LEARNING DISABILITY ELIGIBILITY: COMPARING THE WISC-R AND THE WOODCOCK-JOHNSON TESTS OF COGNITIVE ABILITIES-REVISED

By

DENNIS ANDREW TOMLINSON

Bachelor of Science
University of Oklahoma
Norman, Oklahoma
1973

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1986

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
July, 1993

OKLAHOMA STATE UNIVERSITY

LEARNING DISABILITY ELIGIBILITY: COMPARING
THE WISC-R AND THE WOODCOCK-JOHNSON TESTS
OF COGNITIVE ABILITIES-REVISED

Thesis Approved:

Joseph Pearl
Thesis Advisor

Paul G Ward

D. McCoy

Dale R. Fuqua

Patricia A. Self

Thomas C. Collins
Dean of the Graduate College

ACKNOWLEDGEMENTS

I wish to acknowledge my wife's contribution to achievements in my life. Thanks also to the ABSED faculty who have contributed to my learning.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Background of the Problem

The phrase learning disability evokes multiple images. Groups of characteristics are associated with these images and they define the condition. The multiplicity of images and concomitant definitions create a confusing picture. As a simplification, Sattler (1986) proposes that a learning disability can be visualized in both a broad and a narrow sense.

In the broad sense, a learning disability can be thought of as a learning difficulty that can be associated with any type of factor, including, but not limited to mental retardation, brain injury, sensory difficulties, or emotional disturbance. In the narrow sense, a learning disability can be thought of as the failure, on the part of a child who has adequate intelligence, maturational level, cultural background, and educational experiences, to learn a scholastic skill.

The narrower meaning is what is commonly referred to as specific learning disability, which is defined as follows in Public Law 94-142 (Federal Register, December 29, 1977, p. 65083, 121a.5):

> Specific learning disability means 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 to do mathematical calculations. The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include children who have learning problems which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage.

Public Law 94-142 also specifies that the phrase specific learning disability only applies to children who have a severe discrepancy between achievement and intellectual ability in one or more expressive or receptive skills such as written expression, listening and reading comprehension, or in math calculations or reasoning.

Another definition of learning disability comes from the National Joint Committee for Learning Disabilities (NJCLD). It is a definition that emphasizes that learning disabilities includes any disorder that may seriously handicap an individual in specific areas of functioning (Hammill, Leigh, McNutt, & Larsen, 1981, p.366).

"Learning disabilities is a generic term that refers to a heterogeneous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning, or mathematical abilities. These disorders are intrinsic to the individual and presumed to be due to central nervous system dysfunction. Even though a learning disability can occur concomitantly with other handicapping conditions (e.g. sensory impairment, mental retardation, social and emotional disturbance) or environmental influences (e.g. cultural differences, insufficient/inappropriate instruction, psychogenic factors), it is not the direct result of these conditions or influences."

The NJCLD definition emphasizes that the disorder is caused by factors associated with the individual and not the environment. Presumably the factors are related to central nervous system dysfunction and are not a direct result of other handicapping conditions or environmental deprivation.

Public Law 94-142 and its amended versions state that children who are not learning at a "normal" rate due to reasons other than lack of ability, lack of opportunity, social-emotional disturbance, or sensory defect are referred to as "learning disabled" (Hallahan & Bryan, 1981). This narrower meaning for learning disabilities has helped to define which learners are to be considered disabled enough that they qualify for special services in public schools.

There are, however, several details left unspecified even with this narrower definition of learning disabilities. Notably missing from this definition is some guidance about the degree of ability/achievement discrepancy required to make a student learning disabled (Gearheart, Weishahn, & Gearheart, 1988). The practical problem of how to identify a specific learning disability based on this narrow definition was only addressed in the accompanying federal regulations.

The federal regulations require a severe discrepancy between achievement and intellectual ability as the basis for learning disabilities eligibility. No criteria for defining a severe discrepancy were provided. Even though the discrepancy criteria is not clear, it is clear that a comparison of achievement levels and ability levels will produce data about the discrepancy.

A simple method of making this achievement/ability comparison is to collect test data. Ability levels have commonly been assessed with intelligence tests and achievement with tests designed to measure prior knowledge. There is, once again, no mandate on which ability or achievement tests to use when collecting this information that is needed to evaluate severe discrepancy. Choice of ability and achievement instruments has the potential to alter measured discrepancy.

Multiple problems have been noted with the definition and identification of learning disabilities (Gallagher, 1986; Ysseldyke & Algozzine, 1983; Chalfant & Schefflin, 1969; Clements, 1966; Strauss & Lehtinen, 1947). This research will focus only on the issue of the appropriateness of defining learning disability as a significant discrepancy between ability and achievement. This identification issue alone has the potential for altering who is considered learning disabled.

Significance of the Problem

The use of 'learning disabled' as a diagnostic category has been criticized by many (Chalfant, 1989). One major concern involves the continued increase in the numbers of students classified as learning disabled. The U.S. Department of Education's Tenth Annual Report to Congress (1988) indicated that prevalence has increased dramatically, more than doubling in the previous decade. Almost five percent of all children in the public schools were classified as learning disabled at the time of this report and more than 40% of all pupils served in special education classrooms were so classified.

Differences in classification rate from state to state suggest problems in the definition, assessment, and identification of learning disabled students. States report a range of 26% to 64% in the percentage of the students in special education that are learning disabled (U.S. Department of Education, 1988).

One of the most frequent criticisms about learning disabilities has been the lack of consensus about the definition. This definitional problem has contributed to the rise in identification rate and is related to the wide variation in identification rates among the states.

The greatest divergence of opinion within the field of learning disabilities relates to diagnosis. There is no consensus concerning the diagnostic procedures that should be used to specify the classification of a student as learning disabled. At the heart of this controversy is the criterion for classification. Even with a consensus about definition, even using the federal definition there remains the problem of deciding which of the parts of the definition will be used as criteria: task failure, ability/achievement discrepancy, etiological factors, exclusionary factors, or dysfunctions in one or more of the psychological processes. State guidelines vary about these classification criteria.

Among states that do agree on classification criteria, there is no consensus on choice of diagnostic instrument used to detect or identify the criteria. An array of diagnostic tests are used by practitioners (Perlmutter & Parvus, 1983). Diagnosticians tend to assess the criterion variable using instruments with which they have been trained and feel most comfortable. Some of the wide range of tests used have questionable validity and reliability, and the freedom allowed in test selection keeps the field from defining objective and consistent standards for determining eligibility.

Choice of diagnostic instrument(s) used to identify learning disabled students impacts eligibility. If, for instance, the criteria for eligibility is defined as discrepancy between actual achievement and a predicted level of achievement based on ability, discrepancies will vary based on instrument choice.

The choice of the measure of ability or intelligence has the most potential for creating variance in discrepancy outcomes. There are different models of intelligence and there are instruments designed to assess abilities based on each theory of intelligence. With this multiple-option choice of intelligence tests, different sets of abilities can and are measured.

The use of different intelligence theories and selection of intelligence tests based on these theories poses a significant problem for the process of determining learning disabilities eligibility. Children and their families move, state agencies change personnel and policies, local school districts change policy, tests are revised, new tests are marketed, currently eligible students must be re-evaluated every three years; all these things contribute to the possibility that a child is evaluated with different sets of parameters from time to time. Eligibility for special education services is not static.

Statement of the Problem

The purpose of this study was to determine differences in learning disabilities eligibility upon re-evaluation when a change in use of intelligence test was made. Comparisons were made of ability/achievement discrepancies of LD students who were eligible as the result of an evaluation that included the Wechsler Intelligence Test for Children - Revised (WISC-R) and the same groups' discrepancies upon third year re-evaluation when the Woodcock-Johnson Tests-Revised (WJR) Cognitive Abilities measure was used. Previous research acknowledges acceptable concurrent validity for the WISC-R and both the original Woodcock-Johnson test and the WJR. No previous research exists, though, on what impact actual standard score differences between the WISC-R and the WJR may have on eligibility with a LD sample.

To control for the introduction of the new instrument (WJR); ability/achievement discrepancies of a control group of LD students also was analyzed. These students were also assessed at the three year interval but they were administered the WISC-R at both evaluations.

Objectives

The problem was operationalized through comparisons of evaluation results at two points in time: For what are called the "experimental groups" this involved the most recent evaluation on-record for the learning disabilities student when the WJR was used (1991-1992) and the evaluation done three years previously when the WISC-R was used (1988-1989). For the "control groups" this involved the most recent evaluation when the WISC-R was used (1989-1990) and the evaluation done three years previously when the WISC-R also was used (1986-1987). The following research questions were asked:

1) What are the standard score differences in intelligence from one evaluation to the next for each child?

2) What are the trends in achievement scores over the three year interval?

3) What are the differences in ability/achievement discrepancy scores in reading, math, and written language at successive three year evaluations?

4) What are the effects of size of discrepancy required for learning disabilities eligibility?

5) How are standard score discrepancies between ability and achievement related to age?

Limitations

The following limitations will apply to this study:

1) This study is limited to using the commonly accepted criteria for identifying and determining eligibility for learning disabilities. That is, children with learning disabilities are identified when they are not learning at

a normal rate and they are made eligible if a significant discrepancy between ability and achievement exists.

2) This study is limited to focus on assessment results only when considering factors related to the diagnosis of a learning disability. Other factors such as cultural differences, insufficient/inappropriate instruction, psychogenic factors, and sociological factors such as family functioning will not be investigated on a group or individual basis.

3) This study's scope in terms of evaluating the effect of a model of intelligence and the corresponding method of assessment upon learning disability eligibility is limited to two models. Only the model reflected by the Wechsler instruments and the multiple intelligences model of the Woodcock-Johnson Assessment of Cognitive Ability is investigated.

Definition Of Terms

Intelligence. Two theories of intelligence are involved so it is necessary to define each theory. Operationally, Wechsler (1974, p. 9) defines intelligence as the ability to comprehend and use language and to perform tasks that involve perceptual organization. Woodcock & Johnson (1989, p. 26) define intelligence as the ability to process information through sensory detection, making associations, organizing visual and auditory perception and by educing relations.

Ability. Degree of intelligence(s) as displayed by performance on tests constructed for the purpose of measuring mental development level.

Achievement. Knowledge about or prior learning of facts and applications in a specific content area as measured by norm-referenced tests constructed to assess subject matter.

Potential. A predicted level of achievement corresponding to a given ability level.

Discrepancy. The standard score difference between a student's ability measure score and his/her score on measures of achievement in specific subject matter areas.

CHAPTER II

REVIEW OF THE LITERATURE

Theories of Intelligence(s)

The principal research question addressed by this study involves the impact of choice of intelligence test upon learning disabilities eligibility. So, it is important to detail the theory base for the two ability measures that contribute data to the study. First, however, an overview of definitions of intelligence and measurement of intelligence is necessary. Another goal of this review will be to describe the research and theory that have contributed to the promotion of the use of discrepancy between ability and achievement as the method for determining learning disabilities eligibility. Theory and research in the literature have established no explicit relationship between a particular test used to measure intelligence and LD eligibility.

Definitions of Intelligence

There is no consensus regarding the definition of intelligence, though a variety of definitions exist which have some commonalities. Most definitions of intelligence emphasize abilities to adjust to or adapt to the environment, the ability to learn, or the ability to use abstract thinking with symbols or concepts. Wesman (1968) says the confusion concerning defining and measuring intelligence is linked to the fact that intelligence is an attribute, not an entity, and that it reflects the summation of the learning experiences of the individual. As such, intelligence is known by what it enables us to do (Wechsler, 1958).

Binet, (Binet & Simon, 1916) one of the early theoreticians in intelligence, regarded intelligence as a collection of faculties: judgment, practical sense, initiative, and the ability to adapt oneself to circumstances. Other definitions, (Wechsler, 1958), are similar: "Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with the environment." (p.17).

In a recent survey (Snyderman & Rothman, 1987) of experts in psychology, education, sociology, and genetics a high degree of consensus existed about the important elements of intelligence. Three behavioral descriptions received near unanimous agreement of 96% or higher: Abstract thinking or reasoning, the capacity to acquire knowledge, and problem-solving ability. Seven behavioral descriptions were checked by a majority (60-80%) of respondents: Adaptation to one's environment, creativity, general knowledge, linguistic competence, mathematical competence, memory, and mental speed.

Factor analytic theorists have made significant contributions to the definition of intelligence. Spearman, (1927) an early advocate of a factor analytic approach to intelligence, proposed a two-factor theory of intelligence. Spearman thought intelligence was composed of a general (g) factor plus one or more specific (s) factors that accounted for performance on intelligence tests. The g factor was a general mental energy, with complicated, complex activities such as are found in reasoning, comprehension, and hypothesis-testing tasks containing the greatest amount of g. In contrast, tests with a low g loading are less complex and emphasize processes such as recognition, recall, speed, visual-motor abilities and motor abilities.

Multifactor theories also have been proposed. Thorndike (1927) conceived intelligence to be the product of a large number of interconnected but distinct abilities that combine to form clusters. He identified three of these clusters: Social (ability to deal with people), concrete (ability to deal with things), and abstract (ability to deal with verbal and mathematical symbols). Even though Thorndike's ideas about intelligence are called multifactorial, they were not developed with factor analytic methods. The ideas came from personal perspective.

Thurstone (1938) believed that intelligence could be divided into multiple factors with equal weights, thus eliminating the idea of the g factor. Using factor analytic methods, Thurstone identified the following factors he called the primary mental abilities: verbal, perceptual speed, inductive reasoning, number, rote memory, deductive reasoning, word fluency, and visualization.

Another multifactor theorist, (Guilford, 1967) developed a three-dimensional Structure of Intellect model to depict intellectual factors in a system. One dimension represents the operations involved in processing information, a second dimension represents contents, and a third dimension represents products. With this model, intelligence is seen in terms of the kind of mental operation performed, the type of content on which the operation is performed, and the resulting product. With five types of operations, four types of content, and six types of products, a possible 120 factors exist.

At the same time that the multifactor theorists continued to expand their ideas, other theorists remained aligned with the idea of g or a general ability. One such theorist (Vernon, 1950) devised a hierarchial theory of intelligence. At the top of his hierarchy of intelligence was g, the general ability factor. At the next hierarchial level there were two factors: verbal-educational and spatial-mechanical. Under each of these two factors there were subdivisions of more specific factors related to the two fields. Creative abilities, verbal fluency, and numerical factors belong to the verbal-educational factor.

Spatial, psychomotor, and mechanical information factors belong in the spatial-mechanical factor. The substantial positive intercorrelations among cognitive subtests across a representative population support Vernon's (1965) belief that a general group factor (g) must be considered in any attempt to understand intelligence.

Two theorists (Cattell, 1963; Horn, 1967, 1968, 1978, 1985a; Horn & Cattell, 1967) have developed another innovation on the structure of intelligence. Their theory postulates that there are two types of intelligence - fluid and crystallized. Fluid intelligence is essentially nonverbal, relatively culture-free mental efficiency. Crystallized intelligence involves acquired skills and knowledge that are strongly dependent on exposure to culture. Figure classification, figural analysis, number and letter series, matrices, and paired associates are examples of tasks that use fluid intelligence. Crystallized intelligence is measured with tests of vocabulary, general information, abstract word analogies, and mechanics of language. The Wechsler scales contain measures of both types of intelligence.

Horn (1985b) argues that research does not support the concept of general intelligence. Instead, he believes in a four-level hierarchial model. At the lowest level are visual and auditory sensory detection functions. The second level involves associational processes, both short and long term. At the third level, perceptual organizational processes such as visualization, clerical speed, and auditory thinking are used. The highest level involves the eduction of relations which uses both fluid ability and crystallized ability.

More recently (Sternberg, 1986; Das, 1972; Jensen, 1970; Gardner, 1983) other theorists have contributed to the base of constructions used to view intelligence. Neither Sternberg's dimensions of intelligence, nor Das's simultaneous and successive processing ideas, nor Jensen's associative and cognitive abilities, nor Gardner's multiple intelligences are, however, as relevant to the measures of intelligence that are the focus of this study as the theorists already discussed. Previous theory about intelligence has been included because it was a basis for the structure that was used to devise the Wechsler Intelligence Scale for Children and the Woodcock-Johnson Tests of Cognitive Abilities.

Wechsler Intelligence Scale for Children - Revised

David Wechsler's first scale for measuring children's intelligence (Wechsler, 1949) was developed as a downward extension of his adult intelligence test, the Wechsler-Bellevue Intelligence Scale. A revised version of the children's scale (WISC-R) was published in 1974.

Subsequently, a third edition (WISC-III) of this measure was published in 1991. Results from the WISC-R areused in this study. The WISC-R covers an age range from 6-0 to 16-11 years and contains 12 subtests.

Wechsler states in the manual for the WISC-R (Wechsler, 1974) that the scale is not predicated on any particular definition of intelligence: "Intelligence is the overall capacity of an individual to understand and cope with the world around him." (p. 5). In addition: " Intelligence is an overall or global entity; that is, a multidetermined and multifaceted entity rather than an independent uniquely-defined trait. This avoids singling out any ability, however esteemed, as crucial or overwhelmingly important. Intelligence is best regarded not as a single unique trait but as a composite or global entity" (p. 6).

The previous definition sounds similar to Spearman's general ability (g). Wechsler does not believe this to be the case, though. He states in the manual that he is avoiding equating general intelligence with intellectual ability. Rather,

> "Intelligence is not a kind of ability at all, certainly not in the same sense that reasoning, memory, verbal fluency, etc., are so regarded. Intelligence is something that is inferred from the way these abilities are manifested under different conditions and circumstances. One can infer an individual's intelligence from how he

thinks, talks, moves, almost from any of the many ways he reacts to stimuli of one kind or another," (p. 17).

Mental tests, then, are newer inventions for an observational process that has historically been a method of appraising intelligence.

The WISC-R contains a subdivision of the scale into Verbal and Performance tests. Wechsler thought this dichotomy represented a way of identifying two principal modes by which human abilities express themselves. Also, each of the component tests is equally weighted to obtain the child's IQ. Wechsler states: "This procedure is based on the theory that intelligence measures are best regarded as assortative, not hierarchial" (p. 9).

Factor analysis of the WISC-R standardization group has indicated that three factors could efficiently describe the test (Kaufman, 1975). These factors are labeled Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility. The factor structure of the WISC-R closely agrees with the actual organization of the subtests. Verbal Comprehension describes the hypothesized ability underlying the the factor with reference to item content (verbal) and the mental process (comprehension). Perceptual Organization describes the hypothesized ability underlying the factor with reference to item content (perceptual) and the mental process of organization. Freedom from Distractibility apparently names the ability to concentrate or remain attentive.

Organizationally, the WISC-R is a measure that produces three scores; a Verbal IQ, a Performance IQ, and a Full-Scale IQ. Even though Wechsler views intelligence as a global entity, he asserts that intelligence can be inferred from measuring human verbal abilities and performance on tasks. Scores on this test reflect the multifaceted nature of intelligence that can, however, be reflected in a composite score that is not hierarchially structured or weighted.

Woodcock-Johnson Tests of Cognitive Ability - Revised

The Woodcock-Johnson Tests of Cognitive Ability are the cognitive portion of the Woodcock-Johnson Psycho-Educational Battery first published in 1977 (Woodcock & Johnson, 1977). A revised version of this battery (Woodcock & Johnson, 1989) is the instrument used in this study. The broad battery measures scholastic aptitudes and achievement as well as cognitive abilities.

The revised version of the Woodcock-Johnson Tests of Cognitive Ability (WJ-R COG) is subdivided into a Standard Battery and a Supplemental Battery. The Standard Battery, used for reporting cognitive ability results in this study, is composed of seven subtests. The tests were standardized on subjects aged 24 months to 95 years of age.

The WJ-R COG is described by the authors as an operational representation of a specific theory of intellectual processing. It is a theory that stems from several major theories of intellect (Cattell, 1963; Horn, 1972, 1976, 1985b, 1986, 1988; Horn & Cattell 1966). The term "cognitive ability" is used synonomously with the term "intelligence" by this theory.

One major component of the Horn-Cattell theory is that intelligence should be viewed as a hierarchy. At the top of the hierarchy of functions there is fluid and crystallized intelligence. The eduction-of-relations capacities at this level of the hierarchy are dependent on perceptual organizations (visual and auditory), which, in turn, depend on associational processing, which is based on sensory detection. A developmental hierarchy proceeding from infancy to old age parallels this hierarchy of functions (Woodcock-Johnson, 1989). In other words, sensory detection is an early developmental achievement that is built upon by the infant and later the child until more sophisticated cognitive functioning such as eductions-of-relations are reached in the developmental process.

Eduction-of-relations in this model is demonstrated by the abilities represented in two factors; fluid reasoning and comprehension-knowledge. Fluid reasoning (Woodcock-Johnson, 1989) is said to be the factor in this model most similar to Spearman's g. It is a broad ability to reason and is manifested in drawing inferences and implications and is not heavily dependent on previously acquired knowledge or earlier-learned problem-solving.

For Woodcock-Johnson, the comprehension-knowledge factor represents crystallized intelligence, a person's depth and breadth of knowledge of a culture. This broad intellectual ability is an indicator of individual differences in knowledge, especially in the use of knowledge via verbal abilities and reasoning based on previously learned procedures.

At the second highest level of the Horn-Cattell theory of intellectual processing, perceptual organization is the principal function. Perceptual organization is represented by three abilities: visual processing, auditory processing and processing speed. Broad visualization requires fluent thinking with stimuli that are visual in the mind's eye, that is, the stimuli need not be presented visually, although they usually are. Ability is demonstrated on tasks including recognition of rotations and reversals of figures, finding hidden figures, identifying incomplete or distorted figures, and comprehending spatial configurations.

Auditory processing is the ability to fluently comprehend patterns among auditory stimuli. Tasks requiring this ability include temporal tracking, the perception of speech under distorting or distracting conditions, the detection of transformation of tonal patterns, and the anticipation of an auditory form that can be synthesized from a stream of sounds.

Processing speed, sometimes also called clerical speed, is the ability to work quickly when measured under time pressure to maintain focused attention. Speed of scanning, comparison, writing and printing are examples of tasks that demonstrate this ability.

At the third level of the Horn-Cattell hierarchy of intellect, associational processing is the operative. Short-term memory and long-term retrieval represent this function. Short-term memory involves storing information in the immediate situation and retrieving it within a few seconds at most. Long-term retreival is the ability to retrieve information stored earlier.

At the lowest level of this model of intellectual processing, sensory reception is the function. Visual and auditory sensory detection involve sensitivity to and awareness of visual and auditory stimuli.

To summarize, the WJ-R COG measures seven broad intellectual abilities: fluid reasoning,comprehension-knowledge, visual processing, auditory processing, processing speed, long-term retreival, and short-term memory. The Standard Battery of the test includes one subtest for each of the seven abilities. These seven subtests combine to produce a cluster score, Broad Cognitive Ability, which is a primary basis for test interpretation.

In contrast to the WISC-R, the WJ-R COG conceptualizes intelligence as a hierarchial structure. Intelligence is not a global entity inferred from verbal abilities and performance on tasks but is measured by abilities in a hierarchy in which the distance from top to bottom of the hierarchy is inversely proportional to the magnitude of correlation between abilities. Scores from the top of the hierarchy, however, are not more heavily weighted in the composition of the Broad Cognitive Ability score than are the abilities at the bottom of the hierarchy.

Measuring Intelligence

Since the time of Binet, assessment of mental ability has been feasible. With Binet's scale, higher mental processes were the focus instead of simple sensory functions. Earlier attempts to measure intelligence had focused on psychological processes such as sensation, attention, perception, association, and memory. Intelligence (Binet & Simon, 1916) was reformulated as a shifting complex of inter-related functions that could be practically measured with concern for age-based cognitive development.

The Binet-Simon scales were easily accepted and revised in the United States (Goddard, 1910). Conceptually however, Goddard viewed intelligence differently than Binet. Instead, intelligence was seen as a single, underlying function or faculty by Goddard (Tuddenham, 1962).

Other American investigators (Terman, 1916) were interested in the idea of intelligence as a single property that could be assigned a score. Terman adopted the concept of a mental quotient from Stern (1914). Stern recommended the mental quotient because it was a value that expressed the ratio of mental to chronological age. Terman used the idea and renamed the ratio the intelligence quotient (IQ) for the 1916 revision of the Binet-Simon scales.

Since Binet's test had an age-scale format, the American revisions adopted this approach. The age-scale format (Sattler, 1986) is standardized on groups of children at various age levels. Those items passed by a majority of the standardization sample children at a particular age are assigned to that level. A basal age is established at a level where a child can pass all items of the level. Additional age credit is given for subtests passed beyond the basal level. Age scale subtests are selected on the assumption that important forms of cognitive development appear at specific points in development. The Stanford-Binet scale contains a collection of different tests for different age groups.

There were opponents to the age-scale format (Yerkes, 1917). Yerkes argued that a point-scale was a better alternative. A point-scale format assigns points on the basis of correctness and quality of the response. Raw score totals are converted to standard scores which are then converted into an overall score. Subtests for the point-scale format are selected to measure specific functions. That is, they measure the same aspects of behavior at every age.

Wechsler also was interested in developing a point scale. He selected subtests initially from previous sources; Army Alpha and Beta, Stanford-Binet, Healy Picture Completion Test, Kohs Block Design, and Army Group Examinations. The Wechsler Scales were designed to take into consideration the factors contributing to the total effective intelligence of the individual; no attempt was made to design a series of subtests to measure "primary abilities" or to order the subtests into a hierachy of relative importance.

Beginning in the late 1950's, with the computer as analog, the information processing model began to develop. Swanson (1985) states that the information processing perspective assumes that a number of component operations or processing stages occur between a stimulus and a response. All behavior of a human information processing system is the result of combinations of these various processing stages. Human cognition, then is seen as occurring in a series of discrete stages, with information received being operated on at one stage and then passed on as input to the next stage for further processing.

Woodcock and Johnson have incorporated information processing conceptions of intelligence in their subtest selection. The heirarchy of intellect structure that is the theory base for their test of cognitive ability categorizes mental processes in terms of the different operations performed on the information: sensory reception, associational processing, perceptual organization, and eduction of relations. Subtests have been designed to measure abilities at each of these discrete processing stages.

Validity Issues

As one can see from this limited history of the development of tests of mental ability, different models for the structure of intelligence have guided the construction of particular tests. It is important, then, to examine the validity of these instruments or the extent to which they measure what they are supposed to measure.

Validity is important to determine the appropriateness with which inferences can be made on the basis of the test results. The potential and actual social consequences of using intelligence tests are significant (Messick, 1980). Intelligence tests play a significant role in the diagnosis of learning disabilities when the discrepancy between ability and achievement is the principle assessment concern for establishing the presence of the disability.

Segregation is the actual social consequence of placement in public school classes for learning disabilities. This means less interaction with non-handicapped peers. The potential consequences, in terms of the effects of labeling, self-esteem maintenance, the loss of models of academic success, may be greater.

Intelligence testing has generated much controversy in recent years as evidenced by the Larry P. case and the Parents in Action on Special Education v. Joseph Hannon case (Sattler, 1986). Opponents argue that intelligence testing restricts children's opportunities, places minorities in an unfavorable position, and sorts children into stereotyped categories. Proponents maintain that intelligence testing facilitates movement between social classes, reveals unsuspected talents in many individuals, and assists in the diagnostic process (Sax, 1989).

The most important point to recognize, for this study, is that global scores obtained from different tests of cognitive ability may not be interchangeable. Previous research has established these differences (McGrew, 1986). However, no previous research has shown the effects of these differences on the continued eligibility of an LD sample.

With the WISC-R and the WJ-R COG having different theoretical constructs; construct validity is probably influenced by the different factor structures (Estabrook, 1984). Therefore, concurrent and content validity become especially important considerations because these types of validity establish to some degree whether the two tests are measuring the same things.

Concurrent validity is a measure of whether test scores are related to some currently available criterion measure. Woodcock & Johnson (1989) report a correlation of .685 between the WISC-R Full Scale score and the Standard Battery of the WJ-R COG. No similar measure is reported in the WISC-R manual since it was normed several years before the WJ-R. The literature, as yet, contains no subsequent research since the 1989 publication date of the WJ-R manual about the concurrent validity of the WJ-R COG and other criterion measures.

Considerable research is available, though, on correlations between the WISC-R Full Scale and the original WJ-COG. Strong concurrent validity existed (median correlation of .77 across 21 comparisons; McGrew, 1986). Despite these reasonably high correlation coefficients, the WJ-COG has been the focus of much controversy (Cummings & Moscato, 1984; Thompson & Brassard, 1984; Woodcock, 1984). The primary reason for this controversy was the finding that the WJ-COG provided lower scores than the WISC-R. In two reviews of 21 WISC-R/WJ-COG research comparisons, McGrew (1986) and Woodcock (1984) both noted median mean score discrepancies of five to six scaled score points across all types of samples, in the direction of higher WISC-R scores.

Reports of larger discrepancies in academically handicapped groups (LD in particular) have been a major concern (Bracken, Prasse, & Breen, 1984; Coleman & Harmer, 1985; Hall, Reeve, & Zakreski, 1979; Shinn, Algozzine, Marston, & Yssaldyke, 1982). The range of mean difference scores was from 13 points (Reeve, et. al., 1979) to 5.65 (Coleman & Harmer, 1985). Others (Woodcock, 1984) have reported little mean difference.

A number of content difference hypotheses have been advanced to explain the WISC-R/WJ-COG mean score differences (McGrew, 1987). One suggested difference is in the proportion of general intellectual ability g present in each instrument based on inspection of subtest g characteristics (McGrew, 1984). A more popular content difference hypothesis has been the suggestion that the WJ-COG is saturated more heavily with verbal abilities than the WISC-R (Phelps, Rosso, & Falasco, 1984, 1985).

Ysseldyke, Shinn, and Epps (1980, 1981) and Shinn et al. (1982) have advanced an achievement content hypothesis which suggests that lower WJ-COG scores in academically handicapped samples are due to the fact that the WJ-COG is loaded inappropriately with achievement content. If the WJ-COG Broad Cognitive Ability Score might actually measure achievement, rather than ability, LD students are at a disadvantage and there is confusion in the process of making placement decisions.

Predictive validity is the last validity issue to be raised in this review. Predictive validity refers to the correlation between test scores and performance on a relevant criterion. In the case of LD students, the important question concerns whether the ability test used as part of the diagnostic process accurately predicts potential achievement levels. McGrew and Pehl (1988) noted that the WJ-COG had higher correlations than the WISC-R when compared to reading and math achievementand they conjecture also that this finding is related to the WJ-COG's achievement content. Overall predictive power was generally similar, though, for these two ability measures when a "normal" sample was used.

To summarize the validity issue, the following points are relevant. First, the WISC-R and the WJ-COG have reasonably strong correlations between Full Scale and Broad Cognitive scores across samples. These correlations establish reasonable concurrent validity and possibly allow a legitimate substitution of one measure's score for the other.

The actual score differences that research has generated bring concurrent validity into question, though. Even the smallest score differences reported (five or six standard score points) have the capacity to exclude a significant number of children from LD eligibility if the intelligence test yielding the lower scores is used. These score differences from previous research on the WISC-R and the first edition of the WJ-COG make it important to establish evidence about this comparison using the WJR-COG.

Issues of predictive validity also are important when considering whether these two intelligence tests can be reliably substituted. Predictive validity is the heart of the concept of learning disabilities when the definition of LD involves a comparison of ability and achievement. The measurement question is:

Does a particular intelligence level correlate strongly with subsequent achievement levels? Limited research indicates little difference between the WISC-R and the WJ-COG in this area. Once again, it is important to establish this comparison with the revised Woodcock-Johnson.

Content validity will not be a focus with this study. There is no intention, though, to signify that content validity is not an important consideration when evaluating test score differences on the WISC-R and the WJR-COG.

Issues of validity must be put in perspective. The validity of using the WISC-R or the WJR-COG as a measure of intelligence is an important consideration for this study. This issue is only a component, however, of the issue of the definition of a learning disability. The last portion of this review will turn to the more specific issue of defining a learning disability as a severe discrepancy between ability and achievement.

Ability/Achievement Discrepancy

Bateman (1964) was one of the first to include the idea of discrepancy in a definition of learning disability. The idea has become widely accepted since that time and was included in the discrepancy clause in PL 94-142 regulations and subsequent revisions.

When defining learning disabled students as underachievers, it becomes necessary to quantify the discrepancy between mental ability and achievement.

Financial constraints and a commitment to keep mildly handicapped students in the mainstream have promulgated this need to quantify severe discrepancy (Cone & Wilson, 1981).

A variety of methods have been proposed, have been used, and are still being used to quantify discrepancy. The four most commonly used methods have been; grade level deviation, expectancy formulas, standard-score comparisons, and regression analysis. No method possesses attributes that appeal to all individuals involved in this area of education policy. The four methods that will be described represent an evolution in the sophistication of determining severe discrepancy.

The grade-level deviation method is probably the simplest of the four methods and does not actually include the use of an IQ score in the determination process. The deviation from grade level discrepancy has, however, been frequently utilized to distinguish underachievers. The process involves comparing the student's achievement scores to current grade placement. Achievement scores that are "significantly" below grade placement indicate severe discrepancy. Definitions with a constant deviation criteria specify a particular minimum value the discrepancy must have in order to be considered severe, for example, achievement levels two years below grade placement. A more sophisticated approach uses graduated deviation schedules with an increase in the required magnitude of deviation as grade placement increases.

Expectancy formulas for quantifying discrepant achievement have taken many forms (Kaluger & Kolson, 1969; Bond & Tinker, 1957; Johnson & Myklebust, 1967; Harris, 1970). Harris' formula is as follows and is offered as an illustration even though other formulas differ considerably: Expected Grade Equivalent = [(2MA + CA)/3] - 5. Using this example, the expected grade equivalent for a 6-year-old child with a MA of 6 would be first grade.

A problem with expectancy formulas that use both an MA and some quantified measure of achievement in the formula is that they assume that the correlation between scores on the ability test (where the MA was obtained) and scores on the achievement test (which are predicted) is 1.0, which is rarely the case. Another detraction of this method is the limitation of the MA concept. Also, both of the previous techniques, grade-level deviation and expectancy formulas, have incorporated grade equivalent or age equivalent scores. These derived scores lack equal variability characteristics, are valid only for skills that show a relatively linear growth pattern, and have greatly different meanings depending upon the student's actual age or grade placement.

To avoid the above problems, standard score comparisons have been used to establish discrepancy. A common procedure has been to obtain a standard-score value on a standardized test of mental ability and a comparable (a distribution with the same mean and standard deviation) standard-score value on a standardized test of academic achievement. If the difference between the obtained scores is greater than an established criterion (usually one to two standard deviations), the student is typically considered to be discrepant or underachieving. The standard-score comparison procedure takes into account mental ability and the increased range and variability of achievement scores at the upper grade levels as well as errors of measurement. Regression toward the mean is not addressed, however.

Regression procedures adjust for the well documented phenomenon of regression toward the mean. Regression effects occur when the correlation between two measures is less than perfect. The correlation between ability tests and achievement tests are always less than 1.0, therefore regression effects are always present in the standard-score comparison approach to determining severe discrepancy.

Specific application of the regression procedure was discussed by Cone and Wilson (1980, 1981) for LD identification. Students with IQ scores above average tend to have achievement scores that are not equally above average; achievement regresses toward the mean. When regression effects are not considered for above average intelligence, achievement standards are too high and, therefore, identify too many students as discrepant achievers.

Regression toward the mean also occurs for children with below average intelligence scores. Failure to consider regression effects in these cases results in expectations for achievement which are too low and result in the consideration of too few of this population of students as LD eligible.

The U.S. Department of Education (USDE) Special Education Programs (SEP) Work Group on Critical Measurement Issues In Learning Disabilities and the SEP National Task Force on Eligibility Criteria for Learning Disabilities recommend that regression procedures be used to arrive at a significant discrepancy. Use of the regression procedure requires knowledge of the correlation between the two tests being used. A correlation of .60 is often quoted as a typical correlation between ability and achievement tests. When the correlation is known, a table of expected achievement scores is generated for each ability score. This expected achievement score based on ability is then compared to the actual achievement score obtained to establish level of discrepancy.

Discrepancy formulas are not without drawbacks. One problem is the assumption that the tests used to evaluate a child measure independent functions, when actually achievement and ability tests to some extent measure the same factors. Additionally, the same processing difficulties that reduce achievement scores may reduce intelligence test scores. Also, (Reynolds, 1985) determining a severe discrepancy does not constitute the diagnosis of a learning disability; it only establishes that the primary symptom of LD exists. To many experts, the severe discrepancy is a necessary but insufficient condition for a diagnosis of LD. Reasons for failure to achieve, psychological process disorders, the exclusionary criteria, and medical and developmental histories demand serious consideration also (Chalfant, 1984).

Discrepancy Over Time

Discrepancy between ability and achievement is a complex topic. The addition of a time element to the discrepancy issue increases the complexity not in an additive fashion but in a multiplicative manner. The time lapse that is of interest to this study is the mandated three year interval between evaluations for learning disability eligibility.

As noted earlier, many variables contribute to discontinuity. Families move their children, schools change testing policies, publishers market new tests, etc. As a result of the above factors in addition to problems with defining the nature of the disability, professionals in the LD field express concern that too many of the students being served in classrooms for the learning disabled do not belong there (Perlmutter & Parus, 1983; Poplin, 1981).

Additionally, intraindividual differences due to changing patterns of ability and achievement contribute to discontinuity (White & Wigle, 1985) in magnitude of discrepancy. Movement to different cognitive and social developmental levels over the three year lapse can affect student variables such as motivation and self-esteem that are conjectured to contribute to achievement. With a situation that could, conceiveably, include such a large number of variables that are vaguely causative forces, it is probably best to initially focus on the statistical and psychometric differences found in the test scores from the two points in time.

Summary

This literature review was constructed to highlight the principal research question of: What are the changes in learning disabilities eligibility when a different measure of intelligence is used for two consecutive evaluation points? So, it was necessary to explore some of the relevant history of the following topics: definitions of intelligence, measuring intelligence, the theory and structure of the two intelligence measures utilized in the study, and validity issues relevant to the two instruments.

Also, the method of determining LD eligibility using the severe discrepancy concept were described.

CHAPTER III

METHOD AND PROCEDURES

Introduction

The purpose of this chapter is to describe the methods and procedures of this study. To that end, the subjects involved in the study are described and the method of selection is portrayed. Also, the procedures used to obtain the data are explicated along with the manipulations of the data. Additionally, the analyses of the data related to research questions and hypotheses are described.

Subjects

Subjects for this study were LD students in a small Oklahoma community. The LD population for the subjects' school district was 318 students from a total school population of 5686 at the most recent yearly child count. This proportion of LD students has remained almost constant for at least the past five years in this school district.

For a data collection period of two years (since the WJR was being used), approximately two-thirds of the total current male LD population was available as subjects for the groups receiving the repeated measures of the WISC-R and WJR in that order. Only randomly selected LD males were used as subjects since males have been shown to outnumber females in LD classification at rates as high as 5:1 (Finucci & Childs, 1981). Thirty subjects were selected for each of two groups categorized by school level, elementary and secondary, using a list of district LD students who had been re-evaluated since the WJR was in use.

For this nonequivalent control group design, the control groups were comprised of LD students who had received the repeated measures of two successive WISC-R administrations before the WJR was used. These subjects were randomly selected from archival data in district special education files. Once again, thirty subjects were randomly selected for each of the grouping variables of elementary and secondary students.

The elementary WISC-R x WJR group at the re-evaluation point was comprised of the following subjects; third graders - 3, fourth graders - 3, fifth graders - 10, sixth graders - 14. Mean age for this group was 12 years - 1 month. Age range was 9-11 to 13-6 years.

The elementary WISC-R x WISC-R group at the re-evaluation point was comprised of the following subjects; third graders - 1, fourth graders - 6, fifth graders - 8, sixth graders - 15. Mean age for this group was 11 years - 10 months. Age range was 9-4 to 13-9 years.

The secondary WISC-R x WJR group at the re-evaluation point was comprised of the following subjects; eighth graders - 9, ninth graders - 8, tenth graders - 2, eleventh graders - 5, twelvth graders - 6. Mean age for this group was 16 years - 4 months. Age range was 13-5 to 19-3 years.

The secondary WISC-R x WISC-R group at the re-evaluation point was comprised of the following subjects; eighth graders - 4, ninth graders - 11, tenth graders - 8, eleventh graders - 5, twelvth graders - 2. Mean age for this group was 15 years - 10 months. Age range was 13-11 years to 16-11 years.

Procedure

Coding of the archival data was accomplished by examining the folders of school-identified LD students from district files. The procedure was minimally intrusive since the records that were sought were of tests that had already been administered and recorded as part of special education policy. Personal identification was avoided since psychometric data (test scores) and subject characteristics of grade and age were the only data coded.

Coding was accomplished with the use of a grid to tabulate ability and achievement test scores and corresponding subject characteristics of grade and age. Size of ability/achievement discrepancy for each of the three content areas for each of the evaluations was also calculated and recorded on the grid.

Subjects were included in the study if examination of their special education file disclosed specific psychometric data explained in the following section on instrumentation. Data for the study was collected from special education files in the Spring of 1993.

Instrumentation

LD Students who were evaluated three years previously using the WISC-R as an ability measure and subsequently re-evaluated with the WJR-COG were included in the experimental group unless two poorly correlated achievement tests were used at the two points in time. A poor correlation was said to exist if the two achievement measures did not cover the areas of math, reading, and written language and the two tests did not have cluster scores in each of these academic areas. A cluster score had to be derived from a minimum of two subtests in each of the three content areas.

Subjects included in the nonequivalent control groups were LD students who were administered the WISC-R for both of two successive evaluations. The same stipulations for correlated achievement tests described for the experimental groups also applied for the control groups.

The testing for the subjects in this study was done by Oklahoma State Department of Education psychometrists and school psychologists based at one of the Regional Education Service Centers (RESC). Referral for evaluation forms used by the Oklahoma State Department of Education contained a statement that information from the evaluation could be used for data collection purposes in accordance with federal laws.

Prior to the summer of 1991, RESC personnel were almost exclusively using the WISC-R as an ability measure. The WISC-R was normed on 2200 children representative of the United States population of children stratified as to age, gender, race, geographic region, occupation of head of household, and rural-urban residence from the 1970 census.

Reliability of the WISC-R was evaluated for internal consistency using split-half techniques. Average coefficients across eleven age groups ranged from .70 to .86 for subtests. Full Scale, Verbal, and Performance Scales derived average coefficients across age groups of .90 to .96. Stability was evaluated with a re-test three months later for three age groups. For subtests, average coefficients were in the .65 to .88 range. Verbal, Performance, and Full Scale scores with re-test were in the .90 to .95 range.

Concurrent validity of the WISC-R was evaluated by looking at the correlations with three other individually administered intelligence tests; the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), the Wechsler Adult Intelligence Scale (WAIS), and the Stanford-Binet Intelligence Scale (Form L-M). Full Scale comparisons were as follows; WPPSI = .82, WAIS = .95, SBL-M = .73.

For an achievement assessment prior to 1991, the RESC was using the Woodcock Language Proficiency Battery (WLPB) to assess reading and written language achievement and they were using the Woodcock-Johnson Psychoeducational Battery's (WJPEB) math subtests to measure math achievement. Content of the WLPB is similar to the reading and written language portions of the WJ-R; in fact many of the same items are used on both tests. Cluster scores also are composed in a similar manner for both tests. The math subtests of the WJPEB and the WJ-R also are similar. Cluster scores are structured similarly and many of the same items are used on both math tests.

Beginning with the 1991-1992 school term, the local RESC began using the Woodcock-Johnson Psychoeducational Battery - Revised (WJR) as both an ability and achievement measure. The WJ-R was normed on 6,359 subjects using a stratified sampling design that controlled for the following variables; census region, community size, sex, race, funding of college/university attended, type of college/university attended, education of adults, occupational status of adults, occupation of adults. Over one hundred geographically diverse communities were sampled with selection based on the following SES variables; years of adult's education, household income, labor force characteristics, and occupation of employed adults.

Internal consistency was evaluated using split-half techniques for subtests and for clusters. Nine age ranges were included in the analysis which produced subtest correlations for the seven cognitive subtests (standard battery) within the .69 to .91 range. Cluster score correlation for the Broad Cognitive Ability Scale was .93. No test-retest reliability is reported.

Concurrent validity of the WJ-R Broad Cognitive Ability was evaluated with correlations from two school-age groups. For the Age 9 group, correlations were as follows; K-ABC Mental Processing Composite = .57, SB IV Composite = .69, WISC-R Full Scale = .69. For the Age 17 group, correlations were; SB IV Composite = .65, WAIS-R Full Scale = .64.

Only subjects evaluated with the mixture of WISC-R and WLPB and WJPEB for the evaluation three years ago and the WJ-R for both cognitive and achievement testing at the re-evaluation were included in the experimental groups. Only subjects evaluated with the WISC-R twice successively and with the same achievement mix described above were included in the control groups. All instruments utilized by this study have common means of 100 and similar standard-score deviations of 15.

Measurement Indices

Principal grouping involved a distinction between elementary (grades 1-6 inclusive) and secondary (7-12) students along with designations of "experimental" (WISC-R x WJR) and "control" (WISC-R x WISC-R).

For the ability measure, the Full Scale IQ score from the WISC-R was used and the Broad Cognitive Ability score was used from the WJR. Pairs of ability scores were used for each subject. One score represented the IQ at the first evaluation and the second score was the IQ upon re-evaluation.

For the achievement measure, standard scores from math, reading, and written language clusters were used. Each subject received six achievement scores. Three scores represented the achievement levels at the first evaluation and three scores represented the achievement levels at the re-evaluation point.

Each subject was also given six difference scores. Each of these derived scores represented an ability/achievement discrepancy. Three of the difference scores were calculated by comparing the first IQ score to the reading, math, and written language scores from the first evaluation. The second set of three difference scores were derived from ability and achievement standard scores at the second evaluation.

Also, frequency data was tabulated for learning disabilities eligibility in reading, math, and written language. Students were determined to be eligible in a specific content area if their difference score was 15 or larger. This meant that achievement in reading, math, or written language had to be at least 15 standard score points lower than the IQ score after it was adjusted for regression. Frequencies of eligibility were tabulated for the intervals of 15, 18, and 21 difference score points. This data was aggregated separately for each of the four treatment groups. This was done to reflect policy by the Oklahoma State Department of Education that recommends choosing a discrepancy level from the 15 to 21 point range.

Hypothesis Statement

The following null hypotheses reflects the principal research questions addressed by this study:

Hypothesis :

There are no statistically significant differences in frequencies of subjects eligible for learning disabilities placement between experimental and control subjects or elementary and secondary subjects at each of the two evaluation points.

Analyses

Data from this study was analyzed using SYSTAT: The System for Statistics (Wilkinson, 1984). Significant difference was evaluated with an alpha of .01 for all analyses.

The hypothesis of this study was analyzed with a series of Chi-Squared analyses. The analyses were done to determine if significant differences in frequencies of subjects eligible for LD placement existed at the three criterion levels (-15, -18, -21). Two sets of the

Chi-Squares were done; one set established differences between experimental and control groups subjects' eligibility frequencies. The second set of Chi-Squares were calculated to analyze the distribution of LD eligibility frequencies of elementary and secondary subjects.

CHAPTER IV

RESULTS

Introduction

Chapter four presents the results of the analyses related to the research questions and hypothesis of this study. The chapter is divided into three parts. The first part presents the descriptive statistics for the control groups and the experimental groups in ability, achievement, difference scores (discrepancy), and eligibility. The second part of the chapter presents the outcomes of the analyses done to address the proposed null hypothesis. Finally, a summary of the results is given.

Descriptive Statistics

Ability

All groups, experimental and control, had lower mean ability scores at the re-evaluation. Table I (refer next page) shows the ability results.

TABLE I

MEAN ABILITY SCORES

		Evaluation		Re-Evaluation	
	n	Mean	SD	Mean	SD
All Subjects;	120	95.1	11.0	91.4	10.8
Experimental	60	92.2	8.9	88.6	9.3
Control	60	97.9	12.2	94.3	11.5
Elementary	60	97.1	11.2	92.2	11.1
Secondary	60	93.0	10.5	90.7	10.5
Experimental:					
Elementary	30	93.1	9.5	88.2	10.7
Secondary	30	91.2	8.4	88.9	7.9
Control:					
Elementary	30	101.2	11.5	96.2	10.2
Secondary	30	94.8	12.2	92.4	12.5

The mean ability scores were lower upon re-evaluation for both the elementary and secondary control groups. The elementary group's decline in mean IQ was double that of the secondary group. However, the elementary group's initial IQ was a mean 6.4 standard points higher than the secondary group's initial mean IQ. A Plot of this data was ordinal; no interaction was found between level of education and repeated IQ measures.

The mean ability levels of the experimental groups also were lower upon re-evaluation for both the education levels. Decreases from one evaluation to the next were almost identical to the decreases observed in the control groups.

The major difference between the experimental and control groups was the relative level of ability. Control group ability levels were higher at all evaluation points than experimental ability levels.

The correlation between first and second ability tests was higher for control subjects than for experimental subjects. The highest correlation was found for the elementary control group and the smallest correlation was found for the secondary experimental group. Overall, correlation of successive ability measures were higher for elementary groups than secondary. Table II displays the ability correlations.

TABLE II

ABILITY CORRELATIONS FOR THE TWO EVALUATIONS

	All Subjects	Elementary	Secondary
All Subjects	.744	.797	.688
Experimental	.582	.688	.445
Control	.810	.842	.781

Achievement

A variable pattern of trends existed with the achievement scoring for the control and experimental groups from one evaluation to the next. Figure 1 shows the mean achievement results along with the mean ability score for each of the four groups.

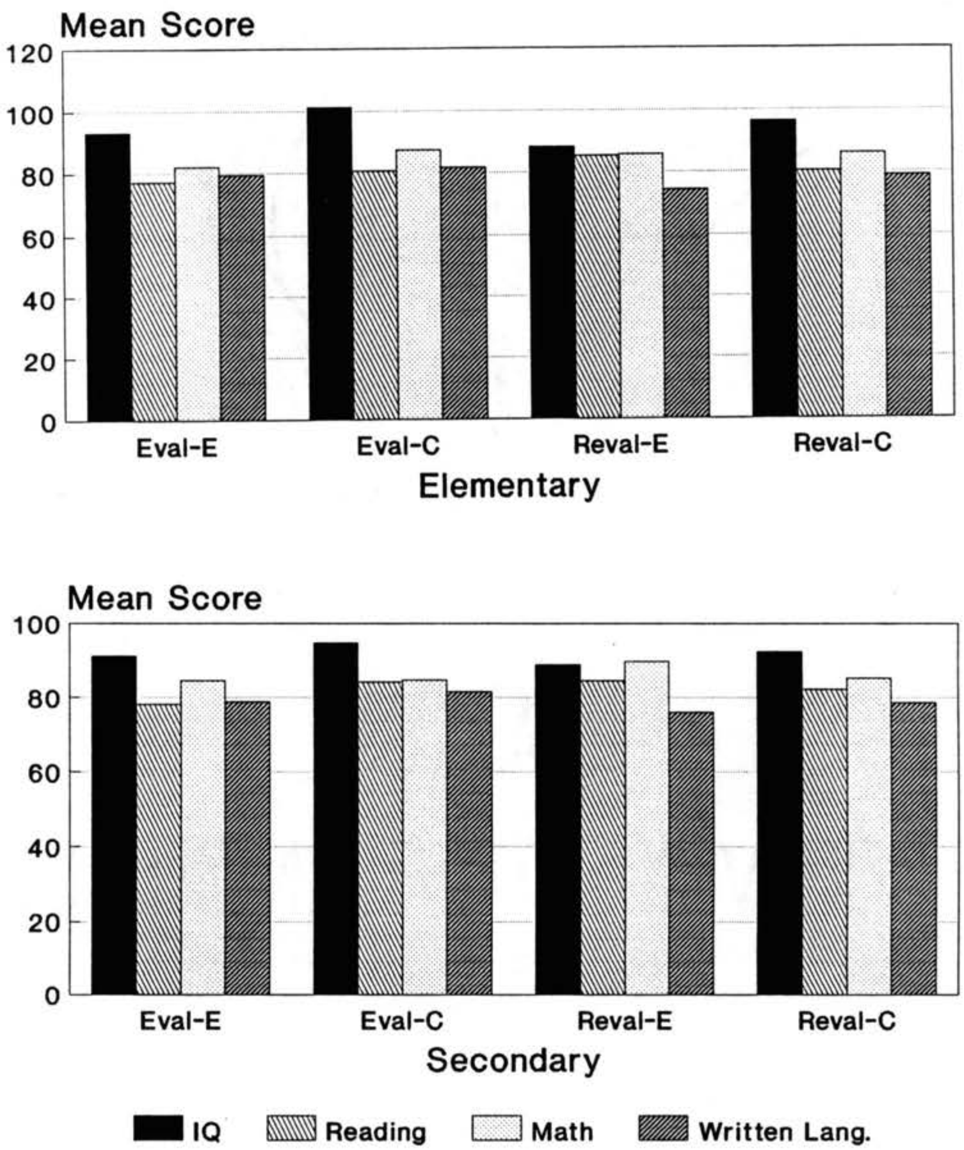

Figure 1. Comparison of Mean Ability and Achievement Scores

For the control groups, all achievement mean standard scores were lower upon re-evaluation except for the math scores of the secondary students which increased slightly. The largest declines were in written language at both the elementary and secondary levels. Elementary reading scores declined the least.

Figures 2 and 3 present elementary and secondary achievement scores for the two evaluations. For the

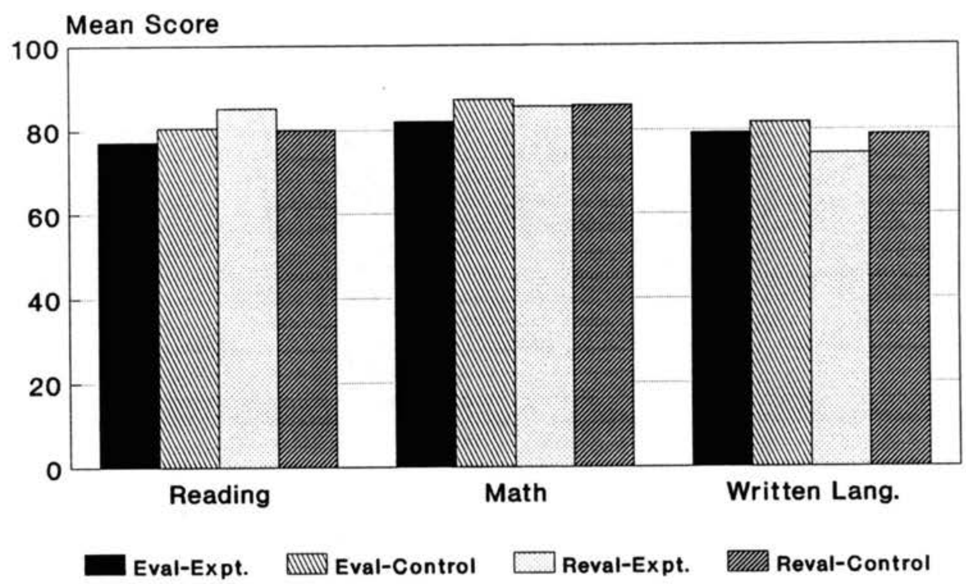

Figure 2. Mean Achievement Scores for Elementary

experimental groups, reading and math achievement was higher at the second evaluation for both elementary and secondary. Written language achievement was lower at the second

TABLE III

MEAN ABSOLUTE DIFFERENCE (DISCREPANCIES)

	Expt.	Control	Elem.	Secondary
1st Evaluation:				
Reading	17.55	17.28	19.30	15.53
Math	12.87	13.55	15.00	11.42
Written Lang.	16.87	18.66	18.42	17.05
Re-Evaluation:				
Reading	10.13	16.15	14.22	12.07
Math	8.73	12.27	12.50	8.50
Written Lang.	17.97	18.90	18.85	18.02

secondary subjects across all three academic areas at both evaluations. Math difference scores were smaller than mean reading and written language difference scores for both elementary and secondary at both evaluations. All mean difference scores were smaller at the re-evaluation except for the secondary and elementary written language scores. The largest magnitude of change was found for elementary reading between the first and second evaluations.

Both the experimental and control groups had smaller mean difference scores in reading and math at the re-evaluation. The written language mean difference scores were larger at the reevaluation for both treatments. Magnitude of change in mean difference scores from one evaluation to the next was largest in reading for the experimental subjects. Changes in written language difference scores were similar for both the experimental and control groups and showed the smallest magnitude of change. All control subjects' mean difference scores were larger than experimental subjects except for reading scores at the first evaluation.

Learning Disabilities Frequencies

Frequencies of learning disabilities eligibility were tabulated at three points in the range suggested by the Oklahoma State Department of Education. Frequencies reported in Table IV represent cumulative counts of LD eligibility at each of the three criterion levels.

TABLE IV

LEARNING DISABILITY ELIGIBILITY FREQUENCIES

	Reading				Math				Written Lang.			
	Ex	C	El	S	Ex	C	El	S	Ex	C	El	S
1st Evaluation:												
> -15	37	33	19	31	25	22	28	19	34	34	35	33
> -18	31	25	23	21	17	17	21	13	26	30	29	26
> -21	24	30	29	15	8	9	11	6	17	22	21	17
Re-Evaluation:												
> -15	12	33	27	18	12	30	21	9	34	43	39	38
> -18	10	27	22	15	8	22	15	4	27	36	33	30
> -21	9	22	16	15	6	10	8	3	19	27	25	21

Ex = Experimental C = Control
El = Elementary S = Secondary

Reading discrepancy scores qualified fewer students in the experimental groups when the WJR was used at both the elementary and secondary levels at all criterion levels. The ineligible students doubled at the -15 and -18 criterion levels for both the elementary and secondary levels in experimental reading groups at re-evaluation when the WJR was used. For control subjects, reading eligibility frequencies did not diminish at the re-evaluation.

Math disabilities eligibility followed a pattern similar to the reading disabilities eligibility. Fewer experimental group students were eligible for LD placement at the re-evaluation when the WJR was used; especially at the -15 and -18 criterion levels. Control groups continued to have more subjects remaining eligible at re-evaluation than experimental groups when the WJR was used despite the fact that there were very small differences between experimental and control groups in math eligibility at the first evaluation when all subjects were evaluated with the WISC-R.

Written language disabilities eligibility did not follow the similar pattern found with reading and math eligibility. Both control and experimental groups had more eligible subjects at the re-evaluation. Use of the WJR at the re-evaluation did not reduce eligibility in written language as was the case in reading and math.

Elementary and secondary groups also showed a pattern of slightly larger eligibility frequencies in written language at the re-evaluation. The total eligible in written language was similar for both evaluation with some variability at the three criterion levels.

Analyses of Hypothesis

Chi-Squared analyses were done to determine if the occurrence of LD eligibility was evenly distributed at each of the criterion levels (-15, -18, -21) of eligibility. Analyses were done to indicate differences between experimental and control subjects and elementary and secondary subjects at each of the evaluations.

TABLE V
CHI-SQUARE RESULTS OF LD ELIGIBILITY FREQUENCIES FOR TREATMENT GROUPS

	Experimental/Control		
	X^2	df	p-value
1st Evaluation:			
Reading:			
> -15	0.549	1	.459
> -18	2.737	1	.098
> -21	0.906	1	.341
Math:			
> -15	0.315	1	.575
> -18	0.000	1	1.000
> -21	0.069	1	.793
Written Language:			
> -15	0.000	1	1.000
> -18	0.536	1	.464
> -21	0.950	1	.330

TABLE V (CONTINUED)

2st Evaluation:			
Reading:			
> -15	15.680	1	.001 *
> -18	11.293	1	.001 *
> -21	7.359	1	.007 *
Math:			
> -15	14.400	1	.001 *
> -18	7.566	1	.006 *
> -21	2.502	1	.114
Written Language:			
> -15	2.936	1	.087
> -18	2.707	1	.100
> -21	2.256	1	.133

* Significance at 0.01 Level

Table V displays the Chi-Square results of LD eligibility frequencies as distributed at the three criterion levels. Comparisons were made of the eligibility occurrences of experimental and control subjects of the two evaluations.

Each Chi-Square reported in Table V represents the four-cell distribution of LD and Non-LD subjects when experimental and control groups' eligibility frequencies were compared at each of the three criterion levels.

At first evaluation, when all subjects were evaluated using the WISC-R, there were no significantly uneven distributions of LD eligibility in any of the three subject areas. However, at the re-evaluation when the WJR was used, Chi-Square analyses were significant (alpha ,.01) in reading at all three criterion levels (-15, -18, -21) and in math at the -15 and -18 criterion levels. Significantly fewer experimental group LD subjects remained eligible in reading and math when the WJR was used in the assessment battery as compared to control subjects who were re-evaluated with the WISC-R. At the re-evaluation, there were no significant differences in the distributions of experimental versus control eligibilities in written language.

Table VI presents the Chi-Square results of the distributions of eligible and ineligible subjects at each of the three criterion levels at each evaluation when elementary and secondary subjects were compared.

TABLE VI
CHI-SQUARE RESULTS OF LD ELIGIBILITY FREQUENCIES FOR EDUCATION LEVELS

	Experimental/Control		
	X^2	df	p-value
1st Evaluation:			
Reading:			
> -15	2.194	1	.139
> -18	4.089	1	.043
> -21	6.125	1	.013
Math:			
> -15	2.833	1	.092
> -18	2.627	1	.105
> -21	1.713	1	.191
Written Language:			
> -15	0.136	1	.713
> -18	0.536	1	.464
> -21	0.373	1	.528

TABLE VI (CONTINUED)

2st Evaluation:			
Reading:			
> -15	2.880	1	.090
> -18	1.915	1	.166
> -21	0.043	1	.835
Math:			
> -15	6.400	1	.011
> -18	7.566	1	.006 *
> -21	2.502	1	.114
Written Language:			
> -15	0.036	1	.849
> -18	0.301	1	.583
> -21	0.564	1	.453

* Significance at 0.01 Level

At the first evaluation, when all subjects were assessed with the WISC-R, no significantly uneven distributions of eligibility frequencies were found in any of the three subjects areas. The size of elementary and secondary eligibility frequencies were not significantly different.

At the re-evaluation, the only significant Chi-Square was found at the -18 criterion level in math (apha < 0.01). Use of the WJR in the assessment battery at the re-evaluation produced an uneven distribution of math eligibility when elementary and secondary eligibility frequencies were compared at the -18 criterion level. At the re-evaluation when the WJR was used, significantly fewer secondary students than elementary students remained eligible. In contrast, there were no significant differences in the frequencies of math eligibility between elementary and secondary students at the first evaluation.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The purpose of this study was to investigate LD eligibility outcomes when a different intelligence test was used for re-evaluation of a previously identified LD sample. Since special education evaluations are three years apart, maturation effects could confound any findings. To reduce this threat to validity, a control group was used that had two successive evaluations with the same intelligence test.

By inspecting Table I of Mean Ability Scores, it can be seen that mean ability standard scores numerically declined for both control and experimental groups from the first evaluation to the re-evaluation. This finding that LD identified students' ability scores trend down is consistent with previous research by Reynolds (1985b) that found declines in IQ averaging two to three points over three years with a LD sample.

Ability scores of elementary subjects in this study declined more than secondary students. This finding reflects the greater variability of younger children's' IQ scores (Wechsler, 1974) that is seen in larger standard errors of measurement for younger children's' scores.

Control groups' numerical declines were almost identical to experimental groups' declines. Regardless of evaluation instrument used, a majority of the subjects in this study had lower IQ's at re-evaluation indicating the possibility of a common characteristic of the majority of LD students. That is, that the process of labeling, placement, and resultant educational services tends to negatively affect self-esteem and thereby, self-efficacy (Mather & Healy, 1990).

The mean IQ of all subjects for both evaluations was 93.25. Once again, previous research and review (Piotrowski & Siegel, 1986) has indicated that 93.25 falls in the upper part of the range commonly reported for LD samples. Most LD samples have included, then, many students in the low average range of intelligence. The elementary control group (mean of 101.2 at first evaluation; 96.2 at re-evaluation) was the only group within the total sample that did not produce mean ability scores similar to previous research findings.

Correlations between the first IQ score and second IQ score were much stronger for control groups (.810) than for experimental groups (.582). This finding should be expected since the control groups were tested twice with the same IQ test. We should expect higher correlations with the oranges to oranges model.

Elementary subjects had stronger correlations (.797) between successive IQ tests than secondary (.688) groups.

A finding of stronger correlations for elementary subjects is not consistent with the principle of larger SEM's for younger children. This finding was strongly influenced by the high correlation of the elementary control group. The elementary control group correlation (.842) was strongest and the secondary experimental (.445) was the weakest.

Though inspection of Figures 2 and 3 it can be seen that mean achievement scores for all control groups, elementary and secondary, numerically declined in reading and written language. Mean declines were not large though (-2.9 to -.5 standard score points). Control groups' math scores changed very little from one evaluation to the next.

Mean achievement scores for the experimental groups changed differently than control groups from one evaluation to the next. Secondary and elementary experimental subjects' mean reading and math scores rose from the first to the second evaluation. The increases in reading and math achievement had greater magnitude (+7.6 to +3.6 standard score points) than the declines seen with the control groups. Written language scores for both elementary and secondary experimental subjects declined from one evaluation to the next.

It is important at this point to summarize the relationships of the ability and achievement scores for the control and experimental groups. These relationships represent the ability/achievement discrepancies of the LD eligibilities.

For the control groups, IQ's went down and reading and written language scores also went down. Declines in IQ score and reading and written language achievement were relatively similar for the secondary subjects. Similar declines in ability and achievement made it likely that ability/achievement discrepancies did not change much from one evaluation to the next. Secondary control subjects' math scores increased while IQ scores went down. The combination of decreased ability and increased achievement narrowed ability/achievement discrepancies.

Elementary control subjects' IQ scores declined more than reading and written language scores making it likely that ability/achievement discrepancies were slightly smaller. Elementary control subjects' math scores also declined, making for a pattern similar to the reading and written language patterns just described.

Experimental groups' relationships of ability and achievement were much different. Once again, IQ scores declined. However, with both secondary and elementary experimental groups, reading and math scores increased. The decrease of ability and increase of achievement made it likely that discrepancies were smaller. The written language scores of experimental subjects declined making for a pattern in which discrepancies were not likely to change much.

With smaller discrepancies there is less LD eligibility. The pattern in which ability and achievement are converging makes for smaller discrepancies. The converging pattern of the experimental groups in reading and math at the re-evaluation made it more likely that LD eligibility frequencies would be smaller than those found for the control groups. This same pattern also made it likely that there would be significant eligibility differences in math between elementary and secondary subjects at the re-evaluation.

This was borne out by the Chi-Square analyses for the reading and math frequency differences between experimental and control groups at the re-evaluation. These groups had shown no significant differences at the first evaluation. At the second evaluation, both the reading and math achievement had increased for the experimental groups and was converging with the declining IQ scores. While, for the control groups, achievement had decreased along with IQ and was not showing the converging pattern that produces fewer eligibilities.

Chi-Square results were significant for the elementary vs. secondary math differences at only one criterion level (-18) at re-evaluation. This third converging pattern did not produce significant differences in eligibility frequencies at two or more criterion levels at the re-evaluation as did the converging patterns mentioned in the previous paragraph.

Conclusions

The functional focus of this study was to gather data about the LD outcomes of substituting one measure of cognitive ability for another. The mix and match practices of combining a variety of assessment instruments at one point in time and also across time produces confusing LD eligibility results.

The data from this study seems to follow some patterns reported in previous psychometric research. The ability scores and some of the achievement scores of the LD students declined over the three year interval. The declines of achievement for LD students have been established by previous research (Badian, 1988; Horn, O'Donnell & Vitulano, 1983; O'Shea & Valacante, 1986). If one assumes that IQ tests are valid predictors of achievement (Hessler, 1987) then the achievement scores have mirrored declines in relative ability. Results from this study indicate that the assumptions of this prediction model were correct for the control groups with the exception of math outcomes for secondary students. The prediction model had limited validity for the experimental groups, though.

Subjects in the experimental group had increased reading and math achievement. This outcome has more than one possible explanation. It is possible that the effectiveness of the interventions for learning disabilities for the experimental groups increased. Empirical evidence of the history of interventions with subjects in the study tends to discount this idea. No large-scale intervention changes were made in the school district.

A more likely explanation for the achievement increases may be related to the previously suggested achievement loading of the WJR cognitive battery. If the WJR is more similar to achievement testing then test administration practice could improve achievement scores. Typically, cognitive assessment is done first during test sessions. If the cognitive assessment includes tasks more similar to tasks used in achievement testing, then the examinee's short-term memory for achievement responses may be involved. That is, the examinee may be "primed" for more successful achievement.

A question that remains to be answered is: Why did the IQ scores of the LD students in this sample systematically decline over the three-year interval? The declines in IQ could be attributed to the normal variability associated with measurement error. The mean declines were all within the range of 90% probability for retest outcomes.

Another answer for the ability declines is that the same information processing problems that interfere with achievement performance also affect performance on IQ measures (Shepard, 1983). These deficits could produce a cumulative effect that depresses the ability scores of LD students relative to standards for "normal" cognitive development. If we agree with this assumption, we have to support the idea that intelligence tests may not be good predictors of future achievement for children with learning disabilities (Mather & Healey, 1990). This assumption would hold true only in the situation where the achievement of LD students improved because of intervention, though. There is little research evidence existing to support the hypothesis that the size of discrepancy is reduced by intervention.

Another important consideration with the ability testing involves the recency of norming. It has been long established (Kaufman, 1975b) that the norms for an intelligence test are less applicable as time increases from the norming point. As Kaufman established with the revising of the WISC, norms underestimated current samples at an increasing rate as the test aged. That is, group IQ means increased over time.

The idea of the aging of norms, if it is applicable for all intelligence tests, does not coincide with the results from this LD sample's control groups. If applicable, the group IQ means would have increased; they decreased. The phenomenon of "aging norms" may be applicable to the experimental group. The WJR had been recently normed and thus a correction reflecting recent population characteristics had been made. As a result, we would expect current sample means to be lower than for subjects who were being tested using much older norms (WISC-R). However, since both control and experimental mean ability decreased it is more likely that the phenomenon of decreasing ability scores is related to the characteristics of learning disabilities samples and not aging norms.

Another phenomenon noticed with this study's sample was that elementary IQ decreased at twice the rate that secondary IQ declined for both experimental and control groups. This finding appears to be independent of differences in choice of IQ test at re-evaluation. Reliability of IQ norms are sometimes questioned with subjects who are near the youngest or oldest ends of the norming group. Virtually none of the elementary subjects from this sample were young enough to have questionable norms.

The principal research question and hypothesis relating to the ability measures of this study appears to be conclusively answered. The WJR Broad Cognitive Ability Scale produced ability scores that trended similarly to the second assessment for the control groups when the WISC-R was used for re-evaluation, it may not be prudent to suggest that it is appropriate to substitute the WJR for the WISC-R. Significant declines in reading and math eligibility were indicated by this study when the WJR was used. Additionally, this sample's psychometric outcomes may not be representative of LD students in general. An example from this study's data that may illustrate the last point is the finding that math and reading achievement increased for the experimental groups; an uncommon finding in LD research.

The finding of increased achievement for the experimental groups probably points out fundamental differences in the characteristics of the experimental and control groups. Also, these differences should make for consideration of one of the principal limitations of doing research with already identified LD subjects.

When research uses already identified LD subjects, the researcher is forced to rely on the judgements of many people as to whether the individual students are appropriately labeled. My experience from attending numerous eligibility meetings is that a wide variety of students are labeled and for a variety of reasons. Sometimes students are labeled because there is evidence of a specific information processing deficit and specific corresponding achievement deficits. In contrast, sometimes students are labeled who show a global deficit, that is, they are performing poorly in all academic areas. Many educators call this student the "slow learner". Even though these two learner profiles are different, they may both demonstrate ability/achievement discrepancy. The use of ability/achievement discrepancy do not allow us to establish if learning disabilities is something different than underachievement.

Control group subjects had significantly higher IQ's than the experimental subjects. If IQ is a predictor of academic success, then the control groups with the higher IQ's should have had more potential for improving achievement. Results from this study did not prove this hypothesis; experimental subjects with lower IQ's made more achievement gains.

Instead, the significant differences in IQ may point to an anomaly of individual subject differences aggregated through the subject selection process. There may have been more "slow learners" in the experimental groups. These "slow learners", even with lower ability levels, may have more potential for academic improvement with intervention than subjects with higher IQ's and presumed neurological deficit that has affected information processing.

Recommendations

This study focused only on psychometric characteristics of LD students. Presence of a severe discrepancy should not be the sole determining criterion for LD. Severe discrepancy should be considered a necessary but not sufficient criterion for determining learning disabilities. Other factors such as motivation, persistence, and interest significantly influence an individual's academic performance. These factors need to be incorporated into our thinking and our practice when determining who is eligible for learning disabilities classification. Future research with already identified LD samples should include measurement of motivation, effort and academic interest to get a broader picture of the factors affecting more easily identified outcomes like discrepancy.

Future studies addressing discrepancy over time should also incorporate some method of assessing the effectiveness of LD intervention. All of the subjects in this study were provided with a similar intervention; pull-out services in which the student went to a LD resource room for individual lesson plans and instruction. Much different outcomes in discrepancy over time are possible with different interventions.

This study should be redesigned to gather additional data about the appropriateness of substituting one ability measure for another when re-evaluating LD eligibility. A more appropriate design would involve re-evaluating LD subjects with both the WJR and the WISC-R or WISC-III at the same time to eliminated some of the maturational threats to validity that are involved in looking at ability/achievement discrepancies over time.

More research is needed to verify if the declines in ability seen with this LD sample are common. If IQ decline is common for LD students, some method of addressing this trend should be included in the process of determining continued eligibility at successive evaluation points. Or the procedure for identifying LD students could be altered to eliminate the use of IQ and severe discrepancy.

BIBLIOGRAPHY

Algozzine, B. & Ysseldyke, J. (1983). Learning disabilities as a subset of school failure: The oversophistication of a concept. Exceptional Children, 50, 242-246.

Badian, N. (1988). The Prediction of good and poor reading before kindergarten: A nine-year follow-up. Journal of Learning disabilities, 21, 98-103.

Bateman, B. (1969). Learning disabilities - yesterday, today and tomorrow. Exceptional Children, 2, 25-31.

Binet, A. & Simon, T. (1916). The Development of Intelligence in Children. (E.S. Kit, translator). Baltimore: Williams & Wilkins.

Bond, G. & Tinker, M. (1957). Reading Difficulties: Their Diagnosis and Correction. New York: Appleton-Century-Crofts.

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

Cattell, R. (1963) . Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54, 1-22.

Chalfant, J. & Scheffelin, M. (1969). Central processing dysfunction in children: A review of research. (NINDS Monograph No. 9) Washington, D.C.: U.S. Dept. of Health, Education & Welfare.

Chalfant, J. (1989). Learning disabilities: Policy Issues and Promising Approaches. American Psychologist, 44, 392-398.

Clements, S. (1966). Minimal brain dysfunction in children: Terminology and identification. Washington D.C.: U.S. Dept. of Health, Education, and Welfare.

Coleman, M. & Harmer, W. (1985). The WISC-R and Woodcock-Johnson Tests of Cognitive Ability: A comparative Study. Psychology In The Schools, 22, 127-132.

Cone, T. & Wilson, L. (1980). Critical Issues in Operationalizing Identification Criteria for Learning Disabilities. Paper presented at meeting of the National Association of School Psychologists, Washington, April 1980.

Cone, T. & Wilson, L. (1981). Quantifying a severe discrepancy: A critical analysis. Learning Disability Quarterly, 4, 359-371.

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

Das, J. (1972). Patterns of cognitive ability in nonretarded and retarded children. American Journal of Mental Deficiency, 77, 6-12.

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

Federal Register. (1977, December 29). Washington, DC.

Finucci, J. & Childs, B. (1981). Are there really more dyslexic boys than girls? In A. Ansara, N. Geschwind, A. Galaburda, M. Albert & N. Gartrell (Eds.) Sex Differences In Dyslexia, pp. 1-9. Towson, MD: Orton Dyslexia Society.

Gallagher, J. (1986). Learning disabilities and special education: A critique. Journal of Learning Disabilities, 10, 595-601.

Gardner, H. (1983). Frames of Mind: The Theory of Multiple Intelligences. New York: Basic Books.

Gearheart, G., Weishahn, M. & Gearheart, M. (1988). The Exceptional Student In the Regular Classroom. Columbus, OH: Merrill Publishing Company.

Goddard, H. (1910). A measuring scale of intelligence. Training School, 6, 146-155.

Guilford, J. (1967). The Nature of Human Intelligence. New York: McGraw-Hill.

Hall, R., Reeve, R. & Zakreski, J. (1979). Validity of the Woodcock-Johnson Tests of Achievement for learning disabled students. Journal of School Psychology, 22, 193-200.

Hallahan, D. & Bryan, T. (1981). Learning Disabilities. In J. Kaufmann & D. Hallahan (Eds.). Handbook of Special Education, pp. 141-164. Englewood Cliffs, NJ: Prentice-Hall.

Hammill, D., Leigh, J., McNutt, G. & Larsen, S. (1981). A new definition of learning disabilities. Learning Disability Quarterly, 4, 336-342.

Harris, A. (1970). How To Increase Reading Abilities. New York: David McKay.

Hessler, G. (1987). Education Issues Surrounding Severe Discrepancy. Learning Disabilities Research, 3(1), 43-49.

Horn, J. (1967). Intelligence - why it grows, why it declines. Transaction, 5, 23-31.

Horn, J. (1968). Organization of abilities and the development of intelligence. Psychological Review, 75, 242-259.

Horn, J. (1972). State, trait, and change dimensions of intelligence. British Journal of Educational Psychology, 42, 159-185.

Horn, J. (1976). Human abilities: A review of research and theory in the early 1970's. Annual Review of Psychology, 27, 437-485.

Horn, J. (1978). Human Ability Systems. In P. Baltes (Ed.), Life-Span Development and Behavior, 1, 211-256. New York: Academic Press.

Horn, J. (1985). Some thoughts about intelligence. In R. Sternberg & D. Detterman (Eds.), What is Intelligence? Contemporary Viewpoints on it's Nature and Definition, pp. 91-96. Norwood NJ: Ablex.

Horn, J. (1985). Remodeling old models of intelligence. In B. Wolman (Ed.), Handbook of Intelligence, pp. 267-300. New York: Wiley

Horn, J. & Cattell, R. (1966). Refinement and test of the theory of fluid and crystallized intelligence. Journal of Educational Psychology, 57, 253-270.

Horn, J. & Cattell, R. (1967). Age differences in fluid and crystallized intelligence. Acta Psychologica, 26, 107-129.

Horn, J. (1988). Cognitive diversity: A framework for learning. In P. Ackerman, R. Sternberg & R. Glaser (Eds.), Learning and Individual Differences. New York: W.H. Freeman.

Horn, W., O'Donnell, J. & Vitulano, L. (1983). Long-term follow-up studies of learning disabled persons. Journal of Learning Disabilities, 16, 542-555.

Jensen, A. (1970). A theory of primary and secondary familial retardation. In N. Ellis (Ed.), International Review of Research in Mental Retardation, 4, 33-105. New York: Academic Press.

Johnson, D. & Myklebust, H. (1967). Learning Disabilities: Educational Principles and Practices. New York: Grune & Stratton.

Kaluger, G. & Kolson, C. (1969). Reading and Learning Disabilities. Columbus, OH: Charles E. Merrill.

Kaufmann, A. (1975a). 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.

Kaufmann, A. (1975b). Intelligent testing with the WISC-R. New York: Wiley & Sons.

Mather, N. & Healey, W. (1990). Deposing aptitude-achievement discrepancy as the imperial criterion for learning disabilities. Learning Disabilities: A Multidisciplinary Journal, 1(2), 40-48.

McGrew, K. (1984). Normative based guides for subtest profile interpretation of the Woodcock-Johnson Tests of Cognitive Ability. Journal of Psychoeducational Assessment, 2, 141-148.

McGrew, K. (1986). Clinical Interpretation of the Woodcock-Johnson Tests of Cognitive Ability. Orlando FL: Grune and Stratton.

McGrew, K. (1987). A multivariate analysis of the Wechsler/Woodcock-Johnson discrepancy controversy. Journal of Psychoeducational Assessment, 5, 49-60.

McGrew, K. & Pehl, J. (1988). Prediction of future achievement by the Woodcock-Johnson Psychoeducational Battery and the WISC-R. Journal of School Psychology, 26, 275-281.

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35, 1012-1027.

O'Shea, L. & Valacante, G. (1986). A Comparison over time of relative discrepancy scores of low achievers. Exceptional Children, 53, 253-259.

Perlmutter, B. & Parus, M. (1983). Identifying children with learning disabilities: A comparison of diagnostic procedures across school districts. Learning Disability Quarterly, 6, 321-328.

Phelps, L., Rosso, M. & 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. & Falasco, S. (1985). Multiple regression data using the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability. Psychology In the Schools, 22, 46-47.

Piotrowski, R. & Siegel, D. (1986). The IQ of learning disability samples: A re-examination. Journal of Learning Disabilities, 19(8), 492-493.

Poplin, M. (1981). The severe learning disabled: Neglected or forgotten? Learning Disability Quarterly, 4, 330-335.

Reynolds, C. (1985a). Measuring the aptitude-achievement discrepancy in learning disabilities diagnosis. Remedial and Special Education, 6(5), 37-48.

Reynolds, C. (1985b). Critical measurement issues in learning disabilities. The Journal of Special Education, 18, 451-476.

Sattler, J. (1986). Assessment of Children, 3rd Edition. San Diego: Jerome M. Sattler.

Sax, G. (1989). Principles of Educational and Psychological Measurement and Evaluation. (3rd Edition). Belmont, CA: Wadsworth Publishing Co.

Shepard, L. (1983). The role of measurement in educational policy: Lessons from the identification of learning disabilities. Educational Measurements: Issues and Practice, 2(3), 4-8.

Shinn, M., Algozzine, B., Marston, D. & Ysseldyke, J. (1982). A theoretical analysis of the performance of learning disabled students on the Woodcock-Johnson Psycho-Educational Battery. Journal of Learning Disabilities, 15, 221-226.

Snyderman, M. & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. American Psychologist, 42, 137-144.

Spearman, C. (1927). The Abilities of Man. New York: Macmillan.

Stern, W. (1914). The Psychological Methods of Testing Intelligence. Baltimore: Warwick & York.

Sternberg, R. (1986). Intelligence Applied: Understanding and Increasing Your Intellectual Skills. San Diego: Harcourt Brace Jovanovich.

Strauss, A. & Lehtinen, L. (1947). Psychopathology and Education of the Brain-Injured Child. New York: Grune & Stratton.

Swanson, H. (1985). Assessing learning disabled children's intellectual performance: An information processing perspective. In K. Gadow (Ed.), Advances in Learning and Behavioral Disabilities, 4, 225-272. Greenwich CT: JAI Press.

Terman, L. (1916). The Measurement of Intelligence. Boston: Houghton-Mifflin.

Thompson, P. & 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.

Thurstone, L. (1938). Primary mental abilities. Psychometric Monographs, No. 1.

Tuddenham, R. (1962). The nature and measure of intelligence. In L. Postman, (Ed.), Psychology In the Making, pp. 469-525. New York: Knopf.

Vernon, P. (1950). The Structure of Human Abilities. New York: Wiley.

Vernon, P. (1965). Ability factors and environmental influences. American Psychologist, 20, 723-733.

Wechsler, D. (1949). Manual for the Wechsler Intelligence Scale for Children. San Antonio: The Psychological Corporation.

Wechsler, D. (1958). The Measurement and Appraisal of Adult Intelligence, 4th Edition. Baltimore: Williams & Williams.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children - Revised. San Antonio: The Psychological Corporation.

Wesman, A. (1968). Intelligent testing. American Psychologist, 23, 267-274.

White, W. & Wigle, S. (1985). Patterns of discrepancy over time as revealed by a standard-score comparison formula. Learning Disabilities Research, 2(1), 14-20.

Wilkinson, L. (1984). SYSTAT: The System for Statistics. Evanston, IL: SYSTAT, Inc.

Woodcock, R. (1984). A response to some questions raised about the Woodcock-Johnson: The mean score discrepancy issue. School Psychology Review, 13, 342-354.

Woodcock, R. & Mather, N. (1989). Woodcock-Johnson Tests of Cognitive Ability: Examiner's Manual. Allen, TX: DLM Teaching Resources.

Yerkes, R. (1917). The Binet versus the point-scale method of measuring intelligence. Journal of Applied Psychology, 1, 111-122.

Ysseldyke, J., Shinn, M. & Epps, S. (1980). A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability. (Research Report #36). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

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

APPENDIX

INSTITUTIONAL REVIEW BOARD FOR HUMAN SUBJECTS RESEARCH APPROVAL

OKLAHOMA STATE UNIVERSITY
INSTITUTIONAL REVIEW BOARD
FOR HUMAN SUBJECTS RESEARCH

Date: 03-24-93 **IRB#:** ED-93-071

Proposal Title: LD ELIGIBILITY: STANDARD SCORE PATTERNS OF DISCREPANCY OVER TIME COMPARING THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN-R AND WOODCOCK-JOHNSON TESTS OF COGNITIVE ABILITY-R

Principal Investigator(s): Joseph Pearl, Dennis Tomlinson

Reviewed and Processed as: Exempt

Approval Status Recommended by Reviewer(s): Approved

APPROVAL STATUS SUBJECT TO REVIEW BY FULL INSTITUTIONAL REVIEW BOARD AT NEXT MEETING.
APPROVAL STATUS PERIOD VALID FOR ONE CALENDAR YEAR AFTER WHICH A CONTINUATION OR RENEWAL REQUEST IS REQUIRED TO BE SUBMITTED FOR BOARD APPROVAL. ANY MODIFICATIONS TO APPROVED PROJECT MUST ALSO BE SUBMITTED FOR APPROVAL.

Comments, Modifications/Conditions for Approval or Reasons for Deferral or Disapproval are as follows:

Signature: Marcia L. Tilley
Chair of Institutional Review Board

Date: March 24, 1993

VITA

Dennis A. Tomlinson

Candidate for the Degree of

Doctor of Philosophy

Thesis: LEARNING DISABILITY ELIGIBILITY: COMPARING THE WISC-R AND THE WOODCOCK-JOHNSON TESTS OF COGNITIVE ABILITIES-REVISED

Major Field: Applied Behavioral Studies

Biographical:

Personal Data: Born in Boise City, Oklahoma in 1951, the son of Bill and Wilmath Tomlinson. Married to Susan Elaine Bunney in 1982.

Education: Graduated from Holdenville High School, Holdenville, Oklahoma; received a Bachelor of Science degree in Education (1973) from University of Oklahoma, Norman; and Master of Science in Applied Behavioral Studies (1986) from Oklahoma State University, completed requirements for the Doctor of Philosophy degree at Oklahoma State University, Stillwater, in July, 1993.

Professional Experience: Social Studies teacher, McLain High School, Tulsa Public Schools, 1974-1977; teacher at Tulsa County Alternative High School, 1978-1979; child-care worker at Tulsa Boys Home, 1979-1981; special education teacher (SED) for Bristow, Oklahoma, Public Schools, 1981-1982; education therapist for in-patient adolescence unit at Willowview Hospital, Spencer, Oklahoma, 1982-1983; teacher at alternative school, Santa Ana, California, 1984; teacher at Central Oklahoma Juvenile Treatment Center, 1986-1989; graduate assistant for ABSED Department, Oklahoma State University, 1989-1991; school psychologist, Ponca City, Oklahoma, Public Schools, 1991 to present.