 3 had low, flat achievement.

Table 14

WISC-R/WJTCA Correlation Matrix for Subtype 4

| | WISC-R Scores | | | WJTCA Score |
|---|---|---|---|---|
| | PIQ | VIQ | FSIQ | WJ BCA |
| PIQ | | .32 | .82 | .22 |
| VIQ | | | .78 | .48 |
| FSIQ | | | | .42 |

Note. Students in Subtype 4 had higher reading than math scores.

Table 15

WISC-R/WJTCA Correlation Matrix for Subtype 5

| | WISC-R Scores | | | WJTCA Score |
|---|---|---|---|---|
| | PIQ | VIQ | FSIQ | WJ BCA |
| PIQ | | .21 | .34 | .02 |
| VIQ | | | .88 | .67 |
| FSIQ | | | | .70 |

*Note*. Students in Subtype 5 had higher math than reading scores.

Subtypes 1 and 3, respectively, to a low of .22 and .02 for Subtypes 4 and 5, respectively.

## Mean Score Difference by Subtype

T-test comparisons for paired observations were performed on the WISC-R and WJ Full Scale scores for each subtype. Table 16 lists the results of these tests. For all five subtypes there was a significant difference between the mean test scores, at an alpha level of .05 or less. For each subtype the WISC-R score was significantly higher than the WJTCA score with differences ranging from 4.7 standard score points for students in Subtype 4, to 8.3 standard score points for students in Subtype 5. Students in Subtypes 1 and 2 had nearly identical mean score differences between the tests, although both measures were approximately 5 points lower in Subtype 2 than Subtype 1. Students in Subtype 3 had the lowest scores on both tests and the second largest mean difference (7.67 points) between the tests.

In order to gain additional information about the differences between the tests across the subtypes, individual performances within each subtype were analyzed. Table 17 presents the results of this analysis. An examination of this table indicates that there is very little difference across the subtypes in terms of the distribution of differences in test scores. Slightly greater than a third of the students in each subtype had WISC-R and WJTCA scores within 5 points of each other (40% of Subtype 1, 38% of Subtype 2, 34% of Subtype 3, 45% of Subtype 4, and 42% of Subtype 5). However, with the exception of Subtype 4 (45%) over half the students in each subtype had WISC-R scores

Table 16

Mean Difference Between WISC-R and WJTCA Full Scale Scores for Subtypes of LD Students

| Subtype | n | Test | Mean | SD | Mean Diff | t-val | Sig |
|---|---|---|---|---|---|---|---|
| 1 | 10 | WISC-R | 96.2 | 13.4 | 5.8 | 2.30 | .05 |
| | | WJTCA | 90.4 | 11.7 | | | |
| 2 | 26 | WISC-R | 91.3 | 8.5 | 5.8 | 3.84 | .01 |
| | | WJTCA | 85.5 | 8.2 | | | |
| 3 | 21 | WISC-R | 82.6 | 10.8 | 7.7 | 5.11 | .01 |
| | | WJTCA | 74.9 | 9.7 | | | |
| 4 | 20 | WISC-R | 87.8 | 7.8 | 4.7 | 2.61 | .05 |
| | | WJTCA | 83.1 | 7.2 | | | |
| 5 | 36 | WISC-R | 95.1 | 10.0 | 8.3 | 6.70 | .01 |
| | | WJTCA | 86.8 | 9.5 | | | |

Note. Subtype 1 had students with high, flat achievement, Subtype 2 had students with moderate, flat achievement, Subtype 3 had students with low, flat achievement, Subtype 4 had students with higher reading than math, Subtype 5 had students with higher math than reading.

Table 17

Distribution of WISC-R/WJTCA Full Scale Score Differences for Subtypes of LD Students

| | WISC-R greater then WJTCA | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1-5pts. | | 6-10pts. | | 11-15pts. | | gt 15pts. | | Total[a] | |
| Subtype | n | % | n | % | n | % | n | % | n | % |
| 1 | 2 | 20 | 2 | 20 | 2 | 20 | 1 | 10 | 7 | 70 |
| 2 | 4 | 15 | 7 | 27 | 5 | 19 | 2 | 8 | 18 | 69 |
| 3 | 5 | 24 | 4 | 19 | 6 | 29 | 3 | 14 | 18 | 86 |
| 4 | 5 | 25 | 7 | 35 | 2 | 10 | 0 | 0 | 14 | 70 |
| 5 | 10 | 28 | 4 | 11 | 13 | 36 | 3 | 8 | 30 | 83 |

| | WISC-R less than WJTCA | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1-5pts. | | 6-10pts. | | 11-15pts. | | gt 15pts. | | Total | |
| Subtype | n | % | n | % | n | % | n | % | n | % |
| 1 | 1 | 10 | 1 | 10 | 0 | 0 | 0 | 0 | 2 | 20 |
| 2 | 2 | 7 | 2 | 7 | 0 | 0 | 0 | 0 | 4 | 15 |
| 3 | 2 | 10 | 1 | 5 | 0 | 0 | 0 | 0 | 3 | 15 |
| 4 | 2 | 10 | 1 | 5 | 1 | 5 | 0 | 0 | 4 | 20 |
| 5 | 2 | 6 | 1 | 3 | 0 | 0 | 0 | 0 | 3 | 9 |

Note. Subtype 1 had students with high, flat achievement (n=10), Subtype 2 had students with moderate, flat achievement (n=26), Subtype 3 had students with low, flat achievement (n=21), Subtype 4 had students with higher reading than math (n=20), Subtype 5 had students with higher math than reading (n=36).

[a] There was 1 student (10%) in Subtype 1, 4 students in Subtype 2 (15%), 2 students in Subtype 4 (10%), and 3 students in Subtype 5 (8%), who had identical WISC-R and WJTCA Full Scale scores.

that were at least 5 points greater than their corresponding WJTCA score.

## WJTCA Cognitive Clusters by Subtype

Table 18 lists the mean scores for the Cognitive Clusters across each subtype as well as the results from a series of one way ANOVAs with Scheffe follow-ups. An examination of the Cognitive Cluster score means across the subtypes reveals some interesting patterns. With one exception, students in Subtype 1 (high, flat achievement) scored higher across all clusters than those in Subtype 2 (moderate, flat achievement) who in turn scored higher across all clusters than those in Subtype 3 (low, flat achievement). The one reversal of this pattern was in the Reasoning Cluster where students in Subtype 2 scored higher than students in Subtype 1. Those who scored much better in math than they did in reading (Subtype 5) did better than those who performed better in reading than on math on 4 of the 6 Cognitive Clusters. The pattern of cluster ranks was very similar for Subtypes 2, 3, and 5 with the major difference being the magnitude of the scores. Subtypes 1 and 4, although similar in that Memory and Verbal Ability were the two highest ranked clusters, had little else in common and seem to represent unique cognitive cluster organizations.

One way ANOVAs with Scheffe follow-ups were conducted across the subtypes for each cluster to determine if the subtypes performed differently on the various measures of aspects of cognitive functioning. As can be seen from Table 18, significant differences were found between the subtypes for each cluster with the exception of the Perceptual Speed

Table 18

Results of One Way ANOVA of WJTCA Cognitive Cluster Scores by Subtypes of LD Students

| Clusters | Subtypes | | | | | $\underline{F}$ | Sig | Scheffe |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | | | |
| Verb Abil | 96.2 | 86.8 | 80.9 | 88.4 | 88.4 | 3.69 | .01 | 1 > 3 |
| Oral Lang | 88.9 | 83.2 | 74.1 | 82.6 | 85.1 | 3.92 | .01 | 1&5 > 3 |
| Reasoning | 88.5 | 91.9 | 84.1 | 85.0 | 94.7 | 3.53 | .01 | 5 > 3 |
| Brd Reason | 87.9 | 87.2 | 78.3 | 82.7 | 91.6 | 5.07 | .01 | 5 > 3 |
| Percpt Spd | 95.9 | 91.7 | 81.6 | 84.9 | 90.2 | 3.43 | .01 | none |
| Memory | 97.7 | 90.2 | 81.0 | 88.5 | 88.4 | 3.46 | .01 | 1 > 3 |

Note. Subtype 1 had students with high, flat achievement, Subtype 2 had students with moderate, flat achievement, Subtype 3 had students with low, flat achievement, Subtype 4 had students with higher reading than math scores, Subtype 5 had students with higher math than reading scores.

Cluster. Although significant differences were found, there was a great degree of overlap between the cluster scores across the subtypes which limited the number of differences. As a result, all the differences involved Subtype 3 (low, flat achievement) scoring significantly lower than Subtype 1 (in Verbal Ability, Oral Language, and Memory) and Subtype 5 (in Oral Language, Reasoning, and Broad Reasoning).

A discriminant analysis was conducted to determine if students' performances on the Cognitive Clusters could accurately predict subtype membership. Table 19 lists the results of this analysis. These results indicate that under half the subjects could be correctly classified into their identified subtypes on the basis of their scores on the Cognitive Clusters. An examination of this table indicates that the greatest success was obtained for subjects in Subtype 1, where 8 of the 10 students' (80%) subtype membership was accurately predicted, and Subtype3, where 14 of the 21 students (67%) were accurately placed into the appropriate subtype on the basis of their Cognitive Cluster scores. The Cognitive Cluster scores were only moderately successful (16 of 36, 44%) in predicting subtype membership for students in Subtype 5 and completely unsuccessful for students in Subtypes 2 (7 of 26, 27%) and 4 (7 of 20, 35%). An analysis of the misclassifications does not indicate any consistent pattern.

Table 19

Classification Results of Discriminant Analysis Using Cognitive Cluster Scores to Predict Subtype Membership

| Actual Subtype | n | Predicted Subtype Membership | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 1 | 10 | 8 | 0 | 0 | 1 | 1 |
| 2 | 26 | 4 | 7 | 5 | 4 | 6 |
| 3 | 21 | 1 | 3 | 14 | 3 | 0 |
| 4 | 20 | 0 | 4 | 5 | 7 | 4 |
| 5 | 36 | 6 | 2 | 6 | 6 | 16 |

Note. Subtype 1 had students with high, flat achievement, Subtype 2 had students with moderate, flat achievement, Subtype 3 had students with low, flat achievement, Subtype 4 had students with higher reading than math scores, Subtype 5 had students with higher math than reading scores.

## CHAPTER 5

## SUMMARY AND DISCUSSION

This study was conducted to examine the performance on the WISC-R and WJTCA by previously identified LD students undergoing a three year re-evaluation (Group 1), and two groups of newly referred students, those subsequently identified as learning disabled (Group 2), and those not identified as learning disabled (Group 3). Previous researchers have questioned the comparability of scores from these tests when used with learning disabled or referred students. It was anticipated that this study would help clarify this issue. In addition, performances on the two tests were examined from an LD subtype framework.

### Summary of Results

The results of this study indicate that while there was a strong correlation between test scores across all three groups, the mean WISC-R Full Scale IQ was significantly higher than the mean WJ Full Scale BCA score for each group. The largest difference was found with the re-evaluated learning disabled students, while the smallest difference was found with the students who were referred but not identified as learning disabled. Less than half the students had WISC-R and WJTCA scores that were within 5 points of each other, while over three quarters of the students had WISC-R scores that were greater than their WJTCA scores.

Subtypes of learning disabled students were formed on the basis of their performance on the reading, math, and written language clusters of the WJTA. Strong correlations between the WISC-R and WJTCA were obtained for Subtypes 1, 3, and 5, while moderate correlations were obtained for Subtypes 2 and 4. Across each subtype there was a significant difference between the Full Scale scores from the two tests. Just less than half the students had WISC-R and WJTCA Full Scale scores that were within 5 points of each other. A discriminant analysis indicated that the LD students' subtypes could be correctly identified in less than half of the cases on the basis of their Cognitive Cluster scores.

## Theories Explaining Mean Score Discrepancy

The results of this study indicate that the WJTCA results in a significantly lower estimate of the cognitive functioning than does the WISC-R when used with learning disabled students and for students referred but not identified as learning disabled. This finding is consistent with virtually every study which has examined this issue. The pattern of results has been so consistent that the question of a difference in test scores for these special groups of students should be put to rest. The difference does exist. What has yet to be determined is the basis for this difference. Various reasons have been proposed which will be examined in light of the findings of this study. Of equal importance is how the difference in test scores and proposed explanations should be interpreted by practitioners, and the resultant effect on their evaluation and identification of special education students.

Theories explaining the differences between the results on the WISC-R and WJTCA for learning disabled or referred students can be grouped into three main classes. The explanations put forth by Woodcock (1980, 1984), because he faults previous research for serious methodological flaws calling into question the validity of the findings, will be examined first.

Woodcock's Perspective

Woodcock has questioned the results of the previous studies because of instrument, score, or sample bias. Woodcock defines instrument bias as occurring when "one of the two instruments to be compared had been used for the selection or subclassification of the sample" (1984, p.342). He states that "referrals, not selectees, will usually be the desired sample populations. Tests to be compared should be administered to the entire referral population, and the discussion and conclusion should be based on this relatively unbiased data" (1984, p.343). Groups 2 and 3 in this study, the referral sample, meet this requirement. In addition, placement decisions (subclassification into LD-Group 2, not LD-Group 3) were based on performance on both tests. Instrument bias is present to a certain extent in Group 1, the re-evaluated LD students. These students were originally diagnosed using WISC-R scores as one of the selection criteria. In discussing previous research, Woodcock's concern in this area is that instrument bias results in the use of scores

> . . . obtained for a subset of students who had been previously placed in LD programs based on relatively high WISC-R scores coupled with relatively low achievement test

> scores. Students with relatively low WISC-R scores were selectively not placed in these LD programs, and consequently their performances on the WISC-R were not part of the comparison data. As a result, the mean of the WISC-R scores was spuriously high. (p.343)

The studies to which he refers had group mean WISC-R Full Scale IQ scores of 98.6 and 100.7. In the present study, it is very unlikely that students in Group 1 had spuriously high WISC-R scores. The mean WISC-R score for this group (88.65) was 5 and 12 points lower than the mean WISC-R score for the other groups.

Score bias, which Woodcock considers an exacerbation of instrument bias, and which he defines as not re-administering the selection test to obtain new scores when it is used as part of the comparison, was controlled for in Group 1 by the re-administration of the WISC-R. Woodcock has also cautioned against subdividing a sample on the basis of a selection test even if it does not involve the elimination of subjects. This aspect of score bias was controlled for across all groups when dividing into LD and non-LD by considering the students' performances on both tests. Neither test was solely responsible for the placement decision.

The third flaw Woodcock has discussed is sample bias or the selection of a sample on some basis other than random sampling from the general population. He states that this bias occurs differently as a function of the correlation between the tests and the sample selection criteria. It is difficult to determine to what extent this occurred in this study. The selection criteria for inclusion in this study for students in Group 1 was their prior inclusion in special education programs for the learning disabled, which implies that there was an identified discrepancy between aptitude and achievement in at least one of several possible areas. Underlying this criteria is the assumption that these students were originally referred for an evaluation because of poor academic achievement in the classroom. There were not any explicit selection criteria for students in Groups 2 and 3. However, it is reasonable to assume that these students were also referred for an evaluation because of poor classroom performance. The WJ Full Scale BCA score has been shown to be more strongly correlated with achievement test scores than the WISC-R Full Scale IQ score. In the present study this was true for students in Groups 1 and 2, particularly in the areas of reading and written language. Students in Group 1 had reading, math, and written language correlations of .19, .51, and .12 with their WISC-R Full Scale IQ score, while the corresponding WJ Full Scale BCA reading, math, and written language achievement scores had correlations of .43, .60, and .41. In Group 2, the reading, math, and written language scores correlated .21, .46, and .25 with the WISC-R and .37, .55, and .47 with the WJTCA. The correlations between the two tests and the achievement measures were much stronger and more consistent for students in Group 3. The WISC-R correlated .68, .62, and .67 with the reading, math, and written language achievement scores while the WJTCA had corresponding correlations of .75, .71, and .69. Students in Groups 1 and 2 had reading, math, and written language achievement scores in the low to mid 80s, while the scores for the students in Group 3 were in the mid 90s.

Woodcock's argument would be that when a sample is chosen on a non-random basis because of poor classroom achievement corresponding to low achievement test scores, the resulting WJ Full Scale BCA score for these students will be lower than their WISC-R Full Scale IQ score because the WJTCA correlates higher with tests of achievement than does the WISC-R. It is doubtful, however, that the mean score difference between the WISC-R and WJTCA can be solely attributed to the slightly higher correlations that exist between the WJTCA and achievement, than between the WISC-R and measures of achievement (e.g., Woodcock, 1984, cites as typical correlations .78 between the WJTCA and reading achievement vs .68 between the WISC-R and reading achievement). While sample bias cannot be discounted as an explanation for the WISC-R/WJTCA mean score discrepancy, it is at best only part of the answer.

## Difference in Test Norms

If one is able to eliminate Woodcock's bias explanations as the major determinant of the mean score difference between the WISC-R and WJTCA for learning disabled or referred students, then alternative explanations must be considered. Questions concerning the issue of a difference in the norms between the two tests as the explanatory mechanism comprise a second major approach to understanding the test score differences. This issue has been raised in two ways. Reeve, Hall, and Zakreski (1979) suggested that the test score differences arose because of errors in calculating the WJTCA norms. Also within this "Woodcock norm error" approach is Cummings and Moscatto's (1984) suggestion that there was a bias in the selection procedures used to obtain the WJTCA norm group. Both of these variants of WJ norm errors can be rejected. As Thompson and Brassard (1984) have suggested, the Reeve et al. hypothesis can be rejected because any calculations involving the WJTCA norms should appear with non-handicapped samples as well as handicapped samples. As Woodcock has pointed out, this "norm error" hypothesis has no empirical basis other than the observation of a difference between the test scores. Cummings and Moscatto suggested that a differential rate of return of parental permission forms from low SES families may have affected the difficulty level of the norms. Again, there are no data to support this conjecture and, Woodcock (1984) believes, it can be partially refuted by examining the summary statistics for the raw normative data. He reports that a slight negative skew in the raw data exists at all grade and age levels rather than the hypothesized positive skew.

The second hypothesis involving a difference in norms as the explanatory mechanism does not involve finding fault with the WJTCA norms. Instead, this hypothesis states that part of the discrepancy between the WJTCA and WISC-R exists simply because different norm groups were used to standardize each test (Sabers, personal communication). Even though the standardization procedures for the tests were conducted within approximately 3.5 years of each other (from 12/71 to 1/73 for the WISC-R and 4/76 to 3/77 for the WJTCA), and the norm groups were intended to be well representative of the U.S. population, Sabers states that the identified score discrepancy is to be expected given that the tests were normed on different groups. Some degree of difference is to be expected. As Cummings and Moscatto (1984) state, "as new instruments are published and old instruments are restandardized, statistically significant mean score differences are invariably reported both for normal and special populations" (p.47). Research conducted with the McCarthy Scales of Children's Abilities (McCarthy, 1972) and the more recently introduced Kaufman Assessment Battery for Children (Kaufman, 1983) demonstrates that each test, like the WJTCA, exhibits some discrepancy between itself and some other, older tests measuring the same construct (e.g., Kaufman, 1982; Naglieri and Haddad, 1984). Each test attempts to measure aspects of cognitive ability in different ways, and each has a different norm group, so it is reasonable to expect differences in scores for particular individuals. Undoubtably, this helps explain a portion of the discrepancy noted between the WISC-R and WJTCA for identified LD and referred students, but some other mechanism must also be operating. The difference between the two test scores for this population is larger than for non-referred, non-identified LD students (Thompson and Brassard, 1984; Woodcock, 1984). If the difference in norm groups alone could account for the test score differences, then similar differences should be obtained regardless of the sample. This is not the case with the WISC-R and WJTCA.

## Difference in Test Content

A third possible explanation looks to the different content of the two tests as the major determinant of the differences between the test scores for the LD and referred students. This hypothesis suggests that the WJTCA, in comparison to the WISC-R, taps skills that are more difficult for low achieving or LD students than for normal achieving students (Reeve, Hall, and Zakreski, 1979; Ysseldyke, Shinn, McGue, and Epps, 1981). Some, such as Reeve et al., have looked to specific subtests of the WJTCA as the area of difference (e.g., the subtests comprising the Perceptual Speed Cluster), while others (Ysseldyke, Shinn, and Epps, 1981; Thompson and Brassard, 1984) believe that the WJTCA is a product dominant test, or one that is dominated by content that is achievement or acquired knowledge in nature. They state that students who are low achievers, as referred or LD students typically are, would do poorer on a test of this nature than on one that is less dominated by such content. In this study, the students with the lowest achievement, those in Groups 1 and 2, also had the largest WISC-R/WJTCA score discrepancy. The corollary to this hypothesis is that students who are high achievers would be expected to do better on a test of this nature. Mather (1984) examined the performance of learning disabled and gifted students on the WISC-R and WJTCA. The gifted students had mean achievement standard scores in math, reading, and written language ranging from 119 to 122, while the learning disabled students had standard scores ranging from 80 to 86. Consistent with other studies, the WJTCA was lower than the WISC-R for LD students (86.1 vs. 91.9). However, the gifted children also had WJTCA scores that were lower than their WISC-R scores, although the scores were closer (126.9 vs. 128.2, respectively). It is possible, however, that the "achievement content" of the WJTCA is such that it is of a level of difficulty that adversely effects the performance of low achieving or learning disabled students in comparison to their performance on the WISC-R, while at the same time not being difficult enough so that high achieving or gifted students would perform better on this test than on the WISC-R.

Woodcock (1984) has responded to the hypothesized "achievement content" of the WJTCA by stating that many authors have confused "correlation with causation" (p.351). While it is true that the WJTCA correlates higher with tests of achievement than the WISC-R, Woodcock correctly points out that this does not mean that the WJTCA has greater achievement content. Rather, Woodcock (1984) states that the correlations are stronger because the subtests of the WJTCA were designed to "measure a heterogeneous mix of cognitive skills directly related to school learning" (p.353).

A review of the subtests would indicate that there is some validity to both arguments. Subtests such as Spatial Relations, Memory for Sentences, Visual-Auditory Learning, Analysis-Synthesis, and Concept Formation seem to have very few items that would be affected by school learning, while seemingly able to assess areas related to school achievement. Other subtests, notably Picture Vocabulary, Quantitative Concepts, Antonyms-Synonyms, and Analogies, because they reflect acquired knowledge, would seem to be directly affected by school learning. That these particular subtests have the highest correlations with achievement (Woodcock, 1984) is supportive of this contention.

Rather than fault the hypothesized achievement content of the WJTCA some researchers (Coleman and Harmer, 1985; Phelps, Rosso, and Falasco, 1984) have suggested that the test is heavily verbally loaded.

Cummings and Moscatto (1984) state that the WJTCA is "assessing a restricted sample of the abilities measured by the WISC-R, most notably, its verbal component" (p.36). The belief is that because LD students typically have greater difficulty with verbal than non-verbal skills, a test that emphasizes verbal skills would be more difficult for them. In support of this hypothesis, these authors have examined the correlations that have been reported between the Wechsler scales and the WJ Full Scale BCA scores. The WJ Full Scale BCA score has consistently been found to correlate much higher with the WISC-R Verbal IQ than with the Performance IQ. Several studies (Ysseldyke, Shinn, and Epps, 1981; Bracken, Prasse, and Breen, 1984; Coleman and Harmer, 1985) have reported correlations that were as strong between the WJ BCA score and the WISC-R Verbal IQ as between the WJ BCA score and WISC-R Full Scale scores. The results of this study do not contradict this hypothesis. In each group the WJ BCA score correlates as well with the WISC-R Verbal IQ as it does with the Full Scale IQ and much less so with the Performance IQ.

An examination of the Cognitive Cluster Scores tends to support the hypothesized achievement/verbal content of the WJTCA. In this study, the re-evaluated LD students and newly referred LD students both obtained their highest scores on the Reasoning Cluster (Analysis-Synthesis, Concept Formation, Analogies, with Antonyms-Synonyms as a suppressor), and lowest scores on the Oral Language Cluster (Picture Vocabulary, Antonyms-Synonyms, Analogies). The Broad Reasoning (Analysis-Synthesis, Concept Formation, Analogies) and Verbal Ability

(Picture Vocabulary, Antonyms-Synonyms, with Analysis-Synthesis as a suppressor) Clusters rank in the bottom for both groups, while the Memory (Numbers Reversed, Memory for Sentences) and Perceptual Speed (Visual Matching, Spatial Relations) Clusters rank in the top half for each group. The clusters in the top rankings are predominantly non-verbal, non-achievement in nature, while the clusters in the bottom rankings are primarily verbal and achievement oriented. In both cases, however, the rankings may be misleading because, with the exception of the Oral Language Cluster, there are fewer than 5 points separating the highest ranked cluster from the lowest.

A review of the literature indicates that similar patterns were found in other studies. The LD children in Mather's study (1984) also had the Reasoning and Oral Language Clusters as their highest and lowest ranked clusters, respectively. Memory, a non-verbal, non-achievement oriented cluster, and Broad Reasoning were in the top ranked clusters while the more verbal and achievement oriented Verbal Ability Cluster along with the Perceptual Speed Cluster were in the bottom ranked. Again, however, the spread between the highest and lowest ranked clusters, with the exception of the lowest ranked Oral Language Cluster, is a relatively small 5.6 points. The LD children in the Ysseldyke, Shinn, and Epps (1981) study had the Reasoning and Perceptual Speed Clusters as top ranked clusters and the Verbal Ability and Memory Clusters as the lowest (the Oral Language and Broad Reasoning Clusters had not yet been developed when the older studies were conducted). Referral students who were identified as LD in the Bracken, Prasse, and

Breen (1984) study had Cognitive Cluster scores that deviated from this pattern. Their students had the Verbal Ability Cluster as the highest ranked followed by the Reasoning, Perceptual Speed, and Memory Clusters. Again, a small 4 point spread separated the highest from lowest ranked cluster. The rankings from the Reeve, Hall, and Zakreski (1979) study were also somewhat different. Although they also found the Reasoning Cluster to be the highest ranked cluster for their LD students, in their study the Verbal Ability Cluster was next highest with the Memory and Perceptual Speed Clusters in the bottom ranks. Thompson and Brassard (1984) examined mild and severe LD students and found differences. The severe LD students had the same pattern of Reasoning and Memory as the highest ranked clusters, while Perceptual Speed and Verbal Ability were the lowest ranked. The mild LD students, however, had a very different pattern. For them, Verbal Ability and Memory were highest ranked while Reasoning and Perceptual Speed were lowest ranked.

In sum, a fairly consistent pattern of rankings for the Cognitive Clusters was found for students identified as LD. Clusters that are primarily non-verbal or non-achievement in nature (Reasoning, Broad Reasoning, Perceptual Speed, and Memory) are typically higher ranked than the clusters which could be considered to be more verbal or achievement oriented (Verbal Ability, Oral Language). However, the significance of this pattern is lessened when the relatively small differences between cluster scores is considered.

A different pattern emerges when the samples of students are those who were referred but not placed, low achieving, gifted, or randomly selected from a regular education class. In the present study, students who were referred but not placed had Broad Reasoning, Reasoning, and Memory Clusters as the highest ranked, while the Verbal Ability, Perceptual Speed, and Oral Language Clusters were the lowest ranked. Fewer than 4 points separated the highest from lowest ranked cluster. Students in the Bracken, Prasse, and Breen (1984) study who were referred but not identified as LD scored highest on the Verbal Ability Cluster, followed by Reasoning, Memory and Perceptual Speed. Just 3.8 points separated the highest from the lowest ranked clusters. The Coleman and Harmer study (1984), with referred students, had cluster scores in the same pattern as the Bracken et al. study although the Verbal Ability Cluster was 4.2 points higher than the next highest cluster, while 4.3 points separated the second and fourth clusters. The referred students in the McGrew (1983) study scored highest on the Reasoning Cluster followed by the Verbal Ability, Memory, and Perceptual Speed Clusters. The Reasoning Cluster was substantially higher than the Verbal Ability Cluster (7.4 points), while 3.2 points separated the second from the fourth clusters.

Mather's (1984) gifted students and Thompson and Brassard's (1984) regular education students had identical patterns that were similar to the non-identified referral population. The gifted students had scores in each cluster that were much higher than any other group, and the difference between scores between each cluster was more evenly distributed. In the Thompson and Brassard study, the regular education students' Memory Cluster scores were 8.1 points higher than the second ranked Verbal Ability Cluster, while the second and fourth ranked clusters were separated by 1.3 points. The low achieving students in the Ysseldyke, Shinn, McGue, and Epps (1981) study had a similar pattern. They scored highest on the Verbal Ability Cluster followed by the Memory, Perceptual Speed, and Reasoning Clusters. In sum, these non-LD students, in comparison to identified or re-evaluated LD students, tended to score higher on the Cognitive Clusters that were verbal and achievement oriented, and lower on the clusters that were more non-verbal or non-achievement in nature.

Reconciliation of Mean Score Discrepancy Hypotheses

Were it not for the WISC-R/WJTCA mean score discrepancy issue, there would be much less controversy about the use of the WJTCA with referred and learning disabled students. Its many strengths would far outweigh its few apparent weaknesses. The crucial issue, then, is whether the differences that have been found between the WISC-R and WJTCA are valid and to be expected, or arise because the WJTCA results in inaccurate estimates of cognitive ability when used with these populations. A decision on the value of the WJTCA depends on which hypotheses to explain the discrepancy one chooses to accept. Of the many hypotheses that have been advanced to explain the WISC-R/WJTCA mean score discrepancy, only one (flawed WJTCA norms) can be safely rejected. Others (use of different norm groups, flawed research) can probably account for a minor portion of the discrepancy. The major issue revolves around the contents of the two tests. Those who believe that the WJTCA results in inaccurately low estimates of cognitive ability for LD or referred students cite what they believe to be the overly verbal and achievement oriented content of the WJTCA as the reason. These students typically are deficient in both these areas and so would obtain a lower score on a test of this nature. Although Woodcock would disagree, the preponderance of the data reported to date would suggest that that is an accurate interpretation.

That does not mean that the WISC-R produces "correct" estimates of cognitive ability while the WJTCA produces "wrong" estimates. Assuming that one believes that "intelligence" can be measured and then refined to a number, one must then determine what questions should be asked and what tasks required. The developers of the WJTCA, in an attempt to "better reflect the skills necessary for success in school" (Woodcock, 1984, p.381) have developed a test with different content than the WISC-R. Research suggests that it is more verbal and achievement oriented than the WISC-R. When does "more" become too much when the aim is an accurate estimate of cognitive ability? Those who believe that the items and subtests on the WJTCA tap those skills that are assumed to comprise intelligence will gladly add the WJPEB to their repertoire of assessment instruments, and will not be alarmed by the WISC-R/WJTCA mean score discrepancy. They also will undoubtedly identify fewer students as learning disabled than will users of the WISC-R (assuming achievement scores are held constant and that the examiner is operating from an aptitude/achievement discrepancy model). Practitioners who believe that the WJTCA is too weighted with items of an achievement or verbal nature to be an accurate measure of cognitive ability should reject the test. The nature of intelligence and intelligence testing is vague enough that either conclusion can be supported. Ultimately, test choice is decided by the theoretical orientation of the examiner in conjunction with personal experience and, hopefully, informed opinion.

## LD Subtypes

One of the recent trends in the field of learning disabilities has been the move toward identifying subtypes of learning disabled students. Two main approaches have been used. The first method involves the use of multivariate statistical techniques such as Q-type factor analysis to identify subtypes. A second approach, and the one used in this study, involves clinical inspection of the psychometric protocols, and then grouping students into subtypes on the basis of perceived similarity. The goal of this procedure is to obtain subtypes of LD students who share common attributes that distinguish them from other subtypes. This was accomplished in this study by examining achievement patterns on the WJTA Reading, Math, and Written Language Clusters. The assumption is that distinct achievement patterns result from particular patterns of strengths and weaknesses across various aspects of cognitive ability. The subtype identification rules established in this study were successful in grouping LD students into 5 distinct subtypes as indicated by the discriminant analysis conducted on these data. Over 82% of the students could be correctly placed into the appropriate subtype on the basis of the patterns of their achievement scores.

Test Score Differences for LD Subtypes

One of the goals of this subtyping was to determine if there were particular subtypes of LD students who scored differently on the WISC-R and WJTCA than did the entire heterogeneous group. McGrew (1983) has suggested that the differences found between the WISC-R and WJTCA may be an accurate reflection of certain of these students' performances. This contention implies that these differences exist only for certain types of students. This idea was somewhat supported in this study, although it would be more accurate to suggest that all the subtypes exhibited significant differences and that only the magnitude of the differences varied across the subtypes. Students whose math scores were at least 10 points better than their reading scores (Subtype 5) had the largest mean score difference, while those whose reading scores were at least 10 points better than their math scores (Subtype 4) had the smallest mean score difference, just over half as large. Those who were lowest in achievement (Subtype 3) had the second largest mean score difference. It would seem that those LD students weakest in reading or much below average in cognitive ability would have more difficulty on the WJTCA in comparison to the WISC-R than would other types of LD students. Nearly half of these students (43% in Subtype 3, 44% in Subtype 5) had WISC-R/WJTCA differences greater than 11 points. Those showing a strength in reading or of low average to average cognitive ability had the most comparable performances on the two tests. Practitioners evaluating LD students known to have reading difficulties should be aware that these students may have a larger discrepancy between their WISC-R and WJTCA scores than do other LD students. The results from this study are not conclusive in this regard but they do warrant consideration.

Cognitive Cluster Differences for LD Subtypes

A second major purpose of this subtyping was to discover if particular subtypes of LD students demonstrate different strengths and weaknesses across aspects of cognitive functioning. The WJTCA, because it provides measures of cognitive functioning in four different areas (Perceptual Speed, Memory, Reasoning or Broad Reasoning, Verbal Ability or Oral Language), would seem to be an ideal instrument for this purpose. This study, however, was not able to support this contention. Although differences were found across the subtypes for 5 of the 6 clusters, they were not particularly revealing. All of the differences involved the lowest achieving and least cognitively able students in Subtype 3 scoring lower than the relatively high achieving students in Subtype 1, or those whose math skills were better than their reading skills (Subtype 5). The simplest explanation for this is that students in Subtype 3 had generally depressed cognitive skills which showed up as significantly lower in a number of areas when compared to the generally more able students in Subtypes 1 and 5. This contention is supported by a review of Table 16 which indicates that the students in Subtypes 1 and 5 had the highest WISC-R and WJTCA scores, while those in Subtype 3 had the lowest. These results would indicate that rather than uncovering specific LD subtypes with particular strengths and weaknesses, the differences found in this study are more indicative of differences in general levels of cognitive ability.

An additional indication of this lack of distinct cognitive strengths and weaknesses across the subtypes was the failure of the discriminant analysis to predict subtype membership on the basis of the cognitive clusters. The subtypes developed in this study did not have distinct cognitive cluster patterns. While it is clear that distinct subtypes could be established on the basis of achievement scores, there was not any corresponding uniqueness in the types of cognitive abilities and disabilities demonstrated by these students. This inability to recover cognitive ability patterns unique to each subtype can be attributed to a number of possible factors.

Breen (1986) was also not successful in identifying patterns of cognitive skills specific to particular subtypes of LD students. He identified possible confounding factors including that the students exhibited a general deficit in areas of language or memory that overwhelmed whatever other differences may have been present, and that subtests of the WJTA are not sensitive enough to uncover what subtle differences may be present. Another possibility is that the subtests of the WJTA do not assess the types of abilities that could be used to subtype LD students. A third possibility is that the method of subtyping may not have been successful in identifying truly unique subtypes. As previously mentioned, the subtypes identified in this study may more accurately be described as reflecting different levels of cognitive ability, as opposed to unique subtypes of learning disabled students.

The inability of this study and the Breen study to identify unique cognitive patterns for subtypes of LD students identified on the basis of achievement scores raises several concerns. The first question that needs to be addressed, and one that is central to this entire study, is whether the students in this sample should even be considered learning disabled.

## Was the Sample LD

The students in this study were drawn from a typical small to medium sized district (3500 students in 4 grade schools, 1 junior high and 1 high school) located on the outskirts of a large city. The criteria for placement as a learning disabled student are based on Arizona State Law mandating that a severe discrepancy between aptitude and achievement be present in order to diagnose a learning disability. As in many districts, however, this district has no formal operationalization of what is meant by a "severe discrepancy." As a result, each Multi-Disciplinary Team is free to make its own determination of when a severe discrepancy is present and, consequently, when students can be diagnosed as learning disabled. This results in numerous students being identified despite the absence of a discrepancy between aptitude (whether measured by the WISC-R or the WJTCA) and achievement.

In this study, discrepancies of 6.9, 3.8, and 7.5 points were noted between the WISC-R Full Scale IQ and reading, math, and written language scores (achievement scores lower) for the re-evaluated students in Group 1. Corresponding differences of -.3, -2.4, (achievement scores higher) and .3 (achievement scores lower) were noted between the WJTCA and achievement in reading, math, and written language, respectively. Even with regression error (which becomes larger as one moves further away from the normative mean of the aptitude scores) considered, the difference between aptitude (whether measured by the WISC-R or the WJTCA) and achievement would not fit most definitions of severe. Newly referred students who were subsequently identified as LD (Group 2) had WISC-R/reading, math, and written language average achievement discrepancies of 11.7, 8.7, and 9.9 points, respectively. Corresponding discrepancies of 5.9, 2.9, and 4.1 points were noted between the WJTCA and the reading, math, and written language achievement scores. Regression error is less of a factor for this group, since the mean WISC-R and WJTCA Full Scale scores were 93.8 and 88.0, respectively, versus 88.7 and 81.5 for the corresponding scores in Group 1. Although the students in Group 3 were not identified as learning disabled, the discrepancies between the WISC-R and the achievement scores (6.7 for reading, 4.7 for math, 5.2 for written language) were nearly the same as for Group 1, while the WJTCA-achievement discrepancies (2.9 in reading, .9 in math, 1.4 in written language) were larger than those in Group 1. Although regression error would be smaller for Group 3 than for either Groups 1 or 2, the major factor would seem to be level of achievement. Achievement cluster scores ranging from 81.2 to 85.1 would certainly make it appropriate to label the students in Groups 1 and 2 as low achieving, but the absence of a severe aptitude/achievement discrepancy makes it questionable as to whether these students are learning disabled.

These results, however, are typical of the other studies discussed in this study. Many of the studies that included LD students failed to include achievement score data, so it is not possible to determine whether the students in their samples exhibited a severe discrepancy between aptitude and achievement. Of those that did report this data, Ysseldyke, Shinn, and Epps (1981) found that their students had an average WISC-R Full Scale score of 100, WJTCA Full Scale score of 92.4 and a PIAT Total Score of 91.9. The Mather study (1984) reported a WISC-R Full Scale IQ score of 91.9, WJ Full Scale BCA score of 86.1 and reading, math, and written language achievement scores of 80.2, 86.6, and 81.7, respectively. The Breen (1986) subtyping study reported achievement scores and WJTCA scores that were nearly identical to the corresponding subtypes in this study. The high-math, low-reading subtype had achievement scores of 91.5 and 75.4, respectively, and a WJTCA score of 86.4. The high-reading, low-math subtype had achievement scores of 91.5 and 73.5, and a WJTCA score of 83.3. The low-reading, low-math subtype had achievement scores of 67.5 in both and a WJTCA score of 72.3.

It would seem that it is the nature of the field that many students who are identified as learning disabled are improperly diagnosed if one of the criteria for identification is the presence of a severe discrepancy between aptitude and achievement. The reality is that a learning disability is whatever the local school's Multi-Disciplinary Team says it is. While not fitting a formal operationalized definition of a learning disability, the LD students in this study, and in the other studies discussed in this report, are representative of the LD student population. This conclusion has important implications for researchers and practitioners.

Researchers who select samples of LD students from school districts need to be aware that these students are more likely to be low achievers than truly learning disabled based on a formal operationalized definition. Either those in the profession responsible for identifying students as learning disabled need to begin basing their identifications on accepted definitions, or those who are influential in formulating these definitions need to shape them to reflect what is happening in the field. The present state of affairs retards progress in research and in education. The development of accurate subtyping methodologies which could be used in an efficient, cost-effective manner by practitioners would be of great value. Of even greater value, and of primary importance, would be the development of a valid definition of learning disability, and a means to accurately operationalize it.

## LIST OF REFERENCES


Adams, K. (1985). Theoretical, methodological, and statistical issues. In B. Rourke (Ed.), Neuropsychology of Learning Disabilities. New York: Guilford Press.

Adelman, H., and Taylor, L. (1985). The future of the LD field: A survey of fundamental concerns. Journal of Learning Disabilities, 18, 423-427.

Algozzine, B., Ysseldyke, J., and Shinn, M. (1982). Identifying children with learning disabilities: When is a discrepancy severe? Journal of School Psychology, 20, 299-305.

Anastasi, A. (1985). Some emerging trends in psychological measurment: A fifty year perspective. Applied Psychological Measurment, 9, 121-138.

Arffa, S., Rider, L., and Cummings, J. (1984). A validity study of the Woodcock-Johnson Psycho-Educational Battery and the Stanford-Binet with black preschool children. Journal of Psychoeducational Assessment, 2, 73-77.

Baglin, R. (1981). Does nationally normed really mean nationally? Journal of Educational Measurement, 18, 97-107.

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

Bannatyne, A. (1974). Diagnosis: A note on recategorization of the WISC-R scaled scores. Journal of Learning Disabilities, 7, 272-273

Bianchini, J., and Loret, P. (1974). Anchor Test Study. Final Report, available as ERIC Documents ED 092 601 through ED 092 631.

Blaha, J., and Vance, H. (1979). The heirarchical factor structure of the WISC-R for learning disabled children. Learning Disability Quarterly, 2, 71-75.

Blashfield, R. (1980). Propositions regarding the use of cluster analysis in clinical research. Journal of Consulting and Clinical Psychology, 48, 456-459.

Blashfield, R., and Aldenfelder, M. (1978). The literature on cluster analysis. Multivariate Behavior Research, July, 271-295.

Bracken, B., Prasse, D., and Breen, M. (1984). Concurrent validity of the Woodcock-Johnson Psycho-Educational Battery with regular and learning disabled students. Journal of School Psychology, 22, 185-192.

Breen, M. (1983). Comparing the Woodcock-Johnson achievement cluster scores and Wide Range Achievement Test in a learning disabled and regular educational population. Journal of Clinical Psychology, 1, 86-90.

Breen, M. (1985). The Woodcock-Johnson Tests of Cognitive Ability: A comparison of two methods of cluster scale analysis for three learning disabilities subtypes. Journal of Psychoeducational Assessment, 3, 167-174.

Breen, M. (1986). Cognitive patterns of learning disability subtypes as measured by the Woodcock-Johnson Psycho-Educational Battery. Journal of Learning Disabilities, 19, 86-90.

Breen, M., and Lehman, J. (1982). Concurrent validity of the Woodcock-Johnson mathematics and reading subtests with a middle school learning disabled population. Educational and Psychological Research, 2, 173-179.

Breen, M., Lehman, J., and Carlson, M. (1984). Achievement correlates of the Woodcock-Johnson Reading and Mathematics Subtests, Key Math, and Woodcock Reading in an elementary aged learning disabled population. Journal of Learning Disabilities, 17, 258-261.

Coleman, M., and Harmer, W. (1985). The WISC-R and Woodcock-Johnson Tests of Cognitive Ability: A comparitive study. Psychology in the Schools, 22, 127-132.

Cummings, J., and Moscatto, E. (1984). Research on the Woodcock-Johnson Psycho-Educational Battery: Implications for practice and future investigation. School Psychology Review, 13, 33-40.

Cummings, J., and Sanville, D. (1983). Concurrent validity of the Woodcock-Johnson Tests of Cognitive Ability with the WISC-R: EMR children. Psychology in the Schools, 20, 298-303.

Dehorn, R., and Klinge, E. (1978). Factor analysis of the WISC-R with behavior disordered children. Journal of School Psychology, 16, 241-246.

Doehring, D., and Hoshko, I. (1977). Classification of reading problems by the Q technique of factor analysis. Cortex, 13, 281-294.

Doehring, D., Hoshko, I., and Bryans, B. (1979). Statistical classification of children with reading problems. Journal of Clinical Neuropsychology, 1, 5-16.

Doehring, D., Trites, R., Patel, P., and Fiedorowicz, C. (1981). Reading disabilities: The interaction of reading, language, and neuropsychological deficits. New York: Academic Press.

Epps, S., Ysseldyke, J., and Algozzine, B. (1983). Impact of different definitions of learning disabilities on the number of students identified. Journal of Psychoeducational Assessment, 1, 341-352.

Estabrook, G. (1984). A correlational analysis of the Wechsler Intelligence Scale for Children-Revised and the Woodcock-Johnson Tests of Cognitive Ability in a sample referred for suspected learning disabilities. Journal of Educational Psychology, 76, 1170-1177.

Fagan, T. (1981). Special education services and the school psychologist. Journal of Learning Disabilities, 14, 383-385.

Fisk, J., and Rourke, B. (1983). Neuropsychological subtyping of learning disabled children: History, methods, implications. Journal of Learning Disabilities, 16, 529-531.

Fleiss, J., and Zubin, J. (1969). On the methods and theory of clustering. Multivariate Behavior Research, April, 235-250.

Fletcher, J., and Satz, P. (1985). Cluster analysis and the search for learning disability subtypes. In B. Rourke (Ed.), Neuropsychology of Learning Disabilities. New York: Guilford Press.

Hall, R., Reeve, R. and Zareski, J. (1984). Validity of the Woodcock-Johnson Tests of Achievement for learning disabled students. Journal of School Psychology, 22, 193-200.

Hayes, W. (1973). Statistics for the social sciences (2nd Ed.). New York: Holt, Reinhart, and Winston.

Ipsen, S., McMillan, J., and Fallen, N. (1982). An investigation of the reported WISC-R/WJ discrepancy. Diagnostique, 13, 327-331.

Jastak, K., and Jastak, S. (1965). Wide Range Achievement Test. Wilmington, DE.: Guidance Associates.

Kaufmann, A. (1975). 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, 43, 135-147.

Kaufmann, A. (1981). The WISC-R and learning disabilities assessment: State of the art. Journal of Learning Disabilities, 14, 520-526.

Kaufmann, A. (1982). An integrated review of almost a decade of research on the McCarthy Scales. In T.R. Kratochwill (Ed.), Advances in School Psychology (Vol. 2). Hillsdale, NJ: Erlbaum.

Kaufmann, A., and Kaufmann, N. (1983). Interpretive manual for the Kaufman Assessment Battery for Children. Circle Pines, MN: American Guidance Service.

Kavale, K., and Forness, S. (1984). A meta-analysis of the validity of Wechsler scale profiles and recategorizations: Patterns or parodies? Journal of Learning Disabilities, 7, 136-156.

Keogh, B., and Hall, R. (1974). WISC subtest patterns of educationally handicapped and educable mentally retarded pupils. Psychology in the Schools, 11, 296-300.

Lennon, R. (1954). A comparison of results of three intelligence tests. Test Service Notebook #11. NY: Harcourt, Brace, and World.

Linn, R. (1975). Anchor Test Study: The long and short of it. Journal of Educational Measurement, 12, 201-214.

Loret, P., Seder, A., Bianchini, J., and Vale, C. (1975). Anchor Test Study: Equivalence and norms tables for selected reading achievement tests. Washington, D.C.: United States Government Printing Office.

Lorr, M. (1983). Cluster Analysis for the Social Scientist. San Francisco: Jossey Bass Publishing Co.

Mather, N. (1984). Performance of learning disability subjects and gifted subjects on the Woodcock-Johnson Psycho-Educational Battery and the WISC-R. Unpublished doctoral dissertation, University of Arizona, Tucson.

McCallum, R., Karnes, F., and Edwards, R. (1984). The test of choice for assessment for gifted children: A comparison of the K-ABC, WISC-R, and Stanford-Binet. Journal of Psychoeducational Assessment, 2, 57-63.

McCarthy, D. (1972). Manual for the McCarthy Scales of Children's Abilities. New York: Psychological Corporation.

McGrew, K. (1983). Comparison of the WISC-R and Woodcock-Johnson Tests of Cognitive Ability. Journal of School Psychology, 21, 271-276.

McGue, M., Shinn, M., and Ysseldyke, J. (1982). Use of cluster scores on the Woodcock-Johnson Psycho-Educational Battery with learning disabled students. Learning Disability Quarterly, 5, 274-287.

McKinney, J., and Feagans, L. (1983). Adaptive classroom behavior of learning disabled students. Journal of Learning Disabilities, 16, 360-367.

Naglieri, J. (1985). Normal children's performance on the McCarthy Scales, Kaufman Assessment Battery, and Peabody Individual Achievement Test. Journal of Psychoeducational Assessment, 3, 175-179.

Naglieri, J., and Anderson, D. (1985). Comparison of the WISC-R and K-ABC with gifted children. Journal of Psychoeducational Assessment, 3, 175-179.

Naglieri, J., and Haddad, F. (1984). Learning disabled children's performance on the Kaufman Assessment Battery for Children: A concurrent validity study. Journal of Psychoeducational Assessment, 2, 49-56.

Naglieri, J., and Pfeiffer, S. (1981). Correlations among scores on the WISC-R and Woodcock-Johnson Achievement Tests for learning disabled children. Psychological Reports, 49, 913-914.

Nie, N., Hull, C., Jenkins, J., Strinbrenner, K., and Bent, D. (1975). Statistical Package for the Social Sciences (2nd Ed). NY: McGraw Hill Book Co.

Petrauskus, R., and Rourke, B. (1979). Identification of subgroups of retarded readers: A neuropsychological multivariate approach. Journal of Clinical Neuropsychology, 1, 117-137.

Phelps, L., Rosso, M., and Falasco, S. (1984). Correlations between the Woodcock-Johnson and the WISC-R for a behavior disordered population. Psychology in the Schools, 21, 442-446.

Phelps, L., Rosso, M., and Falasco, S. (1985). Multiple regression data using the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability. Psychology in the Schools, 22, 46-49.

Porter, J. and Rourke, B. (1985). Socioemotional functioning of learning disabled children: A subtypical analysis of personality patterns. In B. Rourke (Ed.), Neuropsychology of learning disabilities. New York: Guilford Press.

Reeve, R., Hall, R., and Zakreski, R. (1979). The Woodcock-Johnson Tests of Cognitive Ability: Concurrent validity with the WISC-R. Learning Disability Quarterly, 2, 63-69.

Reilly, T., Drudge, O., Rosen, J., Loew, D., and Fischer, M. (1985). Concurrent and predictive validity of the WISC-R, McCarthy Scales, Woodcock-Johnson, and academic achievement. Psychology in the Schools, 22, 380-382.

Reschly, D. (1978). The validity of the WISC-R with minority students. Journal of School Psychology, 17, 257-262.

Reynolds, C., and Gutkin, T. (1980). Patterns of WISC-R scores for black children. Journal of Learning Disabilities, 13, 60-63.

Rourke, B. (1985). Overview of learning disabilities subtypes. In B. Rourke (Ed.). Neuropsychology of learning disabilities. New York: Guilford Press.

Rourke, B., and Finlayson, M. (1978). Neuropsychological significance of variations in patterns of academic performances: Verbal and visual-spatial abilities. Journal of Abnormal Child Psychology, 6, 121-133.

Rourke, B., and Strang, J. (1978). Neuropsychological significance of variations in patterns of academic performance: Motor, psychomotor, and tactile perception abilities. Journal of Pediatric Psychology, 3, 212-225.

Satz, P., and Morris, R. (1980). Learning disability subtypes: A review. In F. Pirozzolo and M. Wottrock (Eds.), Neuropsychological and Cognitive Processes in Reading. New York: Academic Press.

Siegel, S. (1956). Non-Parametric statistics for behavioral sciences. NY: McGraw Hill Book Co.

Snow, J., Cohen, M., and Holliman, W. (1985). Learning disability subgroups using cluster analysis of the WISC-R. Journal of Psychoeducational Assessment, 4, 391-397.

Strang, J., and Rourke, B. (1985). Arithmetic disability subtypes: The neuropsychological significance of specific arithmetical impairment in childhood. In B. Rourke (Ed.), Neuropsychology of learning disabilities. New York: Guilford Press.

Strawser, S., and Weller, C. (1985). Use of adaptive behavior discrepancy criteria to determine learning disabilities severity subtypes. Journal of Learning Disabilities, 18, 205-212.

Sweeney, J., and Rourke, B. (1978). Neuropsychological significance of phonetically accurate and phonetically inaccurate spelling errors in younger and older retarded spellers. Brain and Language, 6, 212-225.

Thompson, P., and Brassard, M. (1984). Validity of the Woodcock-Johnson Tests of Cognitive Ability: A comparison with the WISC-R in LD and normal elementary students. Journal of School Psychology, 22, 201-208.

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

van der Vlugt, H., and Satz, P. (1985). Subgroups and subtypes of learning disabled and normal children: A cross-cultural replication. In B. Rourke (Ed.), Neuropsychology of learning disabilities. New York: Guilford Press.

Woodcock, R., and Johnson, M. (1977). Woodcock-Johnson Psycho-Educational Battery. Hingham, MA: Teaching Resources Corporation.

Woodcock, R. (1980). A response to some concerns raised about the Woodcock-Johnson Psycho-Educational Battery. Leadership series in special education, Minneapolis Public Schools, 10, 107-117.

Woodcock, R. (1984). A response to some concerns raised about the Woodcock-Johnson I. The mean score discrepancy issue. School Psychology Review, 13, 355-362.

Ysseldyke, J., Shinn, M., and Epps, S. (1981). A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability. Psychology in the Schools, 18, 15-19.

Ysseldyke, J., Shinn, M., Mcgue, M., and Epps, S. (1981). Performance of learning disabled and low achieving students on the Wechsler Intelligence Scales for Children-Revised, and the Tests of Cognitive Abilities from the Woodcock-Johnson Psycho-Educational Battery. In W. Cruickshank and A. Silver (Eds.), Bridges to tomorrow: The best of ACLD. Syracuse, NY: Syracuse University Press.