COMPARISON STUDY OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN-THIRD EDITION AND DIFFERENTIAL ABILITIES SCALES IN THE IDENTIFICATION OF STUDENTS WITH LEARNING DISABILITIES

CHRIS CERETTI

To the Graduate and Research Council:

I am submitting herewith a Field Study written by Chris Ceretti entitled "Comparison Study of the Wechsler Intelligence Scale for Children-Third Edition and Differential Ability Scales in the Identification of Students With Learning Disabilities." I have examined the final copy of this Field Study for form and content, and I recommend that it be accepted in partial fulfillment of the requirements for the degree of Education Specialist, with a major in School Psychology.

Dr. Charles Grah, Major Professor

We have read this Field Study
and recommend its acceptance:

Second Committee Member

Third Committee Member

Accepted for the Council:

Dean of the Graduate School

STATEMENT OF PERMISSION TO USE

In presenting this Field Study in partial fulfillment of the requirements for an Education Specialist degree at Austin Peay State University, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this Field Study are allowable without special permission, provided that accurate acknowledgment of the source is made.

Permission for extensive quotation from or reproduction of this Field Study may be granted by my major professor, or in his absence, by the Head of Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this Field Study for financial gain shall not be allowed without my written permission.

Signature Chris Ceretti

Date 8/30/02

COMPARISON STUDY OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN-THIRD EDITION AND DIFFERENTIAL ABILITIES SCALES IN THE IDENTIFICATION OF STUDENTS WITH LEARNING DISABILITIES

A Field Study

Presented for the

Education Specialist

Degree

Austin Peay State University

Chris Ceretti

August 2002

ACKNOWLEDGMENTS

The author sincerely thanks Dr. Charles Grah, my committee chairman and Field Study advisor, for his encouragement, guidance, and support during the course of my graduate study. Appreciation is also extended to Drs. Ann Harris and Jean Lewis for their invaluable assistance and contributions in this study.

The author also extends special thanks and appreciation to her husband and family for their unfailing support, patience and understanding throughout the process of completing this project.

ABSTRACT

The current research investigated the relationship between cluster and composite scores obtained for the Wechsler Intelligence Scale for Children, Third Edition (WISC III) and the Differential Ability Scales (DAS) to determine if these two instruments are comparable when used in the assessment of students referred for possible learning disabilities. Fifty (34 male and 16 female) students of various ethnic origins, attending rural Georgia schools, were initially evaluated using the WISC III and determined to be eligible for special education services through the learning disabilities program. Upon re-evaluation, these same students were assessed using the DAS and the WISC III to determine continued eligibility for special education services and to investigate the possibility of practice effects occurring when the same instrument (the WISC III) was used at the time of initial evaluation and then consistently for re-evaluation.

Using descriptive and inferential statistics, the cluster and composite scores, mean and standard deviations, and significant differences relevant to the identification of students with learning disabilities were obtained for the WISC III and DAS. Scores were examined for correlations and similarities to determine if scores obtained on these two instruments were comparable in assessing students with learning disabilities.

Results of the current study indicated strong correlations between the cluster and composite scores of the WISC III and DAS. Findings supported the hypothesis that these two instruments were comparable for use in evaluation of students with learning disabilities.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Individual intelligence tests are administered in the school setting to help make decisions regarding exceptionality, eligibility and educational placement for students identified as having learning disabilities. In June of 1997, Public Law (P.L.) 101-476, the Individuals with Disabilities Education Act, was reauthorized and amended becoming Public Law (P.L.) 105-17, the Individuals with Disabilities Education Act (IDEA) Amendments of 1997 (IDEA, 1997). P.L. 105-17 defines a specific learning disability as "...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, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia" (IDEA 1997, 34 CFR Subpart A, 300.7).

Approximately 3-5% of public school children in the United States are referred for individual assessment with an estimated total of more than 250 million standardized tests (group and individual) administered on a yearly basis (Smith, Smith, Matthews, & Kennedy, 1993). It is estimated that approximately 73% of students referred for individual testing are found to be eligible to receive special education services (National Information Center for Children and Youth with Disabilities; NICHCHY, 2000). In an article published in *Science*, Roush (1995) stated that "if learning to read, write or do math at expected levels were a disease, American school children would be in the middle of an epidemic" (p. 1896).

In the school setting, IQ tests are administered at the time of the initial referral to help determine eligibility for special education services and then again approximately every three years thereafter to determine continued eligibility for special education (SPED) services. Because students receiving special education services are required to be re-evaluated approximately every three years, questions about administering the same IQ test to the same individual are raised. Are practice effects occurring as a result of re-administration of the same test? Is it necessary to select alternative intelligence tests which are comparable with the original IQ test given? Does administering an alternative, psychometrically comparable, intelligence test help ensure test-retest stability and avoid practice effects which result from repeated administrations of the same test? Are the results of the different tests comparable in the classification process (i.e., are students initially classified as learning disabled also classified as learning disabled on the basis of the second test)?

Practice effects are often difficult to distinguish from changes which may have occurred due to educational interventions, personal experiences, length of time between test administrations, regression to the mean, or age-based item content (test items which tap different abilities at different ages but claim to be measure only one ability despite age differences) (Sattler, 2001). Practice effects may result in inflated scores and/or different diagnostic impressions, which can lead to inaccurate estimates of students' actual cognitive abilities and ineffective educational decisions. Using alternative intelligence tests which are psychometrically comparable with previous IQ tests given may or may not help preserve consistency in the outcome of the classification and evaluation process.

The selection of similar and appropriate alternative instruments for the measurement of cognitive abilities in learning disabled students requires that the two instruments have comparable properties. Using comparable tests may reduce the negative effects of repeatedly administering the same test and help preserve the integrity of the results.

To eliminate practice effects, many school psychologists prefer to give a different intelligence measure at the time of re-evaluation rather than using a previously administered measure. When choosing a different measure, several factors must be considered: the sample sizes should be comparable for the two IQ tests given, the standard deviations and sample distributions should be equivalent, the two different measures should provide similar concurrent and construct validities, the two tests should measure the same abilities, and there should be long-term stability of the test-retest scores for the traits being measured (Dumont, Cruse, Price, & Whelley, 1996).

Selecting the most psychometrically sound and appropriate instrument is critical in providing the most useful information when faced with making educational decisions. Test results are only one part of the eligibility determination process in the identification of students with specific learning disabilities. Placement decisions which are made based on the student's test performance, classroom performance and other available data have a permanent effect on a student's educational future. Two intelligence (IQ) tests with reportedly sound psychometric properties are the popular Wechsler Intelligence Scale for Children, Third Edition (WISC III; Wechsler, 1991) and the less well-known Differential

Ability Scales (DAS; Elliott, 1990a). The WISC III is one of the most widely used individual IQ tests for the identification of students with specific learning disabilities. The WISC III has been reported by Kaufman (1994) to be especially good at detecting patterns of specific learning difficulties. It is considered by many school psychologists to be the most effective and efficient IQ test available, because of its ease of administration and scoring, and its interesting test materials which help make the testing experience more positive for the students being evaluated. Many students who receive special education services have been evaluated using the WISC III one or more times, typically at the time of the initial evaluation and/or again at the time of re-evaluation (Hutton, Dubes, & Muir, 1992; Stinnett, Havey, & Oehler-Stinnett, 1994). Results may be impacted by practice effects that occur with repeated administration of the same test. This can lead to inappropriate and ineffective educational decisions which may negatively impact program planning and the student's academic success.

One of the more recently introduced instruments, which is gaining popularity in educational settings, is the Differential Ability Scales (DAS; Elliott, 1990a). The DAS, developed by Colin Elliott, is based on the British Ability Scales (BAS; Elliott, Murray, & Pearson, 1979) which is widely used in Great Britain as a means of evaluating school-aged children with an instrument sensitive to the British culture and normed on the British student population. Development of the DAS also provided a new paradigm of evaluative practice based on research relating to Piaget's theories of cognitive development and Thurstone's (1938) primary mental abilities.

The Differential Abilities Scales (DAS) was designed by Elliott (1990a) to be a reliable cognitive assessment battery which would provide interpretable data about a student's cognitive processing strengths and weaknesses across a wide range of cognitive domains. The emphasis of the DAS is placed on a student's abilities (Braden, 1992; Platt, Kamphaus, Keltgen, & Gilliland, 1991) rather than a general definition of intelligence as is true with the WISC III. Elliott believes that to accurately assess learning deficits, cognitive processes should be evaluated to identify factors which contribute to the disability. This information is beneficial when designing educational intervention strategies, since it allows for specific cognitive processes to be examined which may negatively impact academic achievement.

This study will explore the positives and negatives of using the WISC III and DAS IQ tests, interchangeably, in the initial assessment and re-evaluation of special education students (SPED) for evaluation, placement and educational planning. Issues which will be explored will include whether or not reliability and validity of test-retest results are possibly compromised when using different IQ tests, possible changes in classification of students which may occur due to test results which may vary, and whether or not identification of specific patterns of learning disabilities change when using different IQ tests.

CHAPTER 2

REVIEW OF LITERATURE

Students who are experiencing significant academic difficulties are often referred for a comprehensive psychoeducational evaluation to determine eligibility for special education (SPED) services. The Individuals With Disabilities Education Act of 1990, 1992, 1997 (IDEA) and P.L.94-142 (Education for All Handicapped Children Act of 1975) and P.L. 105-17 are federal mandates that provide the guidelines for psychoeducational assessment of students with possible learning disabilities.

A comprehensive psychoeducational evaluation includes the administration of an individually administered intelligence test (IQ) to provide an estimate of the student's current level of cognitive functioning and an individually administered achievement test to assess the students level of academic achievement in relation to their cognitive ability. Individually administered IQ tests are administered at the time of the initial evaluation, and again at the time of re-evaluation, in conjunction with individual achievement tests, to determine if a specific learning disability is present. If a learning disability is identified at the time of initial evaluation, a re-evaluation which includes the administration of an individual IQ and achievement tests, is conducted approximately every three years.

Psychoeducational assessments are conducted to provide information needed to help identify specific learning deficit areas, to aid in identifying students who may be experiencing social, emotional and/or behavioral problems, to help guide eligibility determinations for special education (SPED) services or continued SPED services, to provide information used to evaluate and monitor student progress, and to provide information helpful in the modification of Individual Educational Programs (IEP) as needed (Dumont, Cruse, Price, & Whelley, 1996; Ross-Reynolds, 1995).

Most students with learning disabilities are believed to have average IQ scores, (ranging from approximately 85 to109 depending on the particular IQ test administered) suggesting average cognitive abilities and functioning (Swanson,1996; Dumont, Cruse, Price, & Whelley, 1996). The assumption underlying the "average IQ" theory when identifying learning disabled students comes from the idea that children with average IQ scores who experience academic difficulties in reading, writing or math have distinct cognitive processing deficits which result in lower academic achievement scores (academic scores which show a discrepancy between scores obtained for the overall IQ and scores obtained for the achievement test of 15 or more) (Braden, 1992). However, students with low IQ scores (approximately 85 and below) have also shown discrepancies of 15 points or more between the IQ and achievement scores and are not considered to be learning disabled. Traditionally, these students have been considered to be slow or mildly intellectually disabled learners (Swanson, 1996). Thus, the use of the "average IQ" theory and discrepancy scores to identify learning disabled students becomes problematic when delineating between learning disabilities, slow learners and mildly intellectually disabled students (Swanson & Christie, 1994).

In a study by Glutting, McDermott, Konold, Snelbaker and Watkins (1998) the authors noted that strong relationships between general intelligence and school achievement are probably the most documented findings in empirical psychoeducational research. These authors attribute this finding to the reported criterion-related validity of IQ tests and subtest analysis of specific external criteria used to identify patterns of cognitive strengths and weaknesses important to remedial intervention planning.

Although the "average IQ" theory has traditionally been the underlying assumption for determining whether or not a student has a specific learning disability, this theory does little to help explain the differences between cognitive ability and academic performance for the student with a learning disability, or multiple learning disabilities, who doesn't have an average IQ. Significant ability-achievement discrepancies are also frequently found in students with below average and above average IQ scores. In fact, about 3% of school-aged children with specific learning disabilities are also identified as gifted students with above average to superior IQ scores (McDermott & Glutting, 1997).

Typically, students with significant ability-achievement discrepancies tend to experience either specific cognitive processing deficits or neurological impairments which negatively impact their academic performance and intellectual abilities to a marked degree. Other factors commonly associated with significant ability-achievement discrepancies are cultural and environmental disadvantages, economic disadvantage, lack of adequate instruction and education, multicultural backgrounds, emotional and behavioral disorders, medical problems, poor school attendance, school transfers, lack of motivation, and/or hearing and vision difficulties (Glutting, McDermott, Konold, Snelbaker & Watkins, 1998).

Careful consideration of all possible factors and accurately determining the area of cognitive processing deficit contributing to an identified learning disability are critical in the development of appropriate intervention techniques designed to help the student experience academic success. Selecting the most appropriate and psychometrically sound IQ test is certainly one of the most crucial factors in the outcome of psychoeducational evaluations and intervention planning, since not all IQ tests are equal. IQ tests are selected based on their psychometric properties and the skills and abilities they are purported to measure. Psychologists administering IQ tests must be familiar with the properties of the instruments they administer, as well as be keenly aware of which ones are appropriate based on the reason for referral.

While psychologists strive to provide the most accurate and useful information for assessment of children with learning disabilities, controversies over the measurement of human abilities and the use of IQ tests in identifying learning disabled students continue, because no theory or model has established how standardized IQ tests measure "potential", and no single theory or model has gained universal acceptance (Wallace, Larsen & Elksnin,1992, p.106; Waterman, 1994). Yet, IQ tests continue to represent the "measuring stick", because of their diagnostic properties for establishing baseline cognitive abilities against which to compare achievement levels (Detterman & Thompson,1997).

Although strong psychometric relationships for criterion-related validity between scores obtained for overall general intelligence and academic achievement were reported in the study by Glutting, et al (1998), the use of IQ and achievement tests in identification of students with learning disabilities has raised questions about the accuracy of using this method for eligibility and placement decisions (Detterman & Thompson,1997). Often IQ scores fail to adequately and consistently predict achievement levels (Swanson, 1996).

One example is when ability-achievement discrepancies are found between verbally related cognitive abilities and reading achievement levels for students who obtain higher Verbal IQ scores (Lyon, 1996). The Verbal IQ score is typically considered to represent verbal abilities associated with reading and general communication skills (Stanovich & Siegel, 1994). Verbal IQ scores are typically higher for students whose economic, educational and cultural environments are more richly endowed with verbal experiences and communications. Therefore, students with higher Verbal IQ scores are not expected to be identified as having a reading disability when the IQ score is compared to the reading achievement scores. However, the higher Verbal IQ score often results in a significant discrepancy (15 point difference between IQ and achievement score) either in word decoding, phonetic analysis, and/or reading comprehension. When an individual obtains a higher Verbal IQ score and a significantly discrepant reading achievement score, a verbally associated cognitive processing deficit is inferred based on the discrepancy method for identifying learning disabilities. The positively skewed Verbal IQ score will also tend to positively skew the overall or composite IQ score. When this occurs, a learning disability may be identified in other academic areas (e.g., math and/or writing). And when this occurs, the "measuring stick" becomes less reliable in effectively predicting academic achievement and identifying learning disabilities (Detterman & Thompson, 1997; Swanson, 1996).

Brody (1997) noted that IQ scores can change over time, and more importantly, that they are relevant to the individual's educational and environmental opportunities in conjunction with verbally associated cognitive processing abilities. Many students with below average IQ scores, who lack adequate verbal communication experiences through their culture, educational opportunities or environment, also exhibit reading disabilities (Stanovich & Siegel, 1994). The student who obtains a below average IQ score may be experiencing significant verbally associated cognitive processing deficits along with a paucity in verbal communication exposure and experiences.

The majority of students (52%) receiving SPED services through the public schools are identified as having a Specific Learning Disability (SLD) in the areas of reading, math, written expression, oral expression or listening comprehension (USDOE, 1999). Many of the students identified also meet eligibility criteria for more than one learning disability (Swanson, 1996).

With an estimated 73% of referred students nationwide found to be eligible for special education services (NICHCHY, 2000), providing special education services for students identified as having special learning needs is not only federally mandated, but is also big business and an expensive business. The most current educational expenditure information was published in a national study completed by the National Center for Education Statistics (NCES, 1995). Findings of this study estimated that approximately $2,780 a year is spent per pupil in regular education programs in public elementary and secondary schools. Four years later, NCES (1999) reported that the estimated expenditure per special education student was approximately $6,335. The National Center for Education Statistics stopped requiring the collection of state/local educational cost data after 1987-1988 because they decided that the accurate collection of comprehensive educational expenditure data was too difficult to obtain on a national basis

(Chambers, Parrish, & Lieberman, 1999). State and local school systems strive to maintain data for the students in their respective districts; however, accurate national data has not been accurately or responsibly reported since 1988 (USDOE, 1994).

Most of the more current data is reported according to estimates calculated by the National Center for Education Statistics (NCES, 1995) which notes that part of the problem in collecting accurate data is that costs vary considerably across states (e.g., the average expenditure per student ranged from $2,758 in Indiana to $8,501 in Connecticut) and disability categories. The increased costs realized in special education are related to the legal requirements that guide the provision of special education services, including the provision for specially trained educators and specifically designed learning materials used for the delivery of instruction to students with special needs, the requirement that licensed/certified psychologists conduct the psychoeducational assessments, and the expense of the instruments used in conducting these assessments. IQ test materials are extremely costly and the most trusted and widely used instruments also tend to be the most expensive. Ensuring that a student has been appropriately and accurately assessed through the use of psychometrically comparable instruments and providing for the student's unique learning needs are mandated legal regulations in special education, although not always funded mandates. These factors contribute to the selection of the most appropriate and effective measures for the assessment of learning disabilities.

Once a student's cognitive abilities have been measured using an individually administered IQ test, the student's IQ score is compared with their academic achievement scores in one or more of the seven learning areas (Oral Expression, Listening

Comprehension, Basic Reading, Reading Comprehension, Written Expression, Math Calculation and Math Reasoning). Differences of 15 points or more (one standard deviation or more) between cognitive ability and academic achievement scores indicate the presence of a learning disability. For example, if the IQ score is 100 an academic score of 85 (which is one standard deviation below the mean) in one or more of the seven learning areas would be needed to establish that a student has a learning disability. A learning disability is generally defined as a deficit in cognitive processing which negatively impacts a student's ability to learn and achieve at the same rate as his non-learning disabled, age-related peers (Salvia & Ysseldyke, 1995).

According to the interpretation of a learning disability provided by Glutting, McDermott, Konold, Snelbaker and Watkins (1998), a learning disability is the result of a cognitive processing deficit which affects an individual's ability to interpret what he or she sees or hears (visual or auditory perception which are not linked to specific visual or hearing impairments) or his or her ability to link information from different parts of his or her brain in a manner which makes sense (information processing). However, it is important to note that not all learning problems are learning disabilities. Some students develop skills slower or faster than other students and/or may experience maturation delays which can effect academic success (NIMH, 1993).

A careful study of available information and research on the DAS will be explored first, since it is the less well-known instrument for the measurement of cognitive ability. Next, research and the effectiveness of the WISC III in the identification of students with learning disabilities will be discussed, followed by current research on the comparability of the DAS and WISC III in assessing students for possible learning disabilities.

In a study conducted by Platt, et al (1991) the theoretical basis of the DAS and its properties were explored. This study points out that the DAS avoids using the term "intelligence" to describe what it is measuring. Elliott (1990c) believes that the term ability is a more accurate description for explaining the factors which the DAS measures than the term intelligence which is neither clearly defined nor globally accepted. The basic constructs of the DAS are not founded on a loose definition or single model of intelligence, but are based on a hierarchal model of abilities. This approach departs from Wechsler's (1939) definition of intelligence as the "aggregate of global capacity of the individual to act purposefully, to think rationally, and deal effectively with their environment (p.3)." Wechsler's definition of intelligence presumes that general intelligence, or *g*, is more than the sum of its parts, or individual intellectual abilities. Elliott (1990c) believes that Wechsler's beliefs led him to develop intellectual subtests that are not good measures of psychometric *g*. For Elliott, Wechsler's theory of intelligence, which is the foundation for his psychometric tests, is too general and vague and its test items do not adequately assess the domains measured.

The DAS is not based on any single theory of human ability or broad definition of intelligence as is the WISC III. The DAS was designed to assess specific abilities (specific domains of performance) which provide distinct information about an individual's cognitive strengths and weaknesses across a range of cognitive domains (Elliott, 1990b). The WISC III was designed to provide standard or global scores which represent an individual's capacity for purposeful actions (as explained by Wechsler's definition of intelligence).

Elliott avoids the terms intelligence and IQ because the General Conceptual Ability (GCA) score obtained on the DAS is defined somewhat differently than the composite scores of the WISC III and Stanford-Binet Intelligence Test, Fourth Edition (SB:4E) which use the term intelligence in their titles. The Wechsler IQ tests and the Stanford-Binet Intelligence Test, Fourth Edition adopt a relatively broad definition of intelligence as reflected in Wechsler's (1939) statement:

> "One of the greatest contributions of Binet was his intuitive assumption that in the selection of tests, it made little difference what sort of tasks you used, provided that in some way it was a measure of the child's general intelligence" (p.6).

WISC III subtest and composite IQ scores are based on a diverse collection of tasks, some of which have low *g* loadings (Elliott, 1990b). For all ages, the DAS uses a relatively small number of core subtests with high *g* loadings which contribute to the calculation of the General Conceptual Ability (GCA) score. Elliott believes that the GCA score is a more pure and homogeneous score and, therefore, a more interpretable measure of psychometric *g*. Elliott defines psychometric *g* as "the general ability of an individual to perform complex mental processing that involves conceptualization and the transformation of information."

The DAS Handbook (Elliott, 1990b) operationally defines the psychometric properties of *g* as: "[...the first component in a principal-component analysis, the first factor in a common-factor analysis, or the most general factor in a hierarchical factor analysis (Jensen, 1980, 1987)... Tests with the highest *g* loadings are the ones that best define the nature of the underlying variable. The DAS composite GCA score consists of subtests which load highest on the first common factor... Many studies of various batteries of mental tests indicate that the tests with the highest *g* loadings also measure the most complex mental functions.., p. 19]."

For Elliott, in order to effectively interpret individual ability, each cluster score and the GCA score must be derived from a set of subtests whose contents measure tasks with a common (homogeneous) dimension of ability rather than a variety of diverse (heterogeneous) abilities. The WISC III subtest and composite scores were designed to cover a diverse range of tasks, processes and knowledge, according to Elliott. "Subtests that contribute to a composite score should be similar in the sense that they correlate highly with a common group factor or with the instrument's operational definition of psychometric g (Elliott, 1990b, p.19). In contrast to the composite scores of many other individually administered test batteries, which give equal weighting to all subtests, the GCA score of the DAS is derived from only those subtests with high *g* loadings. For Elliott, this feature allows for efficiently obtaining a valid and focused measure of the central component of intellectual ability (Elliott, 1990b).

The DAS norm sample included 3,475 children stratified by age (2.6 to 17.11 years old), sex, race/ethnicity, parent education, geographic region and educational preschool enrollment. The sample also included children identified as having learning disabilities, speech and language impairments, educable mental retardation, gifted abilities, emotional disorders, and those with mild visual, hearing or motor impairments. Bias analyses included the performance of ethnic groups such as African-Americans and

Hispanic children to help identify and eliminate potentially biased items (Elliott, 1990c). The DAS is designed to measure abilities on 17 cognitive and three achievement subtests. Individual subtest scores, the General Conceptual Ability (GCA) score and the Special Nonverbal Ability scores, which measures conceptual and reasoning abilities with minimal verbal components, and cluster scores for measuring more specific abilities, all have a mean of 100 and standard deviation of 15.

Elliott (1990a) looks at the abilities measured by the DAS like pieces of a puzzle, each important in their own right, but coming together to give a global assessment of individual ability. Elliott calls this outcome of ability measurement the General Conceptual Ability (GCA). Elliott notes that the General Conceptual Ability is not the final assessment of an individual's intelligence. For Elliott, the General Conceptual Ability is the top rung of a psychometric hierarchy.

During the development of the DAS, computer modeling was used to rate each item according to its difficulty. Elliott perceives this design as allowing for a more systematic division of subtest items which can be selected to provide a more reliable and statistically accurate estimate of individual ability in specific areas of concern. Average DAS internal consistency reliabilities of subtests range from .70 to .94 and retest reliability correlations for the GCA scores are reported to be .88 to .93 (Elliott, 1990b).

Keith (1990) raised the question of how well the DAS actually measures ability as defined by Elliott (1990b). Keith (1990) attempted to investigate this question by conducting a confirmatory and hierarchical analysis of the DAS using data from the original standardization sample (N = 3475). Keith's findings supported the assertion that the DAS measures a general intellectual ability (*g*) with a reported range of primary factor loadings from .328 to .984. These studies also found some interesting differences in the secondary factor loadings. The main difference between the initial model of the DAS and its final structure was noted to be a loss of a quantitative factor and the addition of a factor Keith (1990) labeled as "Gf" (fluid intelligence which is essentially nonverbal mental efficiency involving adaptive and new learning capabilities further defined by Horn, 1998). Abilities in the areas of verbal reasoning, nonverbal reasoning, memory and speed were retained while the lowest second order loadings were seen in the areas of memory and speed. The results were reported to support the hierarchical structure of the DAS and identification of factors related to the assessment of fluid intelligence ("Gf").

Another question that arose among potential users of the DAS (Elliott, S., 1990) was how well the DAS findings related to the identification of students with specific learning disabilities. Stephen Elliott (1990) investigated the potential of the DAS to provide information about an individual's abilities. Results reportedly allow for more specific interpretations and ability profiling related to cognitive deficits (individual strengths and weaknesses) when predicting ability-achievement discrepancies in the identification of students with specific learning disabilities. The DAS was designed to yield distinct information related to a wide range of common abilities with sufficiently reliable subtest specificity which, according to Elliott (1990b), provides scores that are more individually interpretable. Elliott bases this idea on his belief that differences between DAS scores are more meaningful since scoring is based on the Rasch Model of item response theory (IRT). The Rasch Model of item response theory (IRT) is based on the belief that:

> "One or more characteristics or traits determine a person's observed response to test items. Because these characteristics are not directly observable or measurable, they are termed latent traits (or in the case of cognitive tests, abilities). An IRT model specifies an expected relationship between observable responses on a test and the unobservable traits or abilities assumed to underlie those responses. The trait or ability is a quantitative dimension on which both individuals and test items can be placed." (Elliott, 1990b, p.332)

Latent traits, or abilities, are measured using a latent trait model (LTM) to analyze test items. This method of assessment refers to a measurement procedure developed to provide test items that are believed to have common discriminating abilities (item difficulty and item discrimination) across groups which may differ widely in ability. Item discrimination, item difficulty and the probability of a correct response occurring by chance are measured using IRT or LTM methods and provide important information about the items and responses being evaluated (Sattler, 2001). The probability of a person passing an item on the DAS depends on the individual's ability level and the difficulty of the item. The assumption in this model is that if an individual's ability to solve an item and the difficulty of the item are at the same point on a common scale, the odds of successful solution to the problem are even. If an individual's ability is at a point higher than the difficulty level, the odds are even greater for successfully solving a task. Likewise, if the item's difficulty level is higher than the individual's ability, the odds for a successful solution are lower.

The Rasch model, as applied to the DAS, allows for test score interpretation and analyses using tests of goodness of fit and unidimensionality. Predetermined statistical criterion (e.g., criterion referenced tests) for the DAS were established to measure ability on the test items administered. Dimensionality refers to the number of latent traits or abilities which underlie an individual's performance on test items and helps explain or account for individual differences in performance. Criterion-referenced test scores are based on abilities or behaviors measured rather than comparison to a norm group as with norm referenced tests such as the WISC III (Lyman, 1991). In calculating DAS scores, raw scores are converted to ability scores based on the Rasch model of IRT. The ability and difficulty values within a DAS scale have equal-interval measurement characteristics, for example, item-response theory models such as the Rasch are designed so that any difference between two ability scores, or an ability score and a difficulty value, maintain the same interpretative properties at any point on the measurement scale. The probability of a person's passing a test item depends solely on the ability of that person and the difficulty of the item (Elliott, 1990b). DAS scores are calculated based on the number of items actually administered and does not allow for credit awarded for items not administered or attempted.

Kercher and Sandoval (1991) were also interested in determining the DAS's proficiency in accurately identifying students with specific learning disabilities. Kercher and Sandoval compared the scores obtained on the DAS for 30 learning disabled students (in the area of basic reading) against the scores obtained by 30 non-learning disabled students matched for age, ethnicity and gender. The scores for the learning disabled students were reported to be within 5 points of the mean for the ability subtests, but significantly below those of the non-learning disabled students on the achievement subtests

(Word Reading, $p < .001$; Spelling, $p < .001$; and, Basic Number Skills, $p < .01$). These results supported the authors theory that use of the DAS, both ability and achievement subtests, were accurate and effective in identifying students with reading disabilities.

McIntosh and Gridley (1993) demonstrated similar support for using the DAS to identify students with learning disabilities. Scores obtained for the Generalized Conceptual Ability (GCA) and achievement subtests by 83 of the learning disabled students in the original DAS standardization sample were analyzed using cluster analysis. The analysis yielded six homogenous subgroups identified as generalized (Generalized Conceptual Ability score fell within the below average range with consistently lower achievement scores), high functioning (GCA score fell within the high average range with consistently lower achievement scores), normal (GCA score fell within the average range with consistently lower achievement scores noted), underachievement (GCA score fell within the below average range with achievement scores commensurate with the GCA score), borderline (GCA achievement scores all fell within the borderline range with no significant discrepancies found), and dyseidetic (GCA scores fell within the below average range with significant differences noted between the GCA and Verbal performance scores). Relative strengths were found in verbal ability with significant differences between the GCA and achievement scores in Reading and Spelling for the dyseidetic subgroup.

McIntosh and Gridley (1993) conducted discriminant analyses between the ability and achievement scores to determine the effectiveness of differentially classifying subgroups of learning disabled students. The discriminant function averaged 78% accuracy in correctly classifying learning disabilities in the generalized, high functioning, underachievement and dyseidetic subgroups. The obtained GCA and achievement score discrepancies were found to be consistent with identification of a learning disability. Findings supported the use of the full DAS battery for the identification of students with learning disabilities.

Shapiro, Buckhalt and Herod (1995) also found a significant discrepancy between the DAS General Conceptual Ability (GCA) and achievement scores for the 83 students identified as learning disabled in the DAS standardization sample. The DAS ability scores were reported to be strongly correlated with the archival Wechsler Intelligence Scale-Revised (WISC-R) IQ scores. However, differences were noted in the cognitive processes purported to be measured by the DAS and WISC-R. Shapiro, Buckhalt and Herod (1995) concluded that their results agreed with the findings of Kercher and Sandoval (1991) and McIntosh & Gridley (1993), further supporting the DAS's diagnostic validity and utility for identifying students with learning disabilities.

Glutting, McDermott, Konold, Snelbaker & Watkins (1998) researched the criterion-validity of unusual subtest profiles from the DAS using multivariate-nomothetic, univariate-nomothetic and univariate-ipsative comparisons. The authors were investigating significant differences which might occur between students who had a known disability and those without a known disability. One thousand, two hundred students were selected based on similar demographic information for participation in this study. From the 1,200 students, 60 students were found to have unusual DAS score profiles based on criterion validity. These 60 students were matched by characteristics and GCA against 60 students in a control group. These two groups were compared according to criteria established for eligibility and placement in special education (SPED) classes, three norm referenced achievement tests, and six standardized behavioral scales completed by the students' teachers. The authors hypothesized that they should find criterion differences between the groups. However, no differences were found which suggested that the unusual DAS subtest profiles were not helpful in determining a specific disability category. Glutting, et al determined that a more accurate method of comparison would be one that utilized predictive rather than concurrent criteria. Subtest profile analysis was limited by score comparisons without examination of criterion-related factor deviation quotients (factor deviation quotients, unlike subtest scores, are based on criterion-related validity of their constructs supported by factor analysis).

McDermott (1995) was interested in learning more about children's abilities with similar demographic characteristics, ability, achievement and adjustment profiles. Using the DAS normative sample of 1,200 children, ages 5-17, McDermott evaluated ability scores obtained on the DAS verbal, nonverbal and spatial subtests and their respective achievement scores in the areas of reading and basic math. Adjustment constructs associated with attention deficit hyperactivity disorder, solitary aggressive-provocative, solitary aggressive-impulsive, oppositional defiant, diffident and avoidant disorders were assessed using the Adjustment Scales for Children and Adolescents (ASCA) completed by the student's teachers. Demographic information, ability and adjustment scores were analyzed using canonical and multiple regression procedures. Demographics, notably social class and ethnicity, accounted for approximately 18.9% of the students' ability score variations. Race and ethnicity accounted for 13.5% of variations in ability scores. Social and emotional adjustment were reported to account for only 5.5% of the variance in ability and achievement. Age and gender accounted for 3.1% of the variation in adjustment. Maturation alone was noted to account for 60% of the variations in ability. McDermott concluded that development of separate norms would be beneficial when assessing ability across differing demographic groups.

Youngstrom, Kogos and Glutting (1999) studied the incremental validity of the DAS factor scores in predicting academic achievement. One thousand, one hundred and eighty-five students were grouped according to gender, ethnicity, geographic origin, parental education and educational classification. Descriptive statistical findings revealed that scores obtained fell within the expected levels on all measures of ability and achievement with no significantly differing variability across samples. Standard deviations fell close to the expected value of 15. Associations between ability and achievement were reported as similar across learning and non-learning disabled students. These authors also noted that the DAS GCA provides a more accurate measure for predicting achievement than the DAS factor (ability) scores since multiple factor scores tend to weaken inferential accuracy.

In a study conducted by McDermott and Glutting (1997) hierarchical regression and discriminant models were used to differentiate maximum ability potential, variations in achievement levels, individual learning styles, test-session behavior, and differing classifications of learning disabled students. Ipsative score comparisons revealed that overall cognitive ability scores were more useful in interpreting achievement weaknesses than subtest ability scores. McDermott and Glutting believe that use of subtest scores, instead of using global ability scores, alter the meaningful interpretive information related to individual differences and abilities provided by global ability scores. These authors argue against the use of ipsative subtest score comparisons based on mathematical relationships which they report as showing weak correlations (Holland & McDermott, 1996; McDermott, Fantuzzo, Glutting, Watkins & Baggaley,1992; Watkins & Kush, 1994) between ipsative and conventional IQ subtest scores. For McDermott and Glutting (1997) unusual subtest profiles result from either chance variations of less reliable subtests or administration errors.

The Wechsler Intelligence Scale for Children-Third Edition (WISC III) is the third revision of David Wechsler's intelligence scales for children. The first Wechsler intelligence scale for children was published in 1949. The WISC III is an individually administered intelligence test used in the assessment of intellectual ability in children ages 6 through 16 years old. The WISC III is based on David Wechsler's theory of intelligence which describes intelligence as the total and comprehensive sum of skills and behaviors which contribute to intelligent behavior, rather than a measure of any particular ability. Wechsler believes intelligence is best understood as being "the capacity of an individual to act purposefully, to think rationally, and to deal effectively with his/her environment" (Wechsler, 1991, p.1). He goes on to explain that intelligence is more the act of thinking and acting which reflects an individual's general level of current cognitive ability.

The WISC III consists of six subtests in the verbal scale and seven in the performance scale. Subtest index scores yield a Verbal IQ score, a Performance IQ score and a Full Scale IQ score (Wechsler, 1991). Scoring for the WISC III differs somewhat from the scoring method used for the DAS. WISC III scoring is based on derived scaled scores (often used in norm referenced instruments) which are determined based on a linear standard score rather than raw scores converted to ability scores, as with the DAS.

WISC III derived scores (e.g., standard scores, percentile ranks, stanines, normal curve equivalents, age and grade equivalents, and ratio IQ scores) are calculated from raw scores and represent an individual's level of performance in relationship to scores obtained by the norm group used. WISC III scoring allows for credit to be given for items above the basal which were not actually administered or attempted. Linear standard scores represent the original distribution of raw scores and their standard score equivalents as set by the test author (Lyman, 1991). Using this method allows for transformations from one kind of score (such as derived scaled scores) to another if the assumption is that a normal distribution is used based on the same group of individuals. Derived scores are relatively independent of content difficulty because they are based on the individual's score as compared with the performance of other people in a comparative or normative group. If the test content is designed to be more difficult, an individual's raw score is likely to be lower than it would be on an test with easier content items. Because the difficulty of the test item content influences the scores of all examinees participating in the standardization sample, it is sometimes possible to use the same test for individuals or groups ranging widely in level of ability because the test constructor can aim for test items of about 50 percent difficulty. Fifty percent is believed to be the best difficulty level (from a measurement point of view) since it permits the largest number of inter-individual comparisons and discriminations (Sattler, 2001).

The intellectual factors measured by the WISC III are noted in the manual (1991) to be abstract reasoning, verbal comprehension, perceptual and spatial skills, and processing speed. These factors are identified as providing an estimate of an individual's cognitive strengths and weaknesses. These cognitive strengths and weaknesses are identified through the 12 subtests of the WISC III, which are divided into two groups of tasks - Verbal (expressive and receptive language and memory skills) and Performance (perceptual-motor skills and visual processing). Scaled scores obtained on these specific tasks are converted into three composite standard scores yielding the Verbal, Performance and Full Scale IQ scores. In addition, norms are provided for four factor based index scores which are reported to measure verbal comprehension, perceptual organization, freedom from distractibility, and processing speed (Wechsler, 1991).

Keith and Witta (1997) explored the reliability, validity and constructs purported to be measured by the WISC III. Their research centered around a main question - what constructs does the WISC III actually measure? Four aspects of intelligence were identified by Wechsler and became known as the first order abilities of intelligence: verbal comprehension (verbal ability), perceptual organization (nonverbal ability), freedom from distractibility (attention and concentration), and processing speed (speed of information processing). Keith and Witta note that the WISC III manual fails to provide adequate information (factor loadings and factor correlations) regarding the factor analyses used to confirm these four first order abilities. The authors were also troubled by the fact that the manual fails to report test findings for the actual theoretical structure of the WISC III.

They report that what is actually measured is global cognitive ability, or general intelligence (*g*).The four first order abilities are being measured through a variety of subtests which are derived from Wechsler's informal theory of *g*. Index scores are calculated from raw scores and reported to measure the first order cognitive abilities identified. The Full Scale IQ score is considered to be a measure of *g*. This structure could have been measured according to Keith and Witta (1997), but the WISC III manual fails to use a hierarchical confirmatory factor analyses to measure *g*. Abilities (perceptual skills, processing speed, attention and concentration) which are measured using less cognitively demanding tasks have lower loadings on *g* than factors associated with general intelligence (induction, abstraction, reasoning, and complex mental operations) which typically load highly on *g*. The first order constructs (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed) reportedly measured are residuals of second order general intelligence factors (*g*) which Keith and Witta say are not actually confirmed through factor analysis in the WISC III manual or explained by the Full Scale IQ score.

Lyon (1995) conducted a study designed to compare the IQ scores obtained on the WISC-R and WISC III for learning disabled students. Forty students were given the WISC-R and the WISC III at the time of re-evaluation. All forty students were previously identified as having a learning disability. Results showed significantly lower WISC III Verbal, Performance and Full Scale IQ scores when compared to these same scores obtained on the WISC-R. Differences ranged from approximately one-third to one-half standard deviations (5-8 points). Scores obtained on the WISC III prompted Lyon to recommend caution when conducting re-evaluations using scores obtained on the WISC III to estimate ability-achievement discrepancies when the student was previously tested using the WISC-R. The author points out that these types of differences often occur at the time of restandardization of IQ tests and that performance may often show a deflation in scores when based on outdated normative samples. Since the IQ scores appeared to be deflated, the conclusion that a nonsignificant ability-achievement difference occurred would provide inaccurate eligibility information. Changes made to the WISC III from the WISC-R may potentially disqualify students with learning disabilities at the time of re-evaluation, and may tend to limit the number of students identified as having a learning disability at the time of initial evaluation. In these cases, the examiner would be wise to select a test of cognitive ability other than the WISC III.

Schultz (1997) also examined score differences for students with learning disabilities following changes made between the WISC-R and the WISC III. Sixty-two students identified as having learning disabilities, when initially assessed using the WISC-R and Woodcock Johnson Tests of Achievement-Revised (WJ-R), were administered the WISC III and Woodcock Johnson Tests of Achievement-Revised (WJ-R) at the time of triennial re-evaluation. No significant changes in correlations were found when comparing scores obtained on the WISC-R, WISC III and WJ-R. Eligibility, based on a 15 point score discrepancy between the WISC-R, WISC III and WJ-R, decreased from 86% on the WISC-R to 48% at re-evaluation with the WISC III. Although WJ-R score changes were reported to be statistically significant, WISC-R and WISC III scores for students retaining eligibility were not significantly different from scores obtained by students who were no longer eligible for special education services. Schultz (1997) also noted that score discrepancies between ability and achievement are less frequently found with the WISC III than with the WISC-R. He concluded that changes made to the WISC III were more likely to result in ineligibility for special education services, especially when used with the WJ-R for assessing learning disabilities. Use of the WISC III and the WIAT, when evaluating the possibility of a learning disability, were reported by Schultz (1997) to show more reliable correlations between ability and achievement scores, since these two instruments were co-normed on a common standardization sample.

The potential for error in identification of students with learning disabilities, using the WISC III and the WISC-R, was also explored in a study conducted by Slate and Jones (1997). Slate and Jones hypothesized that "if lower re-evaluation IQ scores actually reflect lower levels of intelligence, comparable declines in IQ scores should also occur when the WISC-III is used for both the initial evaluation and at the time of re-evaluation"(p.200). Stability of WISC III IQ scores obtained at initial evaluation and re-evaluation were compared for a sample of 34 students found to be eligible for special education services (22 of these students were identified as having a learning disability and 12 were identified as mildly intellectually deficient). Paired *t*-tests showed nonsignificant differences between the WISC III Verbal, Performance and Full Scale IQ scores from the time of initial evaluation to re-evaluation. WISC III test-retest scores were also noted to be significantly correlated..

Groth-Marnat (1996) discusses the use of the Wechsler Intelligence Scales as the model for assessment practices in educational and clinical settings. The Wechsler intelligence tests are noted by these authors to have become the most frequently used instruments for the measurement of intellectual abilities. According to Groth-Marnat, the WISC III tasks were designed based on abilities most valued by Western society as representing intelligence, i.e., tasks related to and considered to be predictive of relevant skills. Test-retest reliability coefficients ranged from an average of .95 (VIQ), .91 (PIQ), and .96 (FSIQ) across ages 6 to 16. The greatest fluctuations were noted on the Performance subtests ranging from .54 to .93, across all ages. Groth-Marnat cites an asset of the WISC III as "providing valuable information about a person's cognitive strengths and weaknesses" (p.125) when compared to age-related peers. However, in providing a global assessment of an individual abilities, the WISC III may be somewhat biased toward middle and upper class socioeconomic levels, since the sample used appears to over-represent these individuals and norms may not be applicable to those with lower socioeconomic backgrounds. A lack of sufficient data related to the validity of the WISC III is also noted by Anastasia (1997). Internal consistency reliabilities for the WISC III subtests are somewhat lower than those for the VIQ, PIQ and FSIQ (Sattler, 2001). The average subtest internal consistency reliability coefficients ranged from .69 to .87 across all ages. Criterion validity studies (Wechsler, 1991) between the WISC III and DAS were reported as showing that the WISC III FSIQ correlated highly with the DAS General Conceptual Ability (GCA) at .92. Correlations (Sattler, 2001) for the WISC III and DAS Verbal scales were noted as .71, for the Performance/Nonverbal scales as .81, and for the Spatial Conceptual subtests of the DAS and the Verbal and Performance scales of the WISC III as .66 (Verbal) and .82 (Performance) and .86 (FSIQ).

Macmann and Barnett (1997) conducted a comprehensive study of Kaufman's approach to the reliability of interpretations related to educational decisions based on WISC III IQ test results. The authors believe that Kaufman's model for estimating internal-consistency and test-retest reliability for the WISC III did not adequately provide for various sources of error, e.g., practice effects, scoring errors, examiner differences, etc., which could negatively affect individual profile patterns when evaluating test results. Analysis of the VIQ-PIQ differences, factor index scores and ipsative profile patterns was conducted. Two large independent samples were analyzed for several variables which affect the reliability of composite scores including reliability of profile composite scores, method for determining individual strengths and weaknesses, number of subtests administered, and method of calculating ipsative subtest deviations.

Results showed that Kaufman's (1994, p.6) assumption "that the limitations of IQ testing can be overcome through skilled detective work" did not provide substantial evidence for reliable profile analysis important to educational planning and academic improvement. Use of the WISC III to determine intellectual strengths and weaknesses contributed little to a better understanding of a student's learning difficulties due to various sources of error which Kaufman neglected to include in his approach to explaining the benefits of "intelligent testing" (Macmann & Barnett, 1997). Macmann and Barnett reported that error rates for interpretation of VIQ - PIQ differences were substantial (40.5% of the standardization sample showed VIQ - PIQ differences of 11 or more points in either direction - either VIQ > PIQ or PIQ > VIQ. Kaufmann (1994) recommends a minimum 19 point discrepancy between VIQ - PIQ in order for these scores to be interpretable). Factor index score differences (reliability of differences between factor scores ranged from .56 to .83 for combined effect, test-retest, and parallel forms) and ipsative profile patterns (r =.56 for the mean ipsative composite) on the WISC III raised important questions about the use of the WISC III in making educational decisions.

The WISC III Full Scale IQ (FSIQ) score was found to be the best predictor of achievement levels when compared with achievement scores obtained on its co-normed achievement test, the Wechsler Individual Assessment Test (Glutting, Youngstrom, Ward, Ward, & Hale, 1997). Glutting, et al (1997) evaluated the Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed Index factor scores in conjunction with results of four of the subtests of the Wechsler Individual Assessment Test (WIAT) for a sample of referred (N = 636) and nonreferred (N = 283) students. The authors used the FSIQ to predict the outcome of scores obtained on the WIAT reading, math, writing and language subtests. Results were reported to indicate that the FSIQ allowed for approximately one-third to two-thirds of the variance in achievement test scores for both samples.

Nichols and Ward (1998) conducted a comparison study of the DAS and WISC III scores for evidence of concurrent validity for a small sample (N = 26) of learning disabled students. Scores were obtained at the time of the student's three year re-evaluation. The students average age was 10.7 years old. The authors focused on the differences of outcomes between the WISC III Full Scale IQ (FSIQ) and DAS General Conceptual Ability (GCA) scores and comparison of the WISC III Verbal IQ (VIQ) and Performance IQ (PIQ) scores and DAS Verbal Reasoning (VR), Nonverbal Reasoning (NVR), and Spatial-Conceptual Reasoning (SCR) cluster scores. WISC III mean IQ scores fell within the low average range (80 to 92). Standard deviations were noted to be somewhat higher than expected ranging from 16 to 20. Significant correlations were reported for the WISC III and DAS IQ scores, with the Nonverbal IQ score consistently being the lowest. Significant differences were found between the WISC III FSIQ and DAS GCA scores with the GCA consistently being the lowest (VIQ and NVR = 8.08 point difference), PIQ and NVR = 11.46 point difference (the DAS NVR measures nonverbal fluid intelligence not measured by the WISC III PIQ). FSIQ and NVR = 8.69 point difference, GCA and VIQ = 3.27 point difference, GCA and PIQ = 6.65 point difference, and GCA and FSIQ = 3.89 point difference. Because the scores obtained on the PIQ and NVR subtests represent nonverbal tasks, the 11.46 difference is considered to indicate that these subtests are measuring unique abilities (nonverbal fluid intelligence measured by the DAS NVR) in learning disabled students (the ability to solve abstract problems using nonverbal reasoning skills). Results showed that the DAS provides an appropriate and equitable alternative to the WISC III when assessing students with learning disabilities.

Significant correlations between the DAS and WISC III cluster scores were also found by Dumont, Cruse, Price and Whelley (1996). Fifty-three students identified as having a learning disability based on the WISC III Full Scale IQ score were re-evaluated using the DAS as part of their regular triennial review. Verbal composite scores for both tests were reported to correlate highly ($r = .77$). Correlations for the DAS Nonverbal and Spatial Reasoning IQ scores and the WISC III PIQ ranged from .65 to .67. The DAS

GCA and WISC III FSIQ were reported to be highly correlated ($r = .78$). Of the children retested using the DAS 3 years after initial evaluation, 25 Full Scale IQ score (47%) were found to show no change in their qualitative intelligence classification. Differing qualitative intelligence classifications were noted for 18 children (34%) when the DAS General Conceptual Ability (GCA) score was the only score considered, and 10 (19%) children were found to show higher GCA scores than WISC III FSIQ scores. These differences in classifications were attributed to trying to directly compare labels attached to exact FSIQ and GCA scores rather than considering the score range or standard error of measurement. Intelligence classification errors, as well as eligibility classifications, tend to occur with as little of a difference as one point obtained when comparing levels of intelligence on different measures based on the exact GCA or FSIQ scores. A more reliable comparison is found when scores are reported in confidence intervals. When the 95% confidence interval scores were used, the DAS GCA and WISC III FSIQ scores differed for only 4% (2 students) found to be classified as learning disabled.

Findings reported by Dumont, Cruse, Price and Whelley (1996) suggest that the WISC III and DAS are compatible measures of intelligence. The DAS Nonverbal Reasoning score (a measure of fluid ability not specifically measured by the WISC III Performance subtests) was reported to represent the most frequently found significant difference between the two measures.

The WISC III manual (Wechsler, 1991) reports findings from a concurrent validity study conducted with a small sample (N = 27, ages 7-14) of students. Standard composite scores obtained on the DAS were compared to composite standard scores obtained on the WISC III. WISC III FSIQ scores were reported to be highly correlated with the DAS GCA scores (r = .92). High correlations were also reported for the WISC III VIQ and DAS Verbal Reasoning scores (r = .87) and the WISC III PIQ scores and DAS Nonverbal Reasoning (r = .78) and Spatial Reasoning scores (r =.82).

The current study was conducted to compare scores obtained, over an approximate period of 3 years, on the WISC III and DAS when used interchangeably for the identification and eligibility classification of students with learning disabilities. Scores obtained on initial evaluation using the WISC III were compared to scores obtained on the DAS at the time of re-evaluation. The WISC III was also administered again at the time of re-evaluation to determine whether or not practice effects were apparent. Results were expected to show a high correlation between scores generated from the WISC III and the DAS for students identified with specific learning disabilities.

Elliott believes that the DAS provides a wider range of measurement for general intelligence (*g*) which provides more distinct information about an individual's cognitive strengths and weaknesses. Based on Elliott's belief that the DAS General Conceptual Ability score results in a relatively homogenous score because it was designed to have high loadings on *g*, investigation of the differences in scores obtained for the WISC III and DAS were also examined. Specific cognitive processes identified with the DAS which may affect learning and allow for more effective intervention and educational planning were also explored.

CHAPTER 3

METHOD

Participants

WISC III and DAS IQ scores from 34 male and 16 female students (ranging in age from 9 to 17 years old, in grades 4-12) of various ethnic origins, attending rural schools in Houston County, Georgia, were compared. All students were previously identified as having a learning disability at the time of the initial evaluation (based on scores obtained on the WISC III), approximately three years previously. Identification of a learning disability at the time of initial evaluation was based upon scores obtained from the standard 10 subtests of the WISC III, established state and federal criteria, and determination by the Placement Committee team members. Students were re-evaluated as part of their regularly scheduled triennial review using the six School-Age core subtests of the DAS and the 10 subtests of the WISC III.

Materials

The WISC III is an individually administered measure of intelligence for assessing students ages 6 years through 16 years, 11 months. The five Verbal subtests of the WISC III (Information, Similarities, Arithmetic, Vocabulary, and Comprehension) were designed to measure verbal abilities and the five Performance subtests (Picture Completion, Coding, Picture Arrangement, Block Design, Object Assembly) were designed to measure nonverbal abilities. The WISC III Verbal IQ (VIQ) composite score, Performance IQ (PIQ) composite score, and the Full Scale IQ (FSIQ) composite score were used for comparison in the current study. Standard scores are based on a mean of

100 with a standard deviation of 15.

Reliability coefficients for the mean composite scores of the WISC III (calculated for ages 6 to 16-11 years old, for 200 students) are reported in the manual, for test-retest reliabilities, as: Verbal IQ (VIQ) .95, Performance IQ (PIQ) .91, and Full Scale IQ (FSIQ) .96.

The DAS School-Age battery is an individually administered measure of cognitive ability designed to assess students ranging in age from 6 to 17 years, 11 months. The six cognitive subtests of the DAS School-Age battery, which yield a Verbal Reasoning (VR) IQ composite score, a Nonverbal Reasoning (NVR) IQ composite score, a Spatial-Conceptual Reasoning (SCR) IQ composite score and a General Conceptual Ability (GCA) IQ composite score, were used for comparison in the current study. The DAS Verbal Reasoning ability subtests include Word Definitions and Similarities, the Nonverbal Reasoning ability subtests include Matrices and Sequential and Quantitative Reasoning, and the Spatial-Conceptual Reasoning abilities subtests include Pattern Construction and Recall of Designs. Standard scores are based on a mean of 100 with a standard deviation of 15.

Reliability coefficients for the mean cluster scores of the DAS School-Age IQ test are reported in the DAS manual as: Verbal Reasoning (VR) .88, Nonverbal Reasoning (NVR) .90, Spatial-Conceptual Reasoning (SCR) .92, General Conceptual Ability (GCA) .95 (N = 200 for a sample population ranging in age from 6 to 17 years, 11 months old).

Procedures

Each participant had been previously identified as having a specific learning disability at the time of initial placement using the WISC III. Participants were administered the six standard subtests of the DAS School-Age battery and the 10 standard subtests of the WISC III at the time of re-evaluation as part of the regularly scheduled triennial review. The WISC III was also administered at the time of re-evaluation (within the past year) to investigate for any possible practice effects which may have occurred.

All scores examined in this study were obtained through archival data collected from Houston County student Special Education records. Consent forms were obtained prior to testing as part of the referral process. All participants were evaluated in accordance with the"Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 1992) and Georgia state and federal special education mandates.

CHAPTER 4

RESULTS

Scores obtained on the WISC III and the DAS were examined for correlations between the WISC III IQ composite scores and DAS cluster scores; similarities and differences between the DAS General Conceptual Ability Scores (GCA) and the WISC III Full Scale IQ (FSIQ) scores; intercorrelations between the WISC III Verbal, Performance, Full Scale IQ scores and the DAS Verbal Reasoning (VR), Nonverbal Reasoning (NVR), Spatial-Conceptual Reasoning (SCR) IQ scores; mean differences for the WISC III and DAS cluster scores and significant differences relevant to the identification of students with learning disabilities.

The following descriptive and inferential statistics were used: Pearson correlations of composite and cluster scores, probabilities, paired sample *t*-tests, obtained mean IQ score correlations, and subtest cluster score correlations for the WISC III and DAS. Additionally, investigation of any changes (increases or decreases) in ability as measured by the WISC III VIQ, PIQ, and FSIQ scores, on repeated administration, was examined. An alpha level of .05 was used for all statistical tests.

Means and Standard Deviations were calculated for the initial administration of the WISC III and administration of the DAS on re-evaluation (Table 1). Descriptive statistics revealed mean IQ scores fell within the Average range (90 - 109) except for the DAS NVR IQ score which fell within the Below Average range (80 - 89). The lower DAS NVR IQ scores were consistent with previous research findings for students identified as having learning disabilities. Standard deviations were found to be lower than the expected value of 15 (notably for the WISC III PIQ first administration, WISC III

VIQ for the second administration, the DAS GCA and WISC III FSIQ at the time of re-evaluation with the standard deviations falling approximately 5 points below the expected value of 15).

Table 1

Means and Standard Deviations of WISC III IQ's and DAS Cluster Scores

Subscales	Mean	Standard Deviation
WVIQ	94.22	12.55
WPIQ	95.80	11.85
WFSIQ	93.88	9.92
DASVR	96.06	11.47
DASNVR	89.74	11.73
DASSCR	95.00	12.41
DASGCA	92.18	10.32

Note. WVIQ = WISC III Verbal IQ; WPIQ = WISC III Performance IQ; WFSIQ = WISC III Full Scale IQ; DASVR = DAS Verbal Reasoning IQ; DASNVR = DAS Nonverbal Reasoning IQ; DASSCR = DAS Spatial-Conceptual Reasoning IQ; DASGCA = DAS General Conceptual Ability IQ.

To evaluate if any changes occurred in ability from the first administration of the WISC III compared to the second administration of the WISC III, mean and standard deviations were calculated (Table 2).

Table 2

Means and Standard Deviations for the WISC III at Re-Evaluation

Subscales	Mean	Standard Deviation
WVIQ2	94.40	9.80
WPIQ2	92.80	11.70
WFSIQ2	91.50	10.30

Findings revealed no significant changes in the WISC III VIQ scores across administrations, $t = -0.173$ (49), $p < 0.864$). However, a statistically significant drop in WISC III PIQ scores at the time of the second administration was obtained, $t = 2.563$ (49), $p < 0.013$), as well as a drop in FSIQ, $t = 2.717$, $p < .009$.

No changes in eligibility status for special education services occurred for students participating in the current study. Special education classifications and eligibility status remained the same with no additional areas of disability identified.

Table 3 presents the intercorrelations among the scores obtained for the initial administration of the WISC III and for the DAS administered at the time of re-evaluation.

Table 3

Correlation Matrix for WISC III and DAS Cluster Scores

	WVIQ	WPIQ	WFSIQ	DASVR	DASNVR	DASSCR	GCA
WVIQ	1.000						
WPIQ	0.130	1.000					
WFSIQ	0.783	0.657	1.000				
DASVR	0.643	0.213	0.571	1.000			
DASNVR	0.169	0.495	0.379	0.103	1.000		
DASSCR	0.309	0.629	0.509	0.410	0.522	1.000	
DASGCA	0.478	0.592	0.641	0.682	0.694	0.858	1.000

Pearson correlations of cluster and composite scores obtained for the initial administration of the WISC III and for re-evaluation using the DAS (administered approximately 3 years after initial evaluation), revealed significant correlations between scores for the WISC III and DAS Verbal and Nonverbal subtests as well as for the WISC III Verbal and DAS GCA, WISC III Performance and DAS Spatial-Conceptual Reasoning subtests and DAS GCA; WISC III Full Scale IQ and DAS Verbal Reasoning, Spatial-Conceptual, and GCA IQ scores.

Pearson correlations of composite and cluster scores, obtained for the initial administration of the WISC III and administration of the WISC III at the time of re-evaluation (administered approximately 3 years after initial evaluation) were also calculated (Table 4). Significant correlations were found for the WISC III VIQ, PIQ, and FSIQ across administrations but not for the Verbal compared to the Performance scores. This finding would be expected since these two subtests are measuring different cognitive abilities.

Table 4

Correlation Matrix for WISC III Cluster Scores at Initial Evaluation and Re-Evaluation

	WVIQ1	WPIQ1	WFSIQ1	WVIQ2	WPIQ2	WFSIQ2
WVIQ1	1.000					
WPIQ1	0.130	1.000				
WFSIQ1	0.783	0.657	1.000			
WVIQ2	0.810	0.287	0.747	1.000		
WPIQ2	0.221	0.749	0.574	0.418	1.000	
WFSIQ2	0.636	0.608	0.805	0.799	0.834	1.000

Table 5 presents Pearson correlations obtained for administration of the WISC III and DAS at the time of re-evaluation. Significant correlations between the WISC III and DAS were found for all composite and cluster scores except for the DAS VR and DAS NVR.

Table 5

Correlation Matrix for WISC III (second administration) and DAS Cluster Scores

	DASVR	DASNVR	DASSCR	DASGCA	WVIQ2	WPIQ2	WFSIQ2
DASVR	1.000						
DASNVR	0.103	1.000					
DASSCR	0.410	0.552	1.000				
DASGCA	0.682	0.694	0.858	1.000			
WVIQ2	0.632	0.328	0.375	0.563	1.000		
WPIQ2	0.312	0.476	0.680	0.634	0.418	1.000	
WFSIQ2	0.594	0.456	0.647	0.730	0.799	0.834	1.000

CHAPTER 5

DISCUSSION

The purpose of the current study was to compare the cluster and composite scores obtained for the WISC III and DAS to determine if they are comparable instruments for assessing students with learning disabilities. Significant correlations were found between the cluster and composite scores for the WISC III and DAS which indicates that these two instruments are comparable for use in evaluating students with learning disabilities. The possibility of practice effects occurring when the WISC III is administered during initial evaluation and again at the time re-evaluation was also explored. There were no indications for practice effects found in the current study. In fact, a drop in scores was found on the Performance subtests from the time of initial evaluation to the time of re-evaluation. No changes in eligibility or classification status occurred for students evaluated in the current study. Findings indicate that using either the WISC III or DAS in the process of evaluating students with learning disabilities, does not appear to significantly affect the results or outcome.

Composite scores obtained for the WISC III at the time of initial evaluation and again at the time of re-evaluation were compared to the cluster scores obtained for administration of the DAS at the time of re-evaluation. Pearson correlations revealed significant relationships for scores obtained on the initial administration of the WISC III and re-evaluation using the DAS. In the current study, the DAS SCR and WPIQ mean scores were found to be more closely correlated than the DAS NVR and WPIQ scores which may suggest that the WPIQ is a better measure of spatial-perceptual abilities than nonverbal abilities as reported by Nichols and Ward (1998).

Slate and Jones (1997) reported stability of WISC III scores obtained at initial evaluation and again at the time of re-evaluation for a sample of 34 students identified as having learning disabilities and mild intellectual disabilities. These authors reported that paired *t*-tests showed nonsignificant differences between the WISC III VIQ, PIQ, and FSIQ mean scores from the time of initial administration to re-evaluation. In the current study, paired *t*-tests showed nonsignificant differences between the WVIQ mean scores from the time of initial administration to re-evaluation. Statistically significant drops in WPIQ and FSIQ scores across administrations may suggest a residual effect associated with an increasing range of task difficulty for age, scoring errors, or examiner interpretation differences on the WISC III Performance subtests. Drops noted in the WPIQ may also be attributed to impulsivity, guessing and random responses, or fatigue on the nonverbal items which are more abstract in nature, or a lack of appropriate and effective educational curriculum. Findings of the current study also support the results obtained by Groth-Marnat (1996) which indicated the greatest fluctuations occurred for the WPIQ scores across ages and administrations.

WISC III Full Scale IQ and DAS General Conceptual Ability composite scores were found to provide the most consistent estimate of an individual's intellectual ability across ages and administrations. Use of subtest scores, rather than the FSIQ or DAS GCA scores, in the identification of students with learning disabilities is not recommended since greater fluctuations between subtest scores across ages and administrations fail to provide the most reliable information related to ability and predicted achievement levels. Because score variations across administrations can result in misclassifications and/or inaccurate interventions and educational planning, score differences and fluctuations should be carefully considered when evaluating learning disabled students.

Results of the current study indicate that use of either the WISC III or the DAS composite/cluster scores provide accurate and effective measurements for the identification of students with learning disabilities. Fluctuations noted in WPIQ and WFSIQ scores from initial administration to re-evaluation, which may be associated with increased difficulty levels of tasks presented across ages, sources of error including scoring mistakes and differences in interpretation across examiners should be more closely investigated.

As Nichols and Ward (1998) stated in their study, the DAS Nonverbal Reasoning subtest may be measuring unique abilities associated with nonverbal, fluid intelligence (the ability to solve abstract problems using nonverbal reasoning skills) which the WPIQ subtests do not measure. The subtests of the DAS Nonverbal Reasoning cluster are believed by Nichols & Ward to rely more on the identification of rules and hypotheses testing for more abstract problems than the WPIQ or DAS Spatial-Conceptual Reasoning subtests. In this regard, the DAS Nonverbal Reasoning items may be more sensitive in identification of students who have difficulties with nonverbal reasoning abilities and/or experience possible learning disabilities associated with nonverbal cognitive processing deficits.

Further studies designed to explore the impact of special education services upon IQ scores are needed to help determine if special education interventions and modifications have a negative or positive effect on IQ scores at the time of re-evaluation.

Other possible reasons for changes in scores across administrations (such as personal experiences and learning, regression to the mean, age-based item content and item difficulty, examiner differences, scoring errors, length of time elapsed between evaluations, fatigue, random or impulsive responding, or guessing) should also be more closely evaluated to help determine possible underlying factors which may affect score fluctuations.

LIST OF REFERENCES

REFERENCES

Anastasi, A. (1997). Psychological Testing (7th ed.). New Jersey: Prentice Hall.

Braden, J. P. (1992). The Differential Ability Scales and special education. *Journal of Psychoeducational Assessment, 10,* 92-98.

Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist, 52,* 10, 1046-1050.

Chambers, J., Parrish, T., & Lieberman, J. (1999). *What Are We Spending on Special Education in the United States?* Center for Special Education Finance, Washington, D.C.

Detterman, D. K., & Thompson L.A. (1997). What is so special about special education? *American Psychologist, 52,* 10, 1082-1090.

Dumont , R. P., Cruse, C.L., Price, L., & Whelley, P. (1996). The relationship between the Differential Ability Scales (DAS) and the Wechsler Intelligence Scales for Children-Third Edition (WISC-III) for students with learning disabilities. *Psychology in the Schools, 33,* 203-209.

Elliott, C.D. (1990a). *Differential Ability Scales: Administration and scoring manual.* San Antonio, TX: Psychological Corporation.

Elliott, C.D. (1990b). *Differential Ability Scales: Handbook.* San Antonio, TX: Psychological Corporation.

Elliott, C.D. (1990c).The nature and structure of children's intelligence: Evidence from the Differential Ability Scales. *Journal of Psychoeducational Assessment, 8,* 376-390.

Elliott, C. D., Murray, D.J., & Pearson, L.S. (1979). British Ability Scales. Windsor England, National Foundation for Educational Research.

Elliott, S. N. (1990). The nature and structure of the DAS: Questioning the test's organizing model and use. *Journal of Psychoeducational Assessment, 8,* 406-411.

Federal Register (1997). Part V, 300.7, No.204, Vol. 62, p. 55069.

Glutting, J.J., McDermott, P.A., Konold, T.R., Snelbaker, A.J., & Watkins, M.W. (1998). More ups and downs of subtest analysis: Criterion validity of the DAS with an unselected cohort. *School Psychology Review, 27,* 4, 599-612.

Glutting, J. J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. L. (1997). Incremental efficacy of WISC III factor scores in predicting achievement: What do they tell us? *Psychological Assessment, 9,* 295-301.

Groth-Marnat, G. (1996). Handbook of Psychological Assessment (3rd ed.). New York: John Wiley & Sons.

Holland, A.M., & McDermott, P.A. (1996). Discovering core profile types in the school-age standardization sample of the Differential Ability Scales. *Journal of Psychoeducational Assessment, 14,* 131-146.

Horn, J. L. (1998). A basis for research on age differences in cognitive abilities. In J.J. McArdle & R.W. Woodcock (Eds.), Human cognitive abilities in theory and practice (pp.57-87). Mahwah, NJ: Erlbaum.

Hutton, J.B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. *School Psychology Review, 21,* 2.

Kamphaus, R.W. (1993). Clinical assessment of children's intelligence. Boston: Allyn & Bacon.

Kaufman, A. S. (1994). Intelligent testing with the WISC-III. New York: Wiley.

Kaufman, A.S., & Kaufman, N.L. (1983). *Kaufman Assessment Battery for Children. Manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A.S., & Kaufman, N.L. (1993). *Kaufman Adolescent and Adult Intelligence. Manual.* Circle Pines, MN: American Guidance Service.

Keith, T.Z. (1990). Confirmatory and hierarchical confirmatory analysis of the Differential Ability Scales. *Journal of Psychoeducational Assessment, 8,* 391-405.

Keith, T.Z., & Witta, E.L. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC III: What does it measure? *School Psychology Quarterly, 1, 2,* 89-107.

Kercher, A.C., & Sandoval, J. (1991). Reading disability and the Differential Ability Scales. *Journal of School Psychology, 29,* 293-307.

Lyman, H.B. (1991). Test Scores & What They Mean (5th ed.). Englewood Cliffs, N.J., Prentice Hall, Inc.

Lyon, G.R. (1996). Special education for students with learning disabilities. *The Future of Children, 6, 1*, 121-134.

Lyon, M.A. (1995). A comparison between WISC III and WISC-R scores for learning disabilities re-evaluations. *Journal of Learning Disabilities, 28, 4*, 253-257.

Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman's "Intelligent Testing" approach to the WISC III. *School Psychology Quarterly, 12,* 3, 197-234.

McDermott, P.A. (1995). Sex, race, class and other demographics as explanations for children's ability and adjustment: A national appraisal. *Journal of School Psychology, 33,* (1), 75-91.

McDermott, P.A., & Glutting, J.J. (1997). Informing stylistic learning behavior, disposition, and achievement through ability subtests – or, more illusions of meaning? *School Psychology Review, 26,* (2), 163-175.

McDermott, P.A., Fantuzzo, J.W., Glutting, J.J., Watkins, M.W., & Baggaley, A.R. (1992). Illusions of meaning in the ipsative assessment of children's ability. *Journal of Special Education, 25,* 504-526.

McIntosh, D.E., & Gridley, B.E. (1993). Differential Ability Scales: Profiles of learning disabled subtypes. *Psychology in the Schools, 30,* 11-24.

National Center for Education Statistics (1995). *Statistics of State School Systems: Revenues and Expenditures for Public Elementary and Secondary Education.* U. S. Dept. of Education, Washington, D.C.

National Center for Education Statistics (1999). *Statistics of State School Systems: Revenues and Expenditures for Public Elementary and Secondary Education.* U. S. Dept. of Education, Washington, D.C.

National Information Center for Children and Youth with Disabilities NICHCHY; 2000). *Learning Disabilities,* Fact Sheet #7, 2000. Washington, D.C.

National Institute of Mental Health (1993). *Learning Disabilities*. NIH Publication No. 93-3611. Bethesda, MD.

Nichols, K., & Ward, S. (1998). A Comparison of DAS and WISC-III Scores for Children with Disabilities. Paper presented at the National Association of School Psychologists Annual National Convention, Orlando, FL.

Platt, L.O., Kamphaus, R.W., Keltgen, J., & Gilliland, F. (1991). An overview of the Differential Ability Scales: Initial and current research findings. *Journal of School Psychology, 29*, 271-277.

Ross-Reynolds, J. (1995). Best practices in conducting reevaluations. In A. Thomas & J. Brimes (Eds.), Best Practices in School Psychology-III. Washington, DC: National Association of School Psychologists.

Roush, W. (1995). Arguing over why Johnny can't read. *Science, 267*, 1896-1898.

Salvia, J., & Ysseldyke, J.E. (1995). Assessment (6th ed.). Boston, MA: Houghton Mifflin.

Sattler, J. (2001). Assessment of Children: Cognitive Applications (4th ed.). San Diego, CA: Jerome Sattler, publisher.

Schultz, M. K. (1997). WISC III and WJ-R tests of achievement: Concurrent validity and learning disability identification. *Journal of Special Education, 31, 3*, 377-388.

Shapiro, S.K., Buckhalt, J.A., & Herod, L.A. (1995). Evaluation of learning disabled students with the Differential Ability Scales (DAS). *Journal of School Psychology, 33* (3), 247-263.

Slate, J. R., & Jones, C. H. (1997). WISC-III IQ scores and special education diagnosis. *Journal of Psychology Interdisciplinary and Applied, 131, 1*, 199-202.

Smith, T., Smith, B., Matthews, N., & Kennedy, S. (1993). Subtest scatter and Kaufman regroupings on the WISC-R in non-learning disabled and learning disabled children. *Psychology in the Schools, 30*, 24-35.

Stanovich, K., & Siegel, L.S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86,* 24-53.

Stinnett, T.A., Havey, J.M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12,* 4.

Swanson, H.L. (1996). Classification and dynamic assessment of children with learning disabilities. *Focus on Exceptional Children, 28*, 9, 1-21.

Swanson, H.L., & Christie, L. (1994). Implicit notions about learning disabilities: Some directions for definitions. *Learning Disabilities Research and Practice, 9,* 244-254.

Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1986). Stanford-Binet Intelligence Scale: Fourth Edition. Chicago: Riverside Publishing.

Thurstone, L.L. (1938). Primary mental abilities. Chicago: University of Chicago Press.

Truscott, S.D., McDougal, J.L., & MacKinnon, K.L. (1996). WISC-R and WISC-III V-P Discrepancy and Subtest Stability over Time: Implications for Practice and Research. Paper presented at the National Association of School Psychologists

Annual National Convention, Atlanta, GA.

USDOE (1999). United States Department of Education: National Center for Learning Disabilities. Washington, D.C.

USDOE (1994). Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. Federal Register, 42, 65082-65085.

Wallace, G., Larsen, S.C., & Elksnin, L.K. (1992). Educational assessment of learning problems. Testing for teaching. Boston: Allyn & Bacon.

Waterman, B.B. (1994). Assessing children for the presence of a disability. *National Information Center for Children and Youth with Disabilities (NICHCHY) News Digest, 4,* 1. Washington, D.C.

Watkins, M.W. & Kush, J.C. (1994). Wechsler subtest analysis: The right way, the wrong way, or no way? *School Psychology Review, 23,* 640-651.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children (3rd ed.). Manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1939). The measurement of adult intelligence. Baltimore: Williams & Wilkins.

Youngstrom, E.A., Kogos, J.L., & Glutting, J.J. (1999). Incremental efficacy of Differential Ability Scales factor scores in predicting individual achievement criteria. *School Psychology Quarterly, 14,* (1), 26-39.