RELIABILITY AND SPECIFICITY OF WISC-R SUBTESTS WITH LEARNING DISABLED BOYS

By

BARBARA CLARK

Bachelor of Arts
Framingham State College
Framingham, Massachusetts
1975

Master of Arts
University of Tulsa
Tulsa, Oklahoma
1978

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1987

RELIABILITY AND SPECIFICITY OF WISC-R

SUBTESTS WITH LEARNING DISABLED BOYS

Thesis Approved:

Paul G. Warden
Thesis Adviser

David S. Fournier

Charles R. Davis

[illegible]

Kathryn Castle

Norman N. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

I will be eternally grateful to the members of my com mittee who participated in the starting and completion of my dissertation. Dr. David Lane got me started. He recognized my reluctance and gave me words of wisdom - "Get on the ball!" Dr. Kay Bull provided momentum - "Formulate your problem statement, and I'll tell you if I approve!" Dr. Kathryn Castle gave me gentle guidance - "I think you should tone down this statement just a little." Dr. Bob Davis yanked me toward the finish line - "It's only snowing a little. Let's run to the library, now! Dr. Paul Warden, my major advisor, ran every step of the race with me, but not for me. I'll always appreciate that.

I'd like to thank George Moody for his expertise on the computer, and his amazing willingness to help a friend.

I want to give a very special thanks to my husband, Leo, and my children, Jason, Tim and Jennifer. They have demonstrated amazing patience, and unwavering support. I thank them for always believing that I could succeed. But most of all I thank them for not caring if I failed.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Special Education

Special education is a subsystem of regular education designed and mandated by Public Law 94-142 (Education For All Handicapped Children Act, 1975a) to provide a free and appropriate education to all handicapped children. Each state is held responsible for the implementation of this law and its amendments. In Oklahoma, the law is administered by the Oklahoma State Department of Education (SDE), Special Education Section (SES). The Special Education Section was created and implemented by the State Board of Education and functions to establish services for handicapped children throughout the state.

The Oklahoma SDE publishes the Policies and Procedures Manual for Special Education in Oklahoma (1985). Regulations implementing P.L. 94-142 are included in this manual. The manual's purpose is threefold:

1. To establish minimum standards for program approval.
2. To establish minimum standards for the determination of pupil eligibility.
3. To establish considerations which will lead to appropriate programming within the least restrictive environment.

The manual serves as a guideline for the implementation of P.L.94-142 by local education agencies. Hence, if a child is failing in the classroom in Oklahoma, and is referred for an evaluation, all procedures for the delivery of services to that child are spelled out in P.L. 94-142; the Oklahoma state guidelines for implementation of the law are mandated in the manual.

The Learning Disabled (L.D.)

In most states, more students are labeled learning disabled (L.D.) than any other special education category (Algozzine and Ysseldyke, 1986). Oklahoma is no exception. During the 1985-86 school year, the State of Oklahoma served 63,635 special education students; nearly 28,000 were classified as learning disabled, and placed in special education programs (Oklahoma SDE Statistical Report, 1986, submitted to Regional Education Service Centers).

According to P. L. 94-142 (Education For All Handicapped Children Act, 1975b), a learning disabled child has:

> ...a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. The term includes such conditions as perceptual hand-

> icaps, brain injury, minimal brain disfunction, dyslexia, and developmental aphasia. The term does not include children who have learning problems which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, or of environmental, cultural, or economic disadvantage (p. 13).

The following criteria are designated in 34 CFR 300.541 (Education For All Handicapped Children Act, 1975c) for determining the existence of a specific learning disability:

> The child does not achieve commensurate with his or her age and ability levels in one or more of the areas (listed below) when provided with learning experiences appropriate for the child's age and ability levels; and the team (specified in the law) finds that a child has a severe discrepancy between achievement and intellectual ability in one or more of the following areas: (i) oral expression; (ii) listening comprehension; (iii) written expression; (iv) basic reading skills; (v) reading comprehension; (vi) mathematical calculations; or (vii) mathematical reasoning.
>
> The team may not identify a child as having a specific learning disability if the severe discrepancy between ability and achievement is primarily the result of mental retardation, emotional disturbance, or environmental cultural or economic disadvantage (p. 49).

The law does not define "severe discrepancy." Each state, or in some instances each local district determines their own definition. In Oklahoma, the definition of "severe discrepancy" is left to the discretion of the local school district.

The tests used to diagnose L.D. students attempt to evaluate intelligence, achievement, perception, language, and neurologic functioning (Gearheart, 1977). The goal is to determine the child's strengths and weaknesses which can be used in educational planning.

The Individualized Education Program (IEP)

The law guarantees the identified child an Individualized Education Plan (Education For All Handicapped Children Act, 1975d). Individualized instruction is based on the premise that different children learn in different ways, that these differences can be identified through the psychoeducational assessment, and that instruction can accommodate these differences (Policies and Procedures Manual, 1985, p.29). Section 300.534 stipulates that a child's IEP be reviewed every three years, or more frequently if conditions warrent.

Results of psychoeducational assessments are viewed as sources of data for planning appropriate individualized instruction for each child; i.e. the child's "disorder in one or more of the basic psychological processes" is identified,

and instruction can be matched to learner characteristics (Kirk and Kirk, 1971).

Contents of the IEP must include:

(a) A statement of the child's present levels of educational performance.

(b) A statement of annual goals, including short term instructional objectives.

(c) A statement of the specific special education and related services to be provided to the child, and the extent to which the child will be able to participate in regular educational programs.

(d) The projected dates for initiation of services and the anticipated duration of the services.

(e) Appropriate objective criteria and evaluation procedures and schedules for determining, or at least on an annual basis, whether the short term instructional objectives are met.

The State of Oklahoma IEP form includes a place for a summary statement of the child's strengths and weaknesses (Policies and Procedures Manual, 1985, Form No. 6, p.75). According to this manual :

Information for these statements should be taken from teacher and parent observations, the psycho-educational evaluation and other pertinent sources. These statements should address a child's special learning skills

> and needs; i.e., verbal/language skills, motor skills, behavior, etc (p. 77).

The goal is to use the results of the evaluation, along with the other specified information, to determine the individual learning style of this child so that remediation of academic deficits will be optimal.

This is clearly in line with the Diagnostic-Prescriptive model that has dominated special education for the past twenty years (Arter & Jenkins, 1979). This model emphasizes the use of assessment information for determining underlying abilities; assessed strengths and weaknesses are used to devise instruction. Several authors attempted to design instruments that would tailor instruction to the located deficits (Arter & Jenkins, 1977; Bannatyne, 1968, 1974; Bateman, 1967; Cartelli, 1978; Cartwright, Cartwright, & Ysseldyke, 1973; Cronback & Snow, 1977; Ferinden, Jacobsen & Kovalinski, 1969; Frostig, 1967; Gunnison, Kaufman & Kaufman, 1982; Kirk, 1962; Kirk & McCarthy, 1961; Ysseldyke & Sabatino, 1973).

The diagnostic-prescriptive model differentiated special education from regular education; i.e. regular educator taught reading, writing and arithmetic, special educators focused on underlying psychological abilities and disabilities (Arter & Jenkins, 1979). Wepman (1967), for example, proposed that before deciding on a remedial reading

approach for a particular child, one must first understand his learning type, or his maximum modality. The purpose of diagnosis, according to Bannatyne (1968), was the remediation of deficits. The guiding rule should be, "remediate the deficit areas and reinforce through the intact areas (p. 28).

Kirk and Kirk (1971) felt that in order to develop remedial educational programs, you must first have the results of psychological tests to determine psychological strengths and weaknesses. Ferinden, Jacobson and Kovalinsky (1969) were convinced that data derived from the Wechsler Intelligence Scale for Children (WISC) protocol could be used to plan educational intervention. They published their first book, Educational Interpretation of the Wechsler Intelligence Scale For Children in 1969 and two of the authors updated it in 1974 (Jacobson and Kovalinski). These authors attempted to describe what each of the Wechsler Intelligence Scale for Children-Revised (WISC-R) subtests measure, the effects that a low subtest score might have on classroom performance, and suggestions as to how to remediate the deficit. The Oklahoma State Department of Education publishes a book, Oklahoma Prescriptive Handbook (1985) which includes a section called WISC-R Remediation Techniques. The section bears a striking similarity to Ferinden and Jacobson's work.

It appears that the Oklahoma State Department of Education agrees that the data derived from WISC-R protocols are

indicants of the child's strengths and weaknesses, and that remediation of these strengths and weaknesses enhance academic functioning.

The WISC-R and the Learning Disabled

Oklahoma presently has twenty Regional Education Service Centers (RESC). The centers are to offer school districts professional assistance in a variety of efforts aimed toward the improvement of instruction for students. Each Center provides core services. One of the major functions of the RESCs is to perform the individual psycho-educational evaluations as spelled out in the Policies and Procedures Manual on page 51:

A. Student Appraisals

1. Conduct educational screenings.

2. Provide diagnostic and evaluative services for students who are exhibiting learning problems or exceptionalities.

If an Oklahoma school child is referred for testing, and it is done by the RESC, then it will be done according to the SDE/SES rules and guidelines. Three year reevaluations are done according to the same regulations as the initial evaluations. The minimum test battery for the determination of a learning disability is spelled out by the SES (Policies and Procedures Manual, 1985, p.69). The minimum

test battery includes:

1. Measures of Intelligence
2. Measures of Achievement and Knowledge
3. Measures of Information Processing and Perceptual Impairment.

The procedure for reporting results is also mandated. Results of all testing are reported on standard SDE/SES forms (Policies and Procedures Manual, 1985, pgs. 70-73). In addition to test results, the reports must contain information on:

1. Areas of Strengths
2. Areas of Weaknesses
3. Additional Information
4. Suggestions

An inspection of 100 randomly selected reports of Oklahoma children who were evaluated by RESCs and ultimately placed in learning disability programs revealed that the Wechsler Intelligence Scale for Children-Revised (WISC-R) had been used in all of the evaluations. Ninety seven percent of the reports sent to the schools used WISC-R subtest scores to indicate the child's "strengths and weaknesses" on SDE Form No. 3. Ninety two percent of these reports suggested remediation of WISC-R deficits. The users of the report were referred to Section XX, <u>WISC-R Remediation Techniques</u> in the handbook published by the State Department of

Education entitled Oklahoma Prescriptive Handbook (1985). This section provides suggestions on how to remediate these subtest deficits. For example, if a child received a "low" score on information, the report recommended that this deficit be remediated. It was observed that examiners did not necessarily use Kaufman's suggestions for appropriate interpretation of subtests (Kaufman, 1979b); i.e. a subtest score should deviate from the child's mean scaled score by plus or minus three in order to speculate about strengths and weaknesses. According to Kaufman (1979b):

> even though scaled scores of 11 and 7 may differ significantly from each other, this finding is irrelevant and trivial if both of these scores do not differ significantly from a child's own mean. When a child's scaled scores do differ significantly from the verbal and/or performance means, only then do we have the right to speculate about strengths and weaknesses in abilities that are less global than verbal comprehension and perceptual organization (p. 9).

An examination of 100 randomly selected IEP's from five school systems serviced by the RESC revealed that the "strengths and weaknesses" reported on the RESC's report to the school were subsequently used on 72% of the IEP forms in the place designated for the child's strengths and weaknesses. This widespread use of the WISC-R subtests as in-

dicants of processing strengths and weaknesses has a long standing tradition with educators (Arter and Jenkins, 1979; Coles, 1978). However, use of these subtests as indicators of strengths and weaknesses assumes two things:

1. The subtests have long term stability with learning disabled children. If the subtests lack stability, then an educational program based on assessed strengths or weaknesses may be totally inappropriate.
2. The subtests measure discrete abilities; i.e., they measure some stable psychological process in addition to, but apart from, their contribution to the measurement of intelligence. If the subtests lack specificity, then an educational program based on test results may be meaningless.

Test-Retest Reliability

Central to the diagnostic-prescriptive model is the assumption that the abilities that underlie academic achievement are stable, nonephemeral individual traits. Hence, a test that claims to measure these traits should demonstrate both internal consistency and test-retest reliability for the population for which it is being used. Arter and Jenkins (1979) proposed that for the diagnostic prescriptive model to be effective, it is essential that the tests employed are relatively stable over time.

There is presently little or no conclusive information

to suggest that the WISC-R subtests have long term stability with the learning disabled (Covin, 1977; Gutkin, 1979; Saklofske, Schmidt and Yackulic, 1984; Smith, 1978; Vance, Blixt and Ellis, 1981).

Personal use with the WISC-R over a five year period with learning disabled children has led the author to the conclusion that subtest scores are unstable with this population. A comparison of childrens' WISC-R subtest scores during the process of triannual reevaluations revealed that it was common for the scores to vary considerably from test to retest. A child's subtest score may increase, even when it was clear that no deliberate effort had been made to remediate the previously low score. It was also common for a child's strength on one profile to appear as a weakness on a subsequent profile. There appeared to be no consistency in the direction of the change, and no obvious explanation for the lack of correspondence on test-retest measures. If these subtests lack adequate stability with this population, their use for long term educational recommendations is negligent and should be discontinued.

Acceptable test-retest reliability is difficult to define. Anastasi (1976) proposes .80 as the minimal reliability level. Nunnally (1967) suggests that the reliability level should be a function of the purpose for which the test is used. He proposes reliabilities of greater

than .90 and preferrably above .95 for tests upon which important educational decisions are based. Salvia and Ysseldyke (1978) are in agreement with Nunnally. They recommend reliabilities of .90 for making educational decisions about referred children. Arter and Jenkins (1979) compromised between the various recommended reliabilities and adopted .85 as the minimal reliability level and .90 as a desired level for a test.

Studies that have attempted to investigate WISC-R subtest reliabilities with the learning disabled have reported reliability coefficients as low as .22 (Vance, Blixt and Ellis, 1981; Covin, 1977). Although the methodology of these studies restrict interpretation, they none-the-less indicate that subtest reliabilities with learning disabled may be inadequate for their present use.

Low test-retest reliabilities should not be taken casually. Sedlack and Weener (1973) dramatized the pitfall of low correlation coefficients in an investigation of the ITPA subtest reliabilities:

> Suppose that the bottom 30 percent of first graders in a school is selected for a special remediation program based on their September score on a particular ITPA subtest; how many of this group would be selected for the program based on retesting five months later in February? Sixty-three percent of the group selected in

> September would also be selected in the February testing, but 37 percent diagnosed as "special" in September would be classified as "regular" in the February testing. More than one out of three of the judgements made on the first testing would be considered errors on the basis of the retest which correlates .70 (p. 117).

This exemplifies the grave mistake that can be made by using tests that lack long term stability.

Specificity

Specific variance of these subtests is defined as the amount of reliable, systematic variance not shared with other subtests and not due to random error. There is no formal empirical mandate for interpreting the uniqueness of subtests. Kaufman (1979b) proposes that a subtest's reliable specific variance should equal about .25 or more of the total variance, and it should also exceed its error variance. Kaufman (1979b) cautions the interpretation of subtest-specific skills. First, the proportion of common or shared variance for almost all subtests exceeds the specific variance. Second, although several of the subtests are deemed to have ample specificity others have only adequate specificity, and some have inadequate specificity. Specificity in some instances was a function of age. Kaufman also cautions the user that subtests with ample specificity

can only be interpreted if their scaled score differs from the child's mean by a minimum of three points. According to Kaufman, even a subtest with ample specificity should not be interpreted unless it deviates significantly from a child's verbal or performance scaled score means.

It is important to note that Kaufman's estimates of subtest specificity were based on reliability coefficients obtained from the standardization sample which did not specifically include learning disabled students. Their prevalence in the norm group is unknown; his results may not generalize to this population. Hence, the specificity of the WISC-R subtests with learning disabled children is unclear at this time. The reporting of information that may not be reliable or valid for a child may actually interfere with the child's right to "a free and appropriate education." Although we cannot always know what is appropriate for a child, we have a responsibility to attempt to use educational practices that have scientific integrity.

Problem Statement:

WISC-R subtest scores are being used as indicators of psychological processing strengths and weaknesses. These identified strengths and weaknesses are then used to plan long term educational remediation for learning disabled students. Use of the WISC-R subtests for this purpose necessitates that:

1. The subtests have long term stability with this population.

2. The individual subtests have adequate specific variance to warrant their use for the measurement of specific abilities or disabilities.

There is little or no evidence to suggest that these subtest scores are stable over time with this population. Since specific variance is a function of reliable variance, their use as discrete measures of psychological phenomena may be inappropriate with learning disabled children.

Hypothesis I: The WISC-R subtests scores do not have long term stability when used with learning disabled students.

Hypothesis II: The specific variance of the WISC-R subtests with learning disabled students is too low, according to Kaufman's (1979b) guidelines, for interpretation leading to making remediation decisions.

Hypothesis III: There will be no significant differences in WISC-R stability coefficients across age groups.

CHAPTER II

LITERATURE REVIEW

The WISC-R

The Wechsler Intelligence Scale For Children-Revised (WISC-R) was designed as a test of general intelligence (Wechsler, 1974). It differs from other measures of intelligence in two important ways: (1) it conceives of intelligence as a global, multidetermined and multifaceted entity and (2) it avoids singling out any one particular ability as more important than another. Wechsler characterized intelligence as global because it reflects the individual's behavior as a whole, and multifaceted because it is composed of abilities that are qualitatively differentiable; by measuring these separate abilities, we measure intelligence. A good intelligence test must probe as many of these abilities as possible. Hence, the rationale for the twelve subtests.

Norms for the WISC-R were based on the 1970 census data. The variables used were age, sex, race (white and nonwhite), geographic region, occupation of head of household, and urban-rural residence.

The sample included 100 boys and 100 girls at each of 11 age levels from 6 1/2 through 16 1/2 years of age. Whites,

blacks, American Indians and Orientals were included in the same proportion as in the census data. The sample was restricted to "normal" children, eliminating mental defectives or children with severe emotional disorders.

Reliability and Stability

The WISC-R Manual (Wechsler, 1974) presents split-half reliability coefficients for the Verbal, Performance and Full Scale I.Q.'s and for individual subtests. Average reliability coefficients for the subtests range from .70 to .86. (The average was computed using Fisher's z transformation). All coefficients are split half correlations with the exception of Digit Span and Coding. Test-retest correlations were computed for these two subtests using samples of 50 children for each age group over a one month interval.

Stability coefficients for the twelve subtests were computed using 303 children from the six selected age groups in the standardization sample. Subjects were retested after one month. Obtained coefficients ranged from .55 to .91. Coefficients corrected for the average variability of the normative age groups ranged from .63 to .92.

Validity

The WISC-R manual includes the results of correlating the WISC-R with other measures of intelligence. The correlation coefficients of the Verbal, Performance and Full Scale I.Q.'s with the Stanford-Binet are .71, .60 and .73 respectively. The correlation of the WISC-R Full Scale with the WPPSI Full Scale I.Q. is .82. Similar high correlations are reported for the two Verbal I.Q.'s and the two Performance I.Q.'s. The correlation between the WISC-R and WAIS Full Scale I.Q.'s is .95.

History of WISC-R Subtest Interpretation

In order to understand the present day use of the Wechsler Scales it is necessary to look at the historical perspective of diagnostic psychological testing.

The first of the Wechsler scales was the Wechsler-Bellevue (Wechsler, 1939). Wechsler viewed this test as having the potential not only for psychometric purposes, but as a clinicodiagnostic device to be used for differential diagnosis (Matarazzo, 1972). As expressed in Measurement of Adult Intellignece, (Wechsler, 1944, p.146):

> In point of fact, most intelligence examinations, when administered individually, make available a certain amount of data regarding the testee's mode of reaction,

his special abilities or disabilities and, not infrequently, some indications of his personality traits.

Wechsler went on to propose three diagnostic uses for the Wechsler-Bellevue. First, Wechsler suggested diagnostic interpretation of a verbal-performance discrepancy. He proposed that organic brain disease, psychosis, and psychoneuroses are clinical groups that score higher on verbal tests than non verbal. Psychopaths and mental defectives, on the other hand, score higher on the performance tests.

The second use proposed by Wechsler involved scatter analysis. He suggested that a difference of any more than two points between the subtest scale score and the mean for the subtests can be interpreted as significant. He looked at the performance "patterns" as diagnostic indicators of various pathological groups, mainly organic brain disease, schizophrenia, psychopathic personality, neurotics, and mental defectives. For example, he noted that with schizophrenia, Information and Vocabulary subtests were "high" and Arithmetic was "low". Hence, anyone that had this profile could be placed in the appropriate diagnostic category. This pattern analysis is essentially no different than our present day use of subtest scores to diagnose learning disabilities.

Wechsler also proposed diagnosis of severe psychiatric disorders by analyzing subtests. He theorized that severe

psychiatric disorder was characterized by a deterioration in mental ability, similar to the normal deterioration of aging. His experience with this population led him to formulate a ratio of tests that "hold up with age" (Information, Comprehension, Object Assembly, Picture Completion, and Vocabulary) as opposed to those that do not (Digit Span, Arithmetic, Digit Symbol, Block Design, Similarities, and Picture Arrangement).

The diagnostic use of intelligence tests gained impetus with the publication of Rapaport, Gill & Schafer's Diagnostic Psychological Testing (1945). In a monumental study, Rapaport looked at the scatter indices of 217 patients and 54 normal controls. Although his work was later criticized because of research design flaws, (Rabin and Guertin, 1951), it is credited with furthering clinical psychology in general, and the diagnostic utility of intelligence tests in particular (Matarazzo, 1972).

Although Rabin and Guertin (1951) concluded, rather pessimistically that "the scatter mountain led to a mouse," we can see that the interpretation of scatter is alive and well with those who use the Wechsler scales for the diagnosis of learning disabilities.

The Wechsler-Bellevue was eventually revised twice and is now known as the Wechsler Adult Intelligence Scale-Revised (WAIS-R). Wechsler eventually published a downward extention

of the WAIS, The Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949). The WISC was revised in 1974 and became the WISC-R.

The History of Learning Disabilities

Academic failure has typically been explained in terms of brain damage or dysfunction (Torgeson, 1986). The terms "minimal brain damage" and "neurologically impaired" were originally used to label children who are now called learning disabled. The label has changed, but the explanation for the cause of academic failure has stayed remarkably constant (Gearheart, 1977). The learning disabled child, according to this neuropsychological paradigm, has damage to specific brain functions.

According to neuropsychological theory all learning depends on the functional integrity of the central nervous system. (Hartlage, 1981; Hartlage and Telzrow, 1983; Obrzut, 1981; Obrzut and Hynd, 1983). If a child has learning problems, it is assumed the problem is the result of either naturally occuring variations in neurological substrata or damage to this substrate (Torgeson, 1986). Generalized or specific processes within the child are seen as deficient or dysfunctional and thus causative of learning problems (Quay, 1973). The locus of the problem is within the child. Remediation of academic deficits, then, requires fixing something

within the child, mainly his neurological processing deficits. If the deficit can be identified, then classroom instruction can be tailored to circumvent the processing problem.

Researchers attempted to identify the neurological deficiencies that caused learning disabilities. For example, in 1937, Orton wrote a book discussing his theory about the relationship between brain functioning and reading problems. He viewed dyslexia as the result of interfering competition between different brain hemispheres. Strauss and Lehtinen (1947) theorized about the relationship of brain injury to a particular behavioral syndrome. Kephart (1960) developed a theory relating delayed perceptual motor development to academic achievement. Kirk and Kirk (1971) proposed a relationship between psycholinguistic processing deficits and learning disabilities. The WISC, and subsequently WISC-R subtests fell into use for this purpose. Since its inception, the Weschler Intelligence Scale for Children (Wechsler, 1949) has been used for the identification of cerebral dysfunction (Belmont and Birch, 1966; Graham, 1952; Hartlage, 1981). It has become widely used for educational prescription. Remedial suggestions based on WISC-R subtest scores have flourished (Anderson, Kaufman & Kaufman, 1976; Ferinden,

Jacobsen, & Kovalinski, 1969; Fisk & Rourke, 1983; Glasser & Zimmerman, 1967; Jacobson & Kovalinsky, 1974; Willis & Banas, 1978).

Subtest Specificity

Implicit in the rationale for subtest interpretation is the assumption that a substantial part of a subtest's variance can be attributed to specific functions. Several investigators have attempted to determine the specificity of the Wechsler scales. Cohen (1959), who had previously investigated the factor structure of the Wechsler Bellvue (1952a, 1952b), factor analyzed the Wechsler Intelligence scales for Children, using the subtest's communality as an estimate of common variance. He concluded that a subtest's measurement function is most meaninglyfully described in terms of "g" and the primary abilities. Cohen considered the variance accounted for by the specific subtests to be "essentially uninterpretable" (p. 290).

Kaufman (1975) factor analyzed the WISC-R using Cohen's procedure for determining subtest specificity. He concluded that the analysis supported specific interpretation of the subtests, but urged caution, warning that "one cannot assume the child has an unusual ability or disability in a test without examining his scores on other relevant tests" (p.145).

Kaufman (1979b) later recomputed subtest specificity using squared multiples as estimates of common variance as advocated by Silverstein (1976). His computations for ages 6 1/2 to 16 1/2 revealed specific variance estimates of .19 to .51. Specificity was a function of both the age of the subject, and the particular subtest. In general, seven of the subtests were deemed to have ample specificity across age levels. Only the Object Assembly had inadequate specificity across all ages.

Neither Cohen (1959) nor Kaufman (1979b) adopted a hard and fast rule for the amount of specific variance required for interpretation. Both agreed, however, that the specific variance should equal about .25 or more of the total reliable variance. It must also exceed its error variance.

It should be noted that the specific variance of any WISC-R subtest is a function of the test's reliability, and the test's reliability was computed using a sample of children who were not specifically learning disabled. Interpretation of results with a learning disabled sample may therefore be less tenable, and perhaps totally inappropriate. As Kerlinger (1964) has indicated, a test must be reliable before it is interpretable.

Scholl (1985) has also addressed this issue and concluded that if the norm group of a test is significantly different from the person to be tested, then the results may not

be valid for that child. She stresses the importance of identifying the population on which the instrument was standardized.

Salvia and Ysseldyke (1978) have likewise stressed the importance of establishing whether or not the validity and reliability coefficients of selcted tests for the learning disabled are strong. They point out that if the reliability or validity is poor or unknown, the test should be used cautiously and with these limitations in mind. If the reliability of the WISC-R subtests is poor, then interpretation of individual subtests is foolish.

Wright and Isenstein (1977) argued against using the WISC-R with minority children because only 330 were included in the standardization sample. They reasoned that the inclusion of such a small number would have no significant impact on the test. The same argument applies to use of the WISC-R with learning disabled children. Although random selection of subjects would have perhaps allowed for the inclusion of a proportionate number in the standardization sample, the number would have been too small to impact on the test (Ysseldyke, Algozzine, Regan & Porter, 1980). Thus, WISC-R reliability and specificity may not generalize to the learning disabled.

Section 300.532 of Public Law 94-142 states that tests and other evaluation materials must be validated for the

specific purpose for which they are used (Education For All Handicapped Children Act, 1975f). We do not know how many L.D. children were in the WISC-R standardization sample. Hence, its use for this population is questionable, and perhaps contrary to the law's intent..

Galvin (1981) has pointed out another concern with using the WISC-R with the learning disabled. She argues that although the WISC-R is ordinarily a good predictor of academic achievement with the normal population, it has low predictive validity with the learning disabled, (perhaps due to low reliability). This is an issue that should not be ignored. Use of a test that may not be valid for the population on which it is being used has important ramifications; it may prevent the child from receiving an evaluation that more accurately represents the child's true ability/disability. It may also keep the child from receiving an individualized education program that is accurately "individualized".

When a child is referred for an evaluation, the teacher who referred the child wants educationally meaningful information so she can help the child in the classroom. To make recommendations based on a high or low subtest score that may or may not have ample specific variance to justify interpretation is negligent, and does the child an injustice. The widespread use of WISC-R subtests with the learning disabled

warrants the investigation of the reliability and specificity of these subtests for this population.

Stability

There is suprisingly little research on the stability of WISC-R subtests scores. Only two studies were found that specifically investigated the stability of each of the subtests when used with learning disabled children.

The first of these was conducted by Vance, Blixt and Ellis (1981). They investigated the stability of the WISC-R over a two year interval. Their sample included 30 EMH students and 45 learning disabled. The mean Full Scale I.Q. score for the group was 75.91. Results indicated that the Verbal, Performance and Full Scale I.Q.scores are stable for this population. Coefficients were .80, .91, and .88 respectively. Stability coefficients for the subtests, however, ranged from .53 to .80. Four out of the eleven subtests (36%) had correlations below .60. T-tests for the differences between means of the subtests for the two administrations were computed. Five significant differences were found; four reflected significant decreases in scores and one showed a significant increase. Decreases were found on Similarities, Vocabulary, Digit Span, and Block Design. Scores on Picture Arrangement increased. This contradicts the regression effects that are expected when the scores are

derived from subjects selected on the basis of extreme functioning; i.e., Full Scale I.Q. scores ranged from 55 - 116 with a mean of 75.91.

The authors concluded that the WISC-R Verbal, Performance and Full Scale I.Q. scores are reliable over time for learning disabled and retarded children. Other authors, however, would impose more stringent reliability requirements and conclude differently (Anastasi, 1976; Arter & Jenkins, 1979; Nunnally, 1967; Salvia & Ysseldyke, 1981; Sedlack and Weener, 1973). It should also be noted that the mean Full Scale I.Q. score for this sample was 75.91. This indicates a preponderance of low ability students. EMH students and low achieving (as opposed to learning disabled) students are characterized by their lack of change. Personal experience indicates that the learning disabled child, on the other hand, often demonstrates changeability on a day-to-day basis. Mixing these two groups in a study would obscure results for both groups.

Covin (1977) investigated the stability of the WISC-R with 30 nine year old L.D. children over a one day interval. All children were in the fourth grade, and all were from low income families. Mean Full Scale I.Q. for this group was 87.27. Coefficients were .83 for Verbal I.Q., .84 for Performance, and .85 for Full Scale I.Q. Stability coefficients for the subtests ranged from .22 on Comprehension to .86 on

Block Design. Four of the ten subtests had stability coefficients below .57. Eight of the subtest had coefficients of less than .73. There was a small increase (less than one scaled score) on seven of the subtests, and a decrease on two. These results are difficult to interpret. Anastasi (1976, p. 112) suggests that retesting individuals over a short period of time generally leads to an increase in scores due to practice effects. Right or wrong responses are apt to recur through sheer memory. She maintains that the two administrations of the test cannot be viewed as independent measures, so the correlations should be spuriously high. If this is so, we can conclude that the obtained coefficients in Covin's study would be lower over a longer time interval. The problem of test bias might have also impacted Covin's results. Subjects in this study were all from low income families. Reynolds and Gutkin (1982) have addressed the issue of test bias in educational assessment and concluded that minority group members are over-represented in special class placement. It is also well established that there is a relationship between low income and minority group status. It is possible that Covin's sample was over-represented by minority group members. If their placement was a function of their socioeconomic status, they would not necessarily be representative of the truly learning disabled child. Results may reflect this. Since socioeconomic status is known to be

a source of variance, it should have been controlled.

No other studies could be found that investigated the stability of individual WISC-R subtest scores. Yet these subtest scores are widely used to make long term educational recommendations. If these subtest scores are indeed unstable for learning disabled children, then their continued used is unethical.

Several researchers have investigated the stability of a hypothesized characteristic L.D. profile; i.e. Spatial Ability > Sequencing Ability > Verbal Comprehension > Acquired Knowledge (Gutkin, 1979; Saklofski, Schmidt & Yackulic, 1984; Smith, 1978; Smith, Coleman, Dokecki, & Davis, 1980; Yanagida & Furlong, 1984). Results of these studies indicate that these characteristic profiles remained stable over time, suggesting subtest stability. However, although group means stayed stable, the research design does not indicate how many children within these groups are represented by the group's results.

Alert to this problem, Smith, Coleman, Dokecki and Davis (1977) investigated the problem of analyzing data based on group means. They specifically attempted to address the proportion of individuals within group results that do not represent the results. In their study, they found that although group means indicated that their particular sample of children demonstrated a "typical" L.D. profile, analysis of

individual profiles revealed that 57% did not have the characteristic profile. The research on L.D. profile stability mentioned earlier would be fraught with the same problem. The group means may obscure the functioning of individual children within the group.

As mentioned earlier, use of the WISC-R with learning disabled children presents problems of interpretation because it was not normed on L.D. children. Gutkin (1979) has investigated the use of the WISC-R with an atypical population. He specifically questioned the diagnostic utility of the WISC-R Bannatyne patterns for L.D. children from ethnic minority backgrounds. His study included 53 Caucasian and 87 Mexican-American children. All children had been identified as learning disabled. His results show that the Mexican-American children did not display the typical Bannatyne pattern generally predicted. He also concluded that although group means for the Caucasian L.D. children followed the typical Bannatyne pattern, that only 30% of these children actually demonstrated the pattern. His results call into serious doubts the use of the WISC-R with atypical populations, and the ramifications of making statements about individual children based on group means.

A review of the literature led this author to the conclusion that we currently know little about the stability and specificity of WISC-R subtests with learning disabled

children. Although Kaufman (1979b) suggests that WISC-R subtests have ample specificity for interpretation, his results are based on reliability coefficients derived from non-learning disabled children. They may not generalize to the learning disabled.

Results of the stability research are difficult to interpret. In the two studies that specifically investigated individual subtest stability, the samples limit interpretation. Covin's sample contained only children from low income families. Vance, Blixt and Ellis' sample included a preponderance of low ability students. Hence, interpretation and generalization of results is problematic.

Studies on "profile stability" indicate that the L.D. profile remains stable over time, indicating subtest stability. However, research results for profile stability are based on group means. A large proportion of individual children are not necessarily represented by group results. Hence, at the present time, reliability and specificity of the WISC-R subtests with L.D. children is unclear. Yet these subtests are widely used for important educational decisions. Information on the reliability and specificity of these sutests with the learning disabled is clearly needed.

CHAPTER III

METHODS

The goal of the present study was to determine the long term stability and specificity of the WISC-R subtests when used with the learning disabled student. Stability as a function of age/grade of the student was also tested.

Subjects

Protocols of 200 learning disabled children were selected from the files of 10 school districts in Tulsa and Okmulgee counties. Because the protocols came from different school districts, the subjects had been classified as "learning disabled according to various definitions of "severe discrepancy". However, all children in the study were a minimum of 30% below their expected achievement level; most were 50% or more below.

The goal of the sampling was to obtain four groups of 50 children initially tested in the third, fourth, fifth, and sixth grades, respectively, and retested in approximately three years. Children who were initially tested and placed in first and second grades were eliminated from the sample in an effort to control for the effects of those who were erroneously placed due to developmental delay rather than a

"true" learning disability. Since the majority of L.D. children are placed by the time they are in sixth grade, stability coefficients for these age groups were of paramount interest. Each of the four groups were ultimately comprised of children further restricted by the following standards:

1. All children selected had been in a learning disabilities program for a minimum of three years. None of the children had repeated any grades.
2. The WISC-R was used in the initial evaluation and in the subsequent triannual reevaluation.
3. The Full Scale I.Q. was 85 or above in an effort to eliminate the slow learner from the sample.
4. The same psychometrist administered the test on both occasions in an effort to control for examiner effects.
5. Sampling was limited to urban school districts. Urban is defined as communities with 5000 or more inhabitants. Rural school districts were eliminated because of the difficulty of obtaining sufficient numbers of protocols from these districts to meet sampling requirements.
6. Because of the difficulty of obtaining sufficient numbers of female L.D. students that met sampling requirements, only males were included in the sample.

7. Only students whose Individualized Education Plan did not include remediation of WISC-R deficits were included in the sample. This was determined according to the results of a questionnaire that was submitted to elementary L.D. teachers in the communities included in the sampling. (See Appendix for a copy of the questionnaire). Children were included in the sample if the child's teacher at the time of the initial Individualized Education Plan answered "no" to all three questions on the questionnaire. This restriction was included in an attempt to eliminate the effects of treatment to explain increases in subtest scores and subsequent effects on reliability.

Protocols were randomly selected, using a table of random digits, until the four groups of children who met the above criteria were obtained.

Procedure and Data Analysis

Verbal, Performance, and Full Scale I.Q. scores and subtest scores were recorded for all children for their initial evaluation and subsequent triannual reevaluations. Stability coefficients between the first and second testing were computed for I.Q. scores and subtest scores across the four groups (N = 200) and for each of the four separate groups

(N = 50 each). A test for independent correlations was used to test for the significance of the difference between the correlations for the different age groups.

Using the reliability coefficients computed for the subtests, the specific variance of the subtests was evaluated using the method proposed by Silverstein (1976), and later adopted by Kaufman (1979b); both proposed the use of squared multiples as estimates of common variance. The difference between the common variance and the reliable variance was attributed to specific variance. They further stipulated that to be interpretable, the specific variance must exceed the error variance.

CHAPTER IV

RESULTS

It was hypothesized that the WISC-R subtests do not have long term stability when used with learning disabled boys. Stability coefficients for the 10 WISC-R subtests for each of four groups in the present study confirmed this hypothesis. Results revealed coefficients that ranged from .18 to .59. Stability coefficients for the Verbal, Performance, and Full Scale I.Q. scores for the four groups ranged from .46 to .68 (see Table 1). Stability coefficients across the four groups, for each subtest, were also computed. Coefficients ranged from .35 to .49 for the subtests. Reliabilities for the Verbal, Performance and Full Scale I.Q. were .59, .54 and .57, respectively. Results are presented in Table 2. None of these correlations approach the desired standard of .80 to .95 suggested by several authors (Anastasi, 1976; Arter & Jenkins, 1979; Nunnally, 1967; Salvia and Ysseldyke, 1981; Sedlack and Weener, 1973).

It was hypothesized that there would be no significant differences between the stability coefficients for the four different age/grade groups in the present study. Tests for the difference between independent correlations confirmed this hypothesis. There were no significant differences be-

tween the correlations for the different age groups at the alpha .05 level. For this particular sample, stability did not seem to be a function of the age of subjects at the time of initial placement.

TABLE I

Stability Coefficients Between Two Testings For Four Groups

Subtest	Grade 3 - 6 N=50	Grade 4 - 7 N=50	Grade 5 - 8 N=50	Grade 6 - 9 N=50
Info.	r = .36	.49	.31	.30
Simil.	r = .30	.22	.36	.51
Arith.	r = .30	.51	.52	.50
Vocab.	r = .36	.50	.44	.58
Comp.	r = .32	.42	.51	.37
PicCom.	r = .56	.42	.50	.46
PicAr.	r = .59	.32	.46	.40
BlDes.	r = .51	.27	.37	.35
ObjAs.	r = .55	.32	.41	.44
Cod.	r = .46	.50	.18	.51
V.I.Q.	r = .46	.68	.58	.62
P.I.Q.	r = .62	.52	.55	.48
FS.I.Q.	r = .58	.64	.56	.48

TABLE II

Test Retest Reliabilities

For 200 Learning Disabled Boys

Subtests	r	Subtests	r
Information	.35	Picture Completion	.49
Similarities	.36	Picture Arrangement	.44
Arithmetic	.46	Block Design	.49
Vocabulary	.46	Object Assembly	.43
Comprehension	.41	Coding	.43
	Verbal I.Q.	.59	
	Performance I.Q.	.54	
	Full Scale I.Q.	.57	

It was hypothesized that the WISC-R subtests lack adequate specific variance to warrant meaningful interpretation. According to Kaufman (1979b), to be interpretable, specific variance must account for at least .25 of the reliable variance and exceed the error variance. Results of the present study revealed that none of the subtests met these criteria (See Table 3). Each subtest contained essentially uninterpretable specific factor variance. Most of the reliable variance was accounted for by common factor variance. A large amount of each subtest's total variance was error variance.

TABLE III

Variance Components of the WISC-R Subtests

Subtest	Reliable	Common	Error	Specific
Info.	.35	.44	.65	-.09
Simil.	.36	.32	.64	.04
Arith.	.46	.20	.54	.26
Vocab.	.46	.48	.54	-.02
Comp.	.41	.43	.59	-.02
Pic. Comp.	.49	.32	.51	.17
Pic. Arr.	.44	.25	.56	.19
Bl. Des.	.49	.44	.51	.05
Obj. Asm.	.43	.33	.57	.10
Coding	.41	.22	.59	.19

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Interpretation of WISC-R subtest scores for both diagnosis and prescription has a long standing tradition. Wechsler viewed his tests as clinicodiagnostic devices capable of differential diagnosis (Matarazzo, 1972). Scatter analysis has been used to diagnose a variety of dysfunctions. Although an extensive review of scatter analysis by Rabin and Guertin (1951) led them to conclude that the whole effort had been fruitless, profile analysis continued, and interpretation of deviant subtests became more entrenched. They are presently being used as indicators of processing strengths and weaknesses with L.D. children. Identified strengths and weaknesses are then used to plan long term instruction with the learning disabled. Those who interpret WISC-R subtests do so with the belief that subtest strengths and weaknesses and resultant remediation will ultimately lead to improved academic achievement. Subtest interpretation is based on the assumption that the subtests have long term stability with the learning disabled, and that the subtest have adequate specific variance for interpretation.

The purpose of the present study was to determine whether the WISC-R subtest have adequate stability and

specificity for the purposes for which they are used.

Test-Retest Reliability

Adequate stability was defined according to what other authors had deemed acceptable. Anastasi (1976) suggested .80 as the minimal reliability level. Nunnally (1967) and Salvia and Ysseldyke (1981) proposed reliabilities of .90 for tests upon which important educational decisions are to be made. Arter and Jenkins (1979) investigated test-retest reliabilities of many of the instruments used with the Differential Diagnosis- Prescriptive Teaching model. They proposed that for the model to be effective, the tests that are used must produce results that are relatively stable over time, and decided on .85 as an acceptable level.

Results of the present study indicate that test-retest reliabilities with learning disabled boys are significantly lower than those called for in the literature. Low reliabilities are difficult to explain. Statistical regression to the mean is often a viable explanation for low test-retest correlations. Examination of a sample of 200 individual pairs of scaled scores from 10 subtests in the present study revealed that the preponderance of scores did not regress toward the population mean of the various subtests. Forty-three percent moved toward the mean, but fifty-two percent moved away from, and the others stayed the same.

two percent moved away from, and the others stayed the same. The direction of change appears random. Saklofske, Schmidt and Yakulic (1984) and Vance, Blixt and Ellis (1981) found similar results.

Regression effects are most often associated with extreme functioning. It should be noted that the minimum I.Q. score for the present sample was 85. The maximum was 118. The mean test-retest Full Scale I.Q.scores for the 200 subjects were 92.2 and 92.8. Although these scores do not represent the average WISC-R Full Scale I.Q. score, neither can they be described as extreme. Statistical regression does not seem to be a viable explanation for the low stability coefficients in the present study.

Low test-retest reliabilities are often explained as a result of an inappropriately long time interval between test and retest. The optimum time interval is difficult to determine. According to Bellack and Hersen (1984) the appropriate time interval between tests should be determined by the use of the test. If we are measuring bodily functions such as pulse rate or blood pressure, then the test-retest interval would need to be appropriately brief. However, if we are interested in measuring stable psychological traits, then we would need the measurement instrument to demonstrate stability over a longer time. Intelligence is generally considered to be a stable trait. Extremely important long term

decisions are made about learning disabled children based on intelligence scores. Children who are evaluated in our schools, and placed in Special Education, are seldom re-evaluated in less than three years. We must be confident that the instument that we use to measure these children can demonstrate acceptable stability. Results of the present evaluation suggest that the WISC-R is not such an instrument.

Specificity

The WISC-R subtests are consistently used as indicators of processing strengths and weakness with the learning disabled. Their use for this purpose necessitates that the subtests have adequate specific variance. The criteria for adequate specific variance requires that the subtests specific variance equal .25 or more of the total variance, and it should also exceed its error variance (Kaufman, 1979b). Results of the present study show that in all instances, the error variance associated with the subtests exceeds the reliable variance. Results also reveal that in all but the Arithmetic subtest, most of the reliable variance is common factor variance. It becomes a meaningless exercise to talk about specific variance.

It should also be noted that items on the WISC-R subtests were not factor analytically derived. Hence, results of a second order factor analysis may contain even less

specific variance than the present results indicate.

Kaufman (1979b) has also argued that interpretation of WISC-R subtests is inappropriate unless the subtest score deviates from the child's mean scaled score by plus or minus three. However, when the subtests' correlation coefficients are as low as the present study indicates, then the reliability of the difference between subtests becomes extremely low, or even nonexistant. Interpretation of high and low scores on a profile becomes not only inappropriate, but negligent. Results of the present evaluation indicate that the interpretation of individual subtests as indicants of psychological processing is not based on sound empirical evidence. If subtests are unreliable, yet used in educational planning, than the child who depends on us for help is maligned.

Section 300.532 of Public Law 94-142 states that tests must be reliable and valid for the specific purpose for which they are to be used. The intent of the law is to provide an appropriate education for all children. This should be the intent of those of us who work with children.

Recommendations

Results of the present study revealed Verbal I.Q. scores lower than Performance for all groups on both test and retest (See Table 5). These findings are consistent with previous

unexplained. In the present study, 71% of the students demonstrated the Verbal < Performance discrepancy at the time of the initial testing. At the time of re-test, 94% of these same students still demonstrated the Verbal < Performance discrepancy. Because of the relationship between Verbal I.Q. and academic achievement, research in this area seems essential. Early identification of low verbal skills and ultimate preventive intervention should be a possibility.

TABLE IV

Mean I.Q. Scores for Four Groups

Test-Retest

	Grade 3	6	Grade 4	7	Grade 5	8	Grade 6	9
V.I.Q.=	89.8	90.0	86.9	87.7	87.5	88.1	86.8	86.8
P.I.Q.=	100.0	98.6	100.0	99.6	98.3	98.6	99.9	97.6
FS.I.Q.=	93.9	93.3	92.3	92.3	91.9	92.1	92.1	91.0

Last year, Oklahoma served nearly 28,000 learning disabled children and 11,000 mentally retarded students. Most of these children were assessed with the WISC-R. Results of the present evaluation indicate that the WISC-R lacks long term stability with the learning disabled. Long term stability with EMH students should be investigated. Mislabeling of EMH students with an instrument that lacks reliability with this population can have serious legal ramifications. It also does the child an injustice. Use of the other Wechsler scales with special populations should also be investigated.

Many other instruments are used for diagnostic and remedial purposes with children who are failing in the classroom. Important decisions are based on results. It is essential that reliability and validity studies be done with these instruments for the populations on which they are to be used. For example, the Woodcock Johnson Psychoeducational Battery is widely used in Oklahoma for determining a child's achievement level. A child is most often placed in a learning disability program based on the discrepancy between his I.Q. score and his achievement level. It is entirely possible that many children are placed in Special Education based on the discrepancy between two unreliable scores. It is also entirely possible that very few of these children will be re-evaluated in less than three years. The

reliability of the Woodcock Johnson, and other instruments used with special populations needs to be determined.

The L.D. child has historically been described as having a processing deficit. Identified strengths and weaknesses are used in educational planning. Newcomer and Hammill (1975), and Arter and Jenkins (1977) investigated the relationship between specific abilities and academic instruction. Both concluded that the consistent failure of research to confirm Diagnostic Prescriptive teaching "raises important questions as to its value." It appears time to accept that this model has not worked. We need to concentrate future research toward identification of relationships between specific teaching methods and academic success. Warner and Bull (1986) propose that special educators, like all educators, need to move toward indentifying and integrating a system of educational thought to consistently operate from. These authors, along with others, are concerned that identification of the L.D. child doesn't necessarily lead to successful intervention. Algozzine and Ysseldyke (1986) propose that all children who are failing in the classroom should receive special instruction. Labeling isn't necessary. The "special" education they receive should be based on sound educational strategies that we know work with all children. If a child is failing, he/she needs special help. But if a label is insisted upon in order to receive services,

let's label the child "Teaching Disabled" We can then focus research efforts on identification of teaching disabilities, and allocate funds for enhancing teaching deficits. A profile of teaching strengths and weaknesses may be more diagnostically significant than the child's profile of processing strengths and weaknesses. Although this is somewhat tongue-in-cheek, it makes a point.

If we are to look for changes in the "misuse" of the WISC-R and other testing instruments, conscientious psychometrists and school psychologists must set the pace. They must stop using the WISC-R for purposes that lack emperical justification. WISC-R subtest interpretation has failed to provide information that relates to sound educational strategy. There is growing evidence that the information it does provide is unreliable. Its use for this purpose should be discontinued.

References

Algozzine, B., & Ysseldyke, J. E. (1986). Future of the L.D. field: Screening and diagnosis. Journal of Learning Disabilities, 19, 394-398.

Anastasi, A. (1976). Psychological testing. New York: MacMillan.

Anderson, M., Kaufman, A. S., & Kaufman, N. L. (1976). Use of the WISC-R with the learning disabled population: Some diagnostic implications. Psychology in the Schools, 13, 381-385.

Arter, J., & Jenkins, J. (1977). Examining the benefits and prevalence of modality considerations in special education. Journal of Special Education, 11, 281-298.

Arter, J. & Jenkins, J. (1979). Differential diagnosis, prescriptive teaching. A critical appraisal. Review of Educational Research, 49, 517-555.

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 WISC-R scaled scores. Journal of Learning Disabilities, 7, 272-273.

Bateman, B. (1967). Three approaches to diagnosis and remediation. Academic Therapy Quarterly, 2, 215-222.

Bellack, A. S., & Hersen, M. (1984). Research methods in clinical psychology. New York: Pergamon Press.

Belmont, L., & Birch, H. (1966). The relation of time of life to behavioral consequences in brain damage. Journal of Nervous and Mental Disorders, 131, 91-97.

Cartelli, L. M. (1978). Paradigmatic language training for learning disabled children. Journal of Learning Disabilities, 11, 313-318.

Cartwright, G. P., Cartwright, C. A., & Ysseldyke, J. E., (1973). Two decision models: Identification and diagnostic teaching of handicapped children in the regular classroom. Psychology in the Schools, 10, 4-11.

Cohen, J. (1952a). A factor -analytically based rationale for the Wechsler-Bellvue. Journal of Consulting Psychoogy, 16, 272-277.

Cohen, J. (1952b). Factors underlying Wechsler-Bellvue performance of three neuropsychiatric groups. Journal of Abnormal and Social Psychology,47, 359-365.

Cohen, J. (1959). The factorial structure of the WISC at age 7-6, 10-6, and 13-6. Journal of Consulting Psychology, 21, 283-290.

Coles, G. S. (1978). The learning disabilities test battery: Empirical and social issues. Harvard Educational Review, 48, 313-340.

Covin, T. (1977). Stability of the WISC-R for 9 year olds with learning disabilities. Psychological Reports, 40, 1297-1298.

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods, New York: Irvington Publishers.

Education For All Handicapped Children Act (1975a). United States Statutes at Large, 89, 773-796.

Education For All Handicapped Children Act (1975b). 34 CFR 300.5.

Education For All Handicapped Children Act (1975c). 34 CFR 300.541.

Education For All Handicapped Children Act (1975d). 34 CFR 300.340 - 300.349.

Education For All Handicapped Children Act (1975e). 34 CFR 300.532.

Ferinden, W. E., Jacobson, S. & Kovalinski, J. (1969). Educational interpretation of the Wechsler Intelligence Scale for Children. New Jersey: Remediation Associates.

Fisk, J. L., & Rourke, B.P. (1983). Neurological subtyping of learning-disabled children: History, methods, and implications. Journal of Learnig Disabilities, 16, 529-531.

Frostig, M. (1967). Testing as a basis for educational therapy Journal of Special Education, 2, 15-34.

Galvin, G. (1981). Uses and abuses of the WISC-R with the learning disabled. Journal of Learning Disabilities 4, 326-329.

Gearheart, B. R. (1977). Learning disabilities: Educational strategies, Saint Louis: The C. V. Mosby Company.

Glasser, A. J., & Zimmerman, I. L. (1967). Clinical interpretation of the Wechsler Intelligence Scale for Children. New York: Grune and Stratton.

Graham, E. E. (1952). Wechsler Bellvue and WISC-R scattergrams of unsuccessful readers. Journal of Consulting Psychology, 16, 268-271.

Gunnison, J., Kaufman, A. S., & Kaufman, N. L. (1982). Read ing remediation based on sequential and simultaneous processing. Academic Therapy, 17, 297-307.

Gutkin, T. B. (1979). Bannatyne patterns of caucasian and Mexican-American children. Psychology in the Schools, 16, 178-182.

Hartlage, L. (1981). Neurological assessment techniques. In C. R. Reynolds and T. Gutkin (Eds.), Handbook of school psychology, (pp.296-313). New York: Wiley.

Hartlage, L., & Telzrow, C. F. (1983). Neuropsychological basis of educational intervention. Journal of Learning Disabilities,16, 1983.

Jacobson, S., & Kovalinsky, T. (1974). Educational interpretation of the WISC-R, Linden, New Jersey: Remediation Associates.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. Journal of Consulting and Clinical Psychology, 43, 135-147.

Kaufman, A. S. (1979). WISC-R research: Implications for interpretation. School Psychology Digest, 8, 5-27.

Kephart, N. C. (1960). The slow learner in the classroom. Columbus, Ohio: Charles E. Merrill.

Kerlinger, F. N. (1964). Foundations of behavioral research, New York: Holt, Reinhart, and Winston.

Kirk, S. A. (1962). Educating exceptional children. Boston: Houghton Mifflin.

Kirk, S. A., & Kirk, W. D. (1971). Psycholinguistic learning disabilities: Diagnosis and remediation. Urbana, Illinois: University of Illinois Press.

Kirk, S. A., & McCarthy, J. J. (1961). The ITPA - An approach to differential diagnosis. American Journal of Mental Deficiency, 66, 399-412.

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

Newcomer, P. & Hammill, D. D. (1975). ITPA and academic achievement: A survey. Reading Teacher, 28, 731-741.

Nunnally, J. C. (1967). Psychometric theory. New York: McGraw Hill.

Obrzut, J. E. (1981). Neuropsychological assessment in the schools. School Psychology Review, 10, 331-342.

Obrzut, J. E., & Hynd, G. W. (1983). The neurobiological and neuropsychological foundations of learning disabilities. Journal of Learning Disabilities 16, 529-531.

Oklahoma Prescriptive Handbook, (1982). Oklahoma City: State Department of Education Printing Service.

Oklahoma State Department of Education, (1986). Number of children receiving special education and related services. Oklahoma City: State Department of Education Printing Services.

Orton, S. T. (1937). Reading, writing and speech problems in children. New York: Norton.

Policies and procedures manual, (1985). Oklahoma City: State Department of Education Printing Service.

Quay, H. C. (1973). Special education: Assumptions, techniques, and evaluative criteria. Exceptional Children, 40, 165-170.

Rabin, A. I. & Guertin, W. H. (1951). Research with the Wechsler-Bellevue test, 1945-1950. Psychological Bulletin, 48, 211-248.

Rapaport, D., Gill, M., & Schafer, R. (1945). Diagnostic, psychological testing (2 Vols.). Chicago: Year Book Publishers.

Reynolds, C. R., & Gutkin, T. (1982). Handbook of school psychology, New York: John Wiley and Sons.

Salvia, J. & Ysseldyke, J. C. (1978). Assessment in special and remedial education. Boston: Houghton Mifflin.

Saklofske, D. H., Schmidt, H., & Yackulic, R. (1984). Variations in WISC-R patterns of learning disabled. Perceptual and Motor Skills, 59, 415-421.

Scholl, G. T. (Ed.). (1985). The school psychologist and the exceptional child, Reston, Va.: Council for Exceptional children.

Sedlak, R. A. & Weener, P. (1973). Review of research on the Illinois Test of Psycholinguistic Abilities. In. L. Mann & D. Sabatino (Eds.) The first review of special education, New York: Grune & Stratton.

Silverstein, A. B. (1976). Variance components in the subtests of the WISC-R. Psychological Reports, 39, 1109-1110.

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

Smith, M., Coleman, J. M., Dokecki, P. R., & Davis, E. E. (1980). Recategorized WISC-R scores of learning disabled students. Journal of Learning Disabilities, 13, 60-63.

Strauss, A., & Lehtinen, L. (1947). Psychopathology and education of the brain injured child. New York: Grune and Stratton.

Torgesen, J. K. (1986). Learning disabilities: Its current state and future prospects. Journal of Learning Disabilities, 399-402.

Vance, B., Blixt, S., & Ellis, R. (1981). Stability of WISC-R for a sample of exceptional children. Journal of Clinical Psychology, 37, 397-399.

Warner, M. M., & Bull, K. S. (1986). Grounding LD definitions and practices in systems of educational thought. Journal of Learning Disabilities, 19, 139-144.

Wechsler, D. (1939). Measurement of Adult Intelligence, (2nd ed.). Baltimore: Williams and Wilkins.

Wechsler, D. (1944). Measurement of adult intelligence, (3rd ed.). Baltimore: Waverly Press.

Wechsler, D. (1949). WISC Manual, New York: Psychological Corporation.

Wechsler, D. (1974). WISC-R Manual, New York: Psychological Corporation.

Wepman, J. M. (1967). The perceptual basis for learning. In E. C. Frierson & W. B. Barbe (Eds.) Educating children with learning disabilities: Selected readings, New York: Appleton-Century-Crofts.

Willis, I. H., & Banas, N. (1978). Prescriptions from WISC-R patterns. Academic Therapy 13, 491-495.

Wright, B. J. & Isenstein, V. R. (1977). Psychological tests and minorities. Rockville, Md.: NIMH, DHEW Publication No, (ADM) 78-482.

Yanagda, E., & Furlong, M. J. (1984). Stability of learning disabled children's WISC-R Bannatyne categories. Perceptual and Motor Skills, 59, 832.

Ysseldyke, J. E., Algozzine, B., Regan, R. & Potter, M. (1980). Technical adequacy of tests used in simulated decision making. Psychology in the Schools, 17, 202-209.

Ysseldyke, J. E., & Sabatino, D. (1973). Toward the validation of the diagnostic-prescriptive model. Academic Therapy, 8, 415-422.

Appendix

Questionnaire to Elementary L.D. Teachers

Dear Special Education Teacher,

I am interested in determining whether or not special education teachers believe that remediation of WISC-R deficits is an important part of an Individualized Education Plan for learning disabled children. I need your help. Please answer the following questions and return this in the self-addressed stamped envelope provided. All responses are strictly confidential. Thank you for your help.

Please answer by circling the appropriate response.

1. I believe that WISC-R subtest scores are helpful in planning instruction for L.D. students. Yes No

2. I believe that the remediation of WISC-R deficits should be an important part of a learning disabled child's individualized education program. Yes No

3. I actively attempt to remediate WISC-R deficits.

Yes No

VITA

Barbara Clark

Candidate for the Degree of

Doctor of Philosophy

Thesis: RELIABILITY AND SPECIFICITY OF WISC-R SUBTESTS WITH LEARNING DISABLED BOYS

Biographical:

Personal Data: Born in Acton, Mass. Presently married with three children.

Education: Graduated from Maynard High School, Maynard, Mass. Received Bachelor of Arts Degree in Psychology from Framingham State College, December, 1975. Completed a Master of Arts degree in Psychology at the University of Tulsa in December, 1978. Completed the requirements for the Doctor of Philosophy at Oklahoma State University in December, 1987.

Professional Experience: Teaching Assistant, Dept. of Psychology, Framingham State College, 1974. Research Assistant, Dept. of Psychology, University of Tulsa, 1976. Psychometrist, Oklahoma State Dept. of Education, 1980 - 1985. School Psychologist, Glenpool Public Schools, 1985-1986. Chemical Dependency Counselor, St. John Medical Center, Tulsa, 1986-present.

Volunteer Experience: Tulsa Public Schools High Challenge Program, Children's Medical Center, Tulsa Speech and Hearing Association, The Women's Center.