CONCURRENT VALIDITY OF THE DIFFERENTIAL ABILITY SCALES AND THE WOODCOCK-JOHNSON-III TESTS OF COGNITIVE ABILITIES FOR YOUNG CHILDREN

by

Jennifer Anne Salava

A Research Paper

Submitted in Partial Fulfillment of the
Requirements for the
Masters of Science in Education Degree
With a Major in

School Psychology

Approved: 2 Semester Credits

____
Dr. Mary E. Tusing, Investigation Advisor

The Graduate School
University of Wisconsin-Stout
August, 2001

The Graduate School
University of Wisconsin-Stout
Menomonie, WI 54751

ABSTRACT

	Salava	Jennifer	A.
(Writer)	(Last Name)	(First Name)	(Initial)

Concurrent Validity of the Differential Ability Scales and the Woodcock-Johnson-III
(Title)

Tests of Cognitive Abilities for Young Children

School Psychology	Dr. Mary E. Tusing	08/2001	67
(Graduate Major)	(Research Advisor)	(Mon./Yr.)	(# of Pages)

American Psychological Association (APA) Publication Manual (5^{th} ed.)
(Name of Style Manual Used in this Study)

Validation studies assess how comparable intelligence instruments are with regard to the cognitive abilities they assess. Results derived from such studies are used to make inferences as to how instruments are similar and different regarding the abilities that they are designed to measure. This paper is a review and critical analysis of literature related to the validity of two intellectual assessment instruments often used with young children, the Differential Ability Scales (DAS) and the Woodcock-Johnson-Third Edition Tests of Cognitive Abilities (WJ-III COG). Topics covered in the literature review include issues regarding the assessment of young children, the importance of test validity, and the theoretical foundations of each instrument, as well as their practical use with young children. The purpose of this paper is to propose a study examining the concurrent validity of the DAS and WJ-III COG when used with a sample of young children receiving special education services for speech and language delays.

Table of Contents

List of Tables

Chapter One

Introduction

Traditionally, the use of intelligence assessments has been a major part of the school psychologist's role in the educational community, as the Individuals with Disabilities Education Act (IDEA, 1990, 1997) mandates valid and reliable assessment practices for the identification of children with disabilities (Reschly & Grimes, 1995). The results derived from intelligence assessments are a critical part of educational planning for students with special needs (Dumont, Willis, Farr, McCarthy, & Price, 2000). The assessment process is of particular importance with young children, as the identification of children with or at-risk for developmental delay can be the first step to preventing later academic difficulties. Early intervention is highly desirable because difficulty in the preschool years is often a predecessor of developmental, learning, and behavior problems in subsequent years (Harrison, 1990; cited in McIntosh, Gibney, Quinn, & Kundert, 2000).

In response to research supporting the efficacy of early intervention services, the 1986 revision, as well as subsequent revisions (1990, 1997) of Special Education Law P.L. 99-457 provide an opportunity for special education services to be offered to infants, toddlers and preschool-age children (Demers & Fiorello, 2000). A particular section of IDEA, known as Part B, mandates special education services for children 3 to 5 years of age. Under this section, preschool-age children must be given the same opportunities for identification and special education intervention that is provided to children ages 5 to 21. Further, IDEA requires school districts to make reasonable effort to identify children age

5 and under who may be experiencing or be at-risk for developmental delays (Demers & Fiorello, 2000).

In most states, norm-referenced assessment is a required part of determining a child's eligibility for special education services, including eligibility for early childhood special education. Because of this, school psychologists have experienced an increased role in the assessment of young children. However, considerable debate exists regarding the most appropriate approaches in the assessment of young children (Nagle, 2000). Much of this controversy is centered on the utility of traditional norm-referenced assessment tools in evaluating and making diagnostic decisions for young children (Flanagan & Alfonso, 1995). In general, two opposing viewpoints exist regarding this issue. Practitioners either believe that the use of traditional intelligence tests with young children should be discontinued (Bagnato & Neisworth, 1994), or feel that intelligence test data should only be used as one component of the overall assessment process (Bracken, 1994).

Due in part to this controversy, as well as survey findings indicating that most school psychologists traditionally rely on traditional intelligence tests when assessing young children, Flanagan and Alfonso (1995) reviewed the technical characteristics of a number of intelligence tests for use with this population. They concluded that while assessment tools for young children are generally less technically adequate than those for school-age children, advances in test construction have resulted in an improvement in the quality of newer instruments. Two of the strongest tests identified by Flanagan and Alfonso (1995) for use with young children were the Differential Ability Scales (Elliott, 1990a) and the Woodcock-Johnson Revised Tests of Cognitive Ability (Woodcock and Johnson, 1989).

Recent discussions on the Early Childhood listserve published on the National Association of School Psychologists (NASP) website (www.nasponline.org) suggest that school psychologists continue to rely on norm-referenced assessment tools, including intelligence tests, in the assessment of young children. Further, since Flanagan and Alfonso's review, the Differential Ability Scales appears to have gained increased popularity among practitioners working with young children (NASP, 2001). In addition, the WJ-R has been revised and republished as the Woodcock-Johnson Third Edition Tests of Cognitive Ability (WJ-III; Woodcock, McGrew, & Mather, 2001). Improvements in the Third Edition's test construction and assessment technology suggest that it may be an even more adequate tool for use with young children that the WJ-R. For one, the WJ-III provides additional cluster scores that were not components of the WJ-R. Additionally, the WJ-III provides a Brief Intellectual Ability (BIA) score, which may have efficacy as a screening tool for young children (McGrew & Woodcock, 2001).

Statement of the Problem

Given that school psychologists continue to use intelligence tests in the assessment of young children, and that the tools utilized should be technically adequate for the purpose in which they are used, continued analysis is necessary when new tests are developed. Although the WJ-R was rated favorably by Flanagan and Alfonso (1995), the WJ-III has yet to be evaluated along similar guidelines. Further, outside of the technical manual, little research has examined the relationship between the WJ-III and other commonly used intellectual assessment tools and no research has examined the tool's use with special populations of young children. In particular, no analysis regarding the utility of the tests with young children identified as having speech and language delays has been conducted. Because of the high linguistic demand of many traditional assessment tools, further investigation related to this population of children is necessary.

Additionally, there is a lack of published research available related to specific scores of the DAS and WJ-III when used with young children. For one, there is a need to examine the relationship between the new cluster scores of the WJ-III and other measures used with young children. It is also necessary to examine the usefulness of the WJ-III BIA score as a screening tool for young children. Finally, other than that provided in the DAS technical manual, little data exists to aid in the interpretation of the DAS diagnostic subtests when used with young children. Based upon these factors, further evaluation of the DAS and WJ-III for use with young children is strongly needed.

Purpose of the Study

The purpose of the proposed study is to examine the concurrent validity of two intellectual assessment instruments designed for use with young children, the Woodcock Johnson-Third Edition Tests of Cognitive Abilities (WJ-III) and the Differential Ability Scales (DAS). The study will examine the use of these assessment tools with a special population of young children, those receiving early childhood special education services for speech and language delays. Correlations between broad scores, cluster scores, and subtests measured by each battery will be examined.

Research Questions.

1. The first question addressed in this study will be the relationship between the broad scores of the DAS and the WJ-III. The General

Intellectual Ability-Std (GIA-Std) and General Intellectual Ability-Ext (GIA-Ext) scores of the WJ-III will be compared with the General Conceptual Ability (GCA) score of the DAS. Mean overall composite scores between the two batteries will also be examined.

2. The second question addressed in the study will be the level of concurrent validity between the cluster scores of the WJ-III and the cluster scores of the DAS. Specifically, the strength of the relationship between measures of similar abilities, and the weakness of correlations between dissimilar measure of ability will be examined within each respective battery.

3. The study will also examine the relationship between the diagnostic subtests of the DAS and the cluster scores of the WJ-III. It is expected that the diagnostic subtests will have the strongest correlation with the WJ-III cluster scores that purport to measure similar ability constructs.

4. Finally, the study will examine the relationship between the WJ-III Brief Intellectual Ability (BIA) score and the DAS GCA score. A strong correlation is expected between these two measures, as they both are designed to provide a measure of overall intellectual ability.

Benefits of the Study

A benefit of investigating the validity of the WJ-III and DAS when used with young children is the acquisition of information that will aid practitioners in making appropriate assessment and diagnostic decisions. Examining the concurrent validity of two instruments designed to measure common abilities provides information about the amount of overlap and similarity of the constructs measured across instruments. Additionally, information about the expected differences between scores assessing common constructs may be discovered. This information is of particular importance because the DAS and WJ-III are tests used for diagnostic and classification purposes. Therefore, differences between scores on the two tests, which purport to measure the same abilities, could potentially result in different interpretations and classifications.

A second benefit of the study is that it will examine the use of the DAS and WJ-III with young children receiving special education services for speech and language delays; currently there is no published research in this area. The study will also provide in-depth analysis of how the DAS diagnostic subtests relate to other ability constructs, which will provide practitioners with additional interpretive information for both batteries. Finally, the study will examine the utility of the WJ-III BIA score as an intelligence screener for young children.

Assumptions

In order for the proposed study to adequately address the research questions, the following assumptions are made. First, it is assumed that all administrations of the DAS and WJ-III will follow standardized testing procedures and tests will be scored appropriately. Graduate students specifically trained in administering the DAS and WJ-

III to young children will conduct all assessments. It is also assumed that the instruments being investigated are valid measures of the cognitive abilities that they purport to measure. This will be supported by a review of available construct validity information for each test, which is reported in Chapter Two.

Limitations

One limitation of this study is that children participating will be from a narrow geographic region. Only children from western Wisconsin will be recruited to participate. Therefore, study findings may not generalize to children from other demographic backgrounds. Children currently receiving special education services will be targeted for inclusion in the study. Because of this, the results of the study may not generalize to children who have not been involved in special education. A final limitation of the study is that only children aged 3 years, 6 months to 5 years, 11 months will be included in the sample. Therefore, results of the study may not generalize to other age groups.

Definition of Terms

Concurrent Validity.

Concurrent validity is defined in this study as the comparison of scores obtained on two intelligence tests that are administered to subjects within approximately the same time frame. Examination of the broad and cluster scores of each test, as well as diagnostic subtests, will be conducted in order to establish the convergent and discriminant nature of the abilities measured in each battery.

Intelligence.

Historically, the term intelligence has been subject to many different definitions. Some professionals utilize a very global definition of the term, while others define intelligence in a much more narrow fashion (Sattler, 1992). For the purpose of this study, intelligence refers to those descriptions provided by the test authors of the abilities measured by the assessment tools investigated. These definitions will be further outlined as part of the test descriptions included in Chapter Two. Although it is recognized that certain aspects of intelligence are difficult to measure quantitatively, throughout this paper it is assumed that intelligence, as defined by test authors, can be measured by the standardized intelligence test referenced.

Intelligence Test / Instrument.

Intelligence Test / Instrument is defined in this study as an individually administered norm-referenced assessment instrument that assesses cognitive ability and provides indices of overall intellectual ability and specific cognitive ability domains.

Special Population.

Special population is defined in this study as a group of young children who have been identified by professionals as eligible for special education services due to one or more areas of developmental delays in the following domains: physical development, cognitive development, communication development, social or emotional development or adaptive development.

Young Children.

Young children are defined in this study as children approximately ages 3 to 6 years, consistent with Part B of IDEA (1997). Children in this age range are also often referred to as preschoolers.

Chapter Two

Review of Relevant Literature

The purpose of this chapter is to review literature relevant to the concurrent validity of the Woodcock-Johnson Psychoeducational Battery-Third Edition Tests of Cognitive Abilities (WJ-III; Woodcock, McGrew, & Mather, 2001) and the Differential Ability Scales (DAS; Elliott, 1990a) when used with young children. The chapter will discuss issues involved in the assessment of young children, issues of technical quality for tests used with young children, and history behind the use of the DAS and WJ-III with young children. Current research concerning the concurrent validity of the two instruments will also be reviewed and the theoretical underpinnings of the tests described.

Preschool Assessment

The assessment of young children has received much attention in recent years. Federal legislation, the national agenda of having all children ready to attend school, as well as research demonstrating the importance of early childhood experiences have all influenced this movement (Nagle, 2000). These developments have created a need for assessment services in many areas of early childhood education, thus expanding the role of the school psychologist (Nagle, 2000). In order to provide the best service possible, practitioners must be aware of unique issues pertaining to the assessment of young children.

Unique Characteristics of Young Children: Implications for Assessment.

Many of the characteristics inherent to young children can make assessment of children in this age range a challenge (Bracken & Walker, 1997). The preschool years are a time of tremendous change and growth. Physical, intellectual, and social development proceed at a rapid rate (Feldman, 1997). Issues such as differential learning experiences, developmental changes, newly emerging skills, and the often spontaneous behavior of children in this age range strongly influence the psychometric integrity of assessment procedures used with young children. Because many characteristics typical of young children make reliable and valid assessment challenging (Nagle, 2000) it is critical that individuals working with children in this age range are aware of their unique traits. For example, young children can exhibit a great degree of variability in exposure to environments outside the home, educational experiences, and cultural background. Due to a lack of formal schooling, young children may not have experienced the comparable and somewhat homogeneous experiences that school-age children have as a result of time spent in school (Romero, 1999).

Young children may also tend to have limited attention spans, high energy and activity levels, and a low tolerance for frustration. These characteristics are particularly problematic during lengthy assessment periods (Bracken & Walker, 1997). Dramatic changes in the above-mentioned abilities, particularly attention span, occur as children get older. Children at the upper end of the preschool range tend to have a much longer attention span and the ability to divide their attention between more than one stimulus at a time (Kali, 1997; cited in Feldman, 1997). Thus, it is also important that examiners take these factors into account when interpreting the results of an evaluation.

Another difficulty inherent in the assessment of young children is their level of motivation. Generally, children in this age range do not understand the importance of performing their best and are unable to link their performance to decision making based on assessment outcomes (Bracken & Walker, 1997). Young children also lack experience with a structured question-answer task-reward testing format (Rogers, 1982; cited in Romero, 1999). Unlike the majority of school-age children, for young children motivation and interest in testing is not as dependent upon the mastery of a task as it is upon the intrinsic appeal of the task itself (Romero, 1999). These concerns led Alfonso and Flanagan (1999) to recommend the importance of having attractive, colorful test materials in order to maximize a child's interest and participation in the testing situation.

Romero (1999) also points out that at times, it can be very difficult to distinguish a young child's inability to complete a task from simple refusal to cooperate. Similarly, it can be difficult to interpret nonverbal behaviors, such as a shrug of the shoulders or a silent nonresponse. Although there is no way to know with absolute certainty whether a child is not cooperating, too frightened to respond, or truly does not know an answer, the more experience an examiner has in working with young children will allow for the best interpretation of a child's behaviors. Experience in working with young children, as well as a thorough understanding of assessment instruments, allows an examiner to make appropriate inferences about a child's behaviors (Romero, 1999).

Another aspect of development that can potentially impact the cognitive assessment of young children is their language ability. Children with developmental delays may not possess or display well developed expressive or receptive language skills. This serves to limit their ability to respond to many traditional assessment tools, which rely heavily upon verbal responding and require strong receptive language skills to complete tasks (Bracken & Walker, 1997). Given the importance of understanding a child's language skills when assessing cognitive ability with norm referenced tests, an overview of language development is included below.

Development of Language Skills.

Children must master three areas of language in order to become proficient communicators: form, content and use. Form refers to the structural aspects of language, including the sound system (phonology), sentence structure (syntax) and word structure (morphology). The content area of language is the underlying meanings or concepts expressed through the use of vocabulary and word combinations. Language use, or pragmatics, refers to how speech is used to accomplish communicative acts, such as making requests, expressing desires, and providing information to others (Wyatt & Seymour, 1999). In general, all typically developing children follow a fairly predictable sequence of language development. However, there may be some degree of variability between children in when they reach certain language-related milestones (Hetherington & Parke, 1999; Wyatt & Seymour, 1999). Typical stages of speech and language development are outlined in Table 2.1.

Table 2.1

Milestones for Speech and Language Development

Age	Developmental Milestone
1 to 10 months	Crying serves as the earliest form of communication Babbling stage at 4 to 6 months, begin experimenting with sounds
10 to 12 months	Begin to use nonverbal gestures, such as pointing Emergence of first true words, initial vocabulary of 10 words Understand more words than produced
18 months to 2 years	Begin to produce two-word combinations Vocabulary of approximately 50 words, mainly nouns Frequently over-extend and under-extend the meaning of words Often produce simplified and modified versions of words
2 to 3 years	Begin using key grammatical forms (e.g. plurals, “ing”) Begin developing use of pronouns (e.g. “I,” “mine,” “me”) Begin to use first negatives (e.g. “no,” “not,” “can’t”) Begin to use early helping verbs (e.g. “do,” “be”) Begin asking “what,” “where,” yes/no questions
3 to 4 years	Begin to master grammatical rules of language Consistently producing the following sounds “n,” “m,” “p,” “b,” “h,” “k,” “g,” “f,” “w,” “y,” “t,” and “d” Able to produce words with more adult-like pronunciation
4 to 5 years	Begin to use more complex negatives, questions, pronouns, and helping verbs Average sentence length 4 to 5 words, more complex appearing Starting to master more complex sounds such as “s,” “z,” “l,” “r,” “sh,” and “ch,” however, errors still present Beginning to master conventional rules of speech Begin to adapt language to meet the needs of their listener

Note. Adapted from “Assessing the speech and language skills of preschool children.” (1999). T.A. Wyatt, & H. N. Seymour, In Nuttal, E.V., Romero, I., & Kalesnik, J. (Eds.) *Assessing and Screening Preschoolers: Psychological and Educational Dimensions* (pp. 218-239). Boston: Allyn and Bacon.

Children who experience difficulties in the development of speech and language may be eligible to receive special education services. Difficulties in communication development usually present in one of two ways, either a disorder of language or speech.

A language disorder is defined as the reduced ability to comprehend or express ideas (Early Childhood Special Education, 2001). A disorder of language usually results from inadequate knowledge of the linguistic rules that make up one's language system and/or an inability to express that knowledge. Language disorders are typically classified as disorders of form, content and/or function (Wyatt & Seymour, 1999). In contrast, a disorder of speech refers to difficulty with the mechanics of speech production. This may be observed in voice, articulation, fluency, or any combination of the three (Early Childhood Special Education, 2001).

Speech and language difficulties tend to be due to problems with central nervous system functioning (e.g. language learning disorder), perceptual deficits (e.g. hearing loss), cognitive-intellectual deficits (e.g. mental retardation) or social-emotional development (e.g. autism; Wyatt & Seymour, 1999). Additionally, young children may display language difficulties with no identifiable cause. These children, typically referred to as language delayed, often display limited vocabularies, use short, simple sentences, and make many grammatical errors. They may also have difficulty maintaining a conversation, understanding others, and making themselves understood. In addition to linguistic problems, language delayed children many also have difficulty classifying objects and recognizing similarities and differences (Dumtschin, 1988).

Under IDEA (1997), in order to qualify for special education services for speech and language delays children must consistently exhibit difficulty with the mechanics of speech production or inappropriate use in any of the structures of language. The child's speech and language functioning must be significantly below their cognitive ability, as measured by standardized assessment instruments, and the deficits must have an adverse effect on their educational performance. Additionally, the deficits may not be primarily caused by mental retardation, visual or auditory deficits, environmental or economic disadvantage, or cultural differences (IDEA, 1997).

Implications of Language Skills in the Assessment of Young Children

It has been argued by individuals in the field of psychoeducational assessment that many instruments designed to be used with young children have high linguistic demand, often to a degree that may impede test performance (Flanagan, Mascolo, & Genshaft, 2000). Alfonso and Flanagan (1999) ascertain that many intelligence tests designed for use with young children contain lengthy, complex subtest directions. Additionally, test directions often contain concepts that are unfamiliar to young children. For example, concepts that can be difficult for young children to understand (e.g. “over,” “all,” “before,” and “as many”) are frequently included in subtest directions. Due to an inability to properly understand what is required of them, children may be unable to perform a task optimally, which may result in an underestimate of their true ability (Flanagan et al., 2000).

Many tests of cognitive ability also rely heavily upon verbal response from the child. As young children are beginning to master the words and grammatical structure of their language, they are also beginning to master the sounds. Some sounds are acquired earlier in development than others are. It is not typically until the age of 7 that children have mastered the entire phonological system (Wyatt & Seymour, 1999). Because young children are in the process of developing proficient use of language, grammar and articulation errors are often present in their speech. Although most tests do not penalize a child’s response due to articulation problems, these errors may still interfere with a test examiner's understanding of a child's responses. This may also serve to provide an inaccurate picture of a child's true abilities.

In addition, many intelligence tests for young children contain a vocabulary subtest, which requires the child to provide verbal names for pictured objects. This measure of expressive language skill is dependent upon the sophistication of the child's vocabulary. Young children experience a tremendous amount of growth in their vocabulary between the ages of 3 and 6 years. At this age, children typically understand many more words than they are able to produce (Hetherington & Parke, 1999). Additionally, as a child's vocabulary develops, they often make errors in their use of words, such as under-extending and over-extending the words in their vocabulary (Wyatt & Seymour, 1999). As such, an expressive language subtest may paint an unfair picture of a young child's intellectual development if the child is speech or language delayed. A test less dependent upon expressive language may provide a more accurate picture of young children's abilities (Flanagan et al., 2000).

Because of the concerns related to assessing young children with limited speech and language capabilities, Alfonso and Flanagan (1999) concluded that quality assessment instruments for children in this age range must be sensitive to language development. They suggest that assessment tools should require one-word responses rather than multi-word responses, unless the test is intended to measure expressive language ability. Additionally, they ascertain that tests should not penalize a child for using gestures in order to communicate their responses. Linguistic demand is one of the many properties of a test that must be sensitive to the unique assessment needs of young children. Additional technical qualities to be considered when evaluating assessment tools to be used with young children are discussed below.

Technical Qualities of Assessment Tools for Young Children

Because of the concerns surrounding the utility of intelligence tests with young children, Flanagan and Alfonso (1995) reviewed the Wechsler Preschool and Primary Scales of Intelligence- Revised (WPPSI-R; Wechsler, 1989), the Differential Ability Scales, the Stanford-Binet Intelligence Scale: Fourth Edition (SB:IV; Thorndike, Hagen, & Sattler, 1986), the Woodcock-Johnson Psycho-Educational Battery-Revised: Tests of Cognitive Ability (WJ-R; Woodcock & Johnson, 1989), and the Bayley Scales of Infant Development-II (BDIS-II; Bayley, 1993), all of which are cognitive assessment instruments designed to be used with young children. Technical characteristics important for tests used with young children; namely, standardization characteristics, reliability, test floor, item gradients, and validity were evaluated. Ratings of each domain were based upon criteria suggested by Bracken (1987), and by standards for psychological testing set by the American Educational Research Association (AREA), American Psychological Association (APA), and National Council on Measurement in Education (NCME; Flanagan & Alfonso, 1995).

Standardization.

Psychological tests have no predetermined standards of passing or failing. Rather, an individual's score is interpreted by comparing it with scores obtained by others on the same test. Therefore, it is important that during the process of standardizing a test it is administered to a large, representative sample of the type of persons for whom the test was designed (Anastasi & Urbina, 1997). As such, tests designed to be used with young children should include an appropriate number of children of that age range in the standardization sample in order to provide a representative comparison group. Overall, Flanagan and Alfonso (1995) found all major intelligence test batteries to have acceptable standardization characteristics in terms of representativeness and sample size.

Reliability.

A test's reliability refers to the degree to which a child's score is consistent (internal consistency) and stable across time (test-retest reliability; Anastasi & Urbina, 1997). Given rapid growth and change in abilities during the preschool years, tests used with children in this age range have notoriously lacked adequate reliability (Bracken & Walker, 1997). Adequate internal consistency of subtests and total test scores allows one to assume that the items that make up the tests are related to one another and are measuring similar factors (Bracken & Walker, 1997). Bracken (1987) suggests a median subtest internal consistency criterion of .80. Because total test scores are more often used in making placement decisions, they should be more reliable than subtest scores. Therefore, Bracken (1987) suggests a .90 level of total test internal consistency. Flanagan and Alfonso (1995) found that total test scores of all instruments reviewed met or approached the desired reliability of $r = .90$.

Test-retest reliability provides information regarding the stability of test scores over time. The test-retest sample should closely approximate the population for which the test is intended to be used. Thus, tests that are designed to be used with young children should include an appropriate number of children in this age range in the test-retest sample (Bracken, 1987). Although stability coefficients of around .90 were found with all batteries, many of the tests reviewed by Flanagan and Alfonso (1995) had inadequate representation of young children in the test-retest sample. Therefore, although it appears that the test remains stable over time this information cannot be generalized to all young children (Flanagan & Alfonso, 1995).

Test Floor.

A major concern regarding assessment tools for young children is whether or not the test demonstrates adequate floor. Test floor refers to the extent to which there are enough easy items to discriminate between lower levels of test performance, and thus distinguish children at lower levels of cognitive functioning. Inadequate test floors may result in an overestimation of ability, as the instrument does not contain enough easy items to detect subtle differences across ability levels (Flanagan & Alfonso, 1995). It is recommended (Bracken, 1987) that subtest and total test floors extend at least two standard deviations below the mean, with an ideal goal of three standard deviations below the mean. In other words, on a subtest with a mean of ten and a standard deviation of three, a raw score of one must be associated with a standard score of three or less in order to ensure differentiation among children who function well below the normative mean. Although subtest floor is an important test consideration, the floor of the total test score is most important. This is due to the fact that placement decisions are made based largely upon a child's overall perceived level of functioning (Flanagan & Alfonso, 1995).

Test floor is generally one of the weakest characteristics of assessment instruments designed to be used with young children. In general, test floor tends to improve as the age with which the test is used increases (Flanagan & Alfonso, 1995). Flanagan and Alfonso (1995) found that the majority of subtests of the test batteries they reviewed did not have adequate floor at the lower end of the age range for which they were intended to be used. Overall, with a few exceptions, it was discovered that subtest, scale and total test floor of intelligence tests for very young children are poor.

Item Gradients.

A related characteristic to be considered when evaluating assessment tools for young children is the test's item gradients. Item gradient refers to the magnitude of change in a child's standard score resulting from small changes in raw score points. The better item gradients a test has, the more sensitive it is to minor differences in a child's true ability (Flanagan & Alfonso, 1995). Bracken (1987) suggests that changes in single raw score points should be less than or equal to a change of one-third of a standard score standard deviation. In other words, it should take a minimum of three raw score points to equal a standard score increase of one. When this is not the case item gradient "violations" occur.

Item gradient violations that occur closer to the mean of a test are most problematic. This is because "violations around the mean are not only insensitive to differences in ability within the average and low average range of functioning (within one standard deviation of the mean), but they also cannot detect differences in ability greater than one standard deviation below the mean" (Flanagan & Alfonso, 1995 p. 71). The majority of cognitive assessment instruments evaluated by Flanagan and Alfonso (1995) were found to have poor item gradients at the lower end of the age range for which they are intended to be used. Additionally, the violations began to decline and to occur further away from the mean with increasing age.

Validity

Validity, the idea that a test is measuring what it intends to, is a core concept of all assessment practice. As the present study focuses on the concurrent validity of two assessment instruments, the concept of validity will be discussed in depth in the following pages. Determining the validity of a test is a lengthy, complex process. Validation of an intelligence test generally requires gathering information about three specific types of validity: construct, content, and criterion-related. The construct validity of a test describes the extent to which the test measures a theoretical construct or trait. Confirming this aspect of test validity requires gathering information from a variety of sources. Construct validity is further supported by evidence from content and criterion-related validation studies (AERA, APA, NCMA, 1985; Anastasi & Urbina, 1997).

Concurrent validity, along with predictive validity, make up the broader construct of criterion-related validity. Criterion-related validity studies provide evidence that test scores are related to another set of performance, or outcome, criteria. The main difference between concurrent and predictive validity is the time the outcome criteria are obtained. Studies of predictive validity are aimed at determining the correlation of a measure with performance on another measure at some point in the future. Concurrent validity studies examine the correlation of performance on two measures obtained simultaneously (Kazdin, 1992). Thus, concurrent validity studies attempt to gain evidence supporting whether or not an instrument measures what it is intended to measure by comparing it with another instrument thought to measure a similar construct (Anastasi & Urbina, 1997).

In the case of intelligence tests, the intercorrelations of subtest and total test scores are typically examined to establish two features of concurrent validity, convergent and discriminant validity. Convergent validity involves determining the extent to which two measures assess similar or related constructs. It is expected that measures that assess similar domains will correlate strongly. Conversely, discriminant validity is the correlation between measures that are not expected to be related to each other. Weaker correlations are expected between measures that assess conceptually distinct elements. The overall validity of an instrument is suggested if the measures show high correlations between similar constructs and low correlations between dissimilar constructs (AERA, APA, NCMA, 1985; Kazdin, 1992; Anastasi & Urbina, 1997). Typically concurrent validity studies (c.f. McIntosh, Brown, & Ross, 1995; McIntosh, Wayland, Gridley, & Barnes, 1995; McIntosh, Gibney, Quinn, & Kundert, 2000) also compare the overall range of scores obtained on the two tests. Studying the mean score difference between two tests aids in determining if the tests have comparable performance outcomes for examinees (Anastasi & Urbina, 1997).

Comparing the scores of two intelligence tests aids in the determination that the two tests, which were designed to measure similar aspects of cognitive ability, have comparable outcomes. Studies that provide information about the similarities and differences between tests are important for several reasons. For one, they provide data about the amount of overlap and similarity of the constructs measured across instruments. They may also predict expected differences between scores assessing common constructs across instruments. This information is of particular importance with regard to tests that are used for diagnostic and classification purposes. Differences between scores on two different tests that purport to measure the same abilities could potentially result in different interpretations and subsequently potentially different eligibility determinations (Dumont et al., 2000).

In general, it is difficult to set acceptable validity criteria for assessment instruments. This is mainly due to the unclear nature of the criteria that specifies the conditions under which an instrument is considered valid (Bracken, 1987; Flanagan & Alfonso, 1995; Bracken & Walker, 1997). According to Bracken and Walker (1987), it is the test author's responsibility to provide validity evidence so examiners are able to evaluate a test in order to determine whether a particular instrument is suitable for the purpose and population they have intended. Based upon information provided in technical manuals and in the assessment literature, Flanagan and Alfonso (1995) concluded that most all intelligence tests for young children reportedly had adequate overall validity. However, few of the validity studies reported had been conducted with children under the age of 4 or 5 years. Therefore, generalization of these findings to young children is difficult.

After their evaluation of tests according to the technical characteristics reported above, Flanagan & Alfonso (1995) determined that the DAS and the WJ-R were two of the most appropriate intelligence tests for use with young children, yet continued study of the validity of both tests with young children was needed. However, since Flanagan and Alfonso's 1995 review, the WJ-R has undergone major revisions and has been published as the WJ-Third Edition. Therefore the DAS and the WJ-III were chosen for review for the proposed study. A discussion of the structure and theoretical underpinnings of both tests follows. Additionally, concurrent validity studies involving the two test batteries are also presented.

The Woodcock Johnson-III Tests of Cognitive Abilities

Theoretical Background.

The WJ-III is a revised and updated version of the WJ-R. The WJ-III like the WJ-R, was developed to match a specific factor analytic theory of intelligence, namely the Carroll-Horn-Cattell (CHC) theory of cognitive abilities. CHC theory represents an integration of two independently derived theories of intelligence, *Gf-Gc* theory (Horn & Cattell, 1967) and Carroll's Three Stratum theory of intelligence (Carroll, 1993). This combination of two very similar theories is felt to be the most comprehensive and empirically supported framework available for understanding human cognitive abilities (McGrew & Woodcock, 2001).

CHC theory is a hierarchical multiple ability theory. Thus, the WJ-III cluster scores follow a similar pattern. The WJ-III includes a higher order factor measure of *g*, the General Intellectual Ability-Standard (GIA-Std) and General Intellectual Ability-Extended (GIA-Ext) scores. Unlike some ability measures, the WJ-III strongly emphasizes the importance of *g* in the overall ability score. That is, individual subtests' contribution to the GIA score are weighted, with the subtests identified as the strongest measure of *g* contributing the most to the overall GIA score. Thus, and WJ-III GIA scores are both well saturated measures of *g.*

The WJ-III also provides CHC factor scores that represent the broad ability areas defined in CHC theory. These include Comprehension-Knowledge (*Gc*), Long-Term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv*), Auditory Processing (*Ga*), Fluid

Reasoning (*Gf*), Processing Speed (*Gs*), and Short-Term Memory (*Gsm*). Additionally, each subtest contributing to a broad ability cluster is designed to be individually interpreted as a unique and distinct narrow ability measure of the *Gf-Gc* ability it purports to assess (McGrew & Woodcock, 2001).

Validity.

The WJ-III was built upon the extensive exploratory and confirmatory factor analyses of the Woodcock Johnson Psychoeducational Battery (WJ; Woodcock & Johnson, 1977) and WJ-R normative data, which supported the *Gf-Gc* broad factor structure of the test battery. Confirmatory factor analyses (CFA) of the WJ-III provided support for the test at both the broad and narrow factor levels (McGrew & Woodcock, 2001). In general, the WJ-III is accepted as a well validated measure of intelligence. However, examination of the technical manual reveals that a majority of the studies related to the internal structure of the WJ-III focused on individuals ages six to adult. Because no younger children were included in this research it is difficult to generalize factor analytic findings supporting test structure to children under the age of six.

However, the structure of the WJ-III was also confirmed by examining developmental patterns with the use of growth curves. According to McGrew and Woodcock (2001), this is a relatively new concept in test development. It allows for the examination of the relationship between abilities and different age levels. The existence of different growth patterns across abilities measured can provide additional evidence that the abilities measured by the test are unique and distinct from one another (Carroll, 1993). Test authors examined the changes in score patterns beginning at age five years, continuing through age 90. It was discovered that all tests and clusters displayed average score changes consistent with the developmental growth and decline of distinct CHC cognitive abilities across the life span (McGrew & Woodcock, 2001). Thus, this information provides addition overall support for the seven CHC factor scores of the WJ-III and extends the findings to a younger age range. Nonetheless the findings cannot be generalized to children younger than age five.

Correlations of the WJ-III COG with other Cognitive Measures

As with the DAS, a number of studies reported in the technical manual of the WJ-III were conducted to determine the similarities and differences between the WJ-III COG and other measures of cognitive ability. Two studies addressed cognitive measures appropriate for young children. In one study the WJ-III COG, WPPSI-R and the DAS were administered to a group of children ages 1 year, 9 month to 6 years, 3 months. A similar study was conducted administering the WJ-III and SB-IV to a sample of children ages 3 years to 5 years, 10 months. The GIA-Std and GIA-Ext correlated positively with composite scores of other intelligence batteries, with correlations of *r* = .67 or higher. According to McGrew and Woodcock (2001), correlations of this magnitude are similar to those reported between the composite scores of other intelligence test batteries, as well as those reported for the WJ-R. Overall results of the studies are summarized in Table 2.2.

Table 2.2

Correlations of the WJ-III with other Measures of Cognitive Ability

WJ-III	WPPSI-R FSIQ	SB-IV Composite SAS	DAS GCA
GIA-Standard	.73	.76	.67
GIA-Extended	.74	.71	.76

Note. Adapted from *Woodcock-Johnson III Technical Manual.* (2001). K.S. McGrew & R.W. Woodcock. Itasca, IL: Riverside Publishing. FSIQ = Full Scale IQ; SAS = Standard Age Score; GCA= General Conceptual Ability.

Patterns of correlation of the cognitive clusters of the WJ-III provide evidence of convergent and discriminate validity, supporting the notion that the test measures several distinct abilities. The WJ-III Verbal Ability-Std, Verbal Ability-Ext, Comprehension-Knowledge (*Gc*), and Knowledge (*Gc*) scores all have positive correlations with verbal cluster scores of the WPPSI-R, DAS, and SB-IV, generally $r = .60$ or higher. According to McGrew and Woodcock (2001) these strong correlations provide support for interpretation of the WJ-III Verbal Ability, Comprehension-Knowledge, Verbal Comprehension and Knowledge measures as valid measures of verbal knowledge and comprehension abilities (*Gc*).

Because the Thinking Ability-Std and Thinking Ability-Ext scores are comprised of multiple broad abilities (*Glr, Gv, Ga, Gf*) the test authors expected that these composites would function similarly to other test scores based on multiple abilities, namely full-scale scores. As expected, the Thinking Ability clusters consistently had the highest correlations with the DAS GCA score, the SB-IV Test Composite SAS, and the WPPSI-R Full Scale IQ ($r = .68/.63$, .73/.69, .68/.64, respectively; McGrew & Woodcock, 2001).

Interpreting the convergent and discriminant validity for Visual-Spatial Thinking *(Gv)* and Fluid Reasoning (*Gf)* cluster scores was restricted by the fact that most other test batteries either do not provide specific measures of these abilities or they provide composites that contain combinations of these broad abilities (e.g., the WPPSI-R PIQ is thought to be a mixed measure of broad *Gf, Gv,* and *Gsm* abilities; McGrew & Woodcock, 2001). Findings reported in the technical manual suggest that these cluster scores were consistently correlated with the measures that were theoretically considered to be most similar within other cognitive batteries. For example, the Visual-Spatial Thinking (*Gv*) score of the WJ-III was found to correlate most strongly with the WPPSI-R Performance IQ (r =.53). Similarly, the Fluid Reasoning (*Gf*) score was found to have its highest correlation (r = .40) with the DAS Nonverbal Ability score. These moderate correlations suggest that the WJ III may provide a more specific measure of these domains, while other instruments appear to be assessing a more mixed measure of the abilities.

An absence of specific cluster scores for Long-Term Retrieval (*Glr*), Auditory Processing *(Ga*), Short-Term Memory (*Gsm*), Processing Speed (*Gs*), and Phonemic Awareness (*Ga*) across other intelligence tests hampers the interpretation of the concurrent validity of these WJ-III clusters. Similarly, lower correlations were obtained between the Cognitive Efficiency-Std and Cognitive Efficiency-Ext scores of the WJ-III with other batteries. This reinforces the notion that the WJ-III measures unique abilities typically not measured by other intelligence tests (McGrew & Woodcock, 2001).

The Differential Ability Scales

Theoretical Background.

Based upon its European predecessor, the British Ability Scales (BAS; Elliott, Murray, & Peterson, 1979), the DAS was designed to assess specific abilities in addition to the general idea of "intelligence" (Elliott, 1997). According to Elliott (1990b), numerous controversies in the area of human ability have precluded any single theory or model of intelligence to find universal acceptance. Because of this, he argues that it is a mistake to base a cognitive test battery on any single theory. Thus, "the DAS is built upon a variety of subtests that sample a range of abilities felt to be useful in assessing children, particularly those with learning difficulties" (Elliott, 1997, p.184).

In a general sense, Elliott adopted Thurstone's view that intelligence is multidimensional and consists of several "primary mental abilities." The influence of this theory can be seen in the development of individual subtests, particularly the diagnostic subtests, that provide narrow, distinct indicators of multiple cognitive abilities (Elliott, 1990b). Similarities between the cluster scores of the DAS with the various broad and narrow abilities of Horn and Cattell's *Gc-Gf* theory (1967) have also been noted in research (Elliott, 1990b; McGrew & Flanagan, 1998). As such, the DAS is an instrument that can be interpreted from a variety of theoretical perspectives including CHC theory (Elliott, 1990b; 1997). In general, the DAS follows a hierarchical structure of abilities. There is a higher order factor measure of *g* (GCA score), group factors that represent broad ability areas (cluster scores) and specific abilities (individual subtest scores; Elliott, 1990b).

Elliott also points out that the GCA score provided by the DAS is defined differently when compared to overall ability scores of other cognitive measures, such as the Wechsler and Stanford-Binet scales, which adopt a relatively broad definition of intelligence. This is evident in the fact that the composite scores of many other test batteries give equal weight to all subtest scores, including those that have low *g* loadings. In contrast, the GCA, like the WJ-III COG GIA score, is considered a well-saturated measure of *g* because only the subtests that have the highest *g* loadings, referred to as core subtests, contribute to its estimation. Additionally, the diagnostic subtests of the DAS are completely independent of the GCA score and only provide measures of the distinct abilities they were designed to measure (Elliott, 1997). As the DAS GCA score and the WJ-III GIA scores are both well saturated measures of *g* it is expected that they would be strongly correlated.

Validity.

Exploratory and confirmatory factor analyses conducted by Elliott (1990b) suggest that a one-factor (general ability) solution provides the best fit of data for the core subtests for the Lower Preschool Level. In contrast, a two-factor (Verbal-Nonverbal) solution provides the best fit of data for the Upper Preschool Level of the DAS. This progressive differentiation of ability with increasing age is consistent with research from a number of sources (c.f. Anastasi, 1970 for a review; cited in Elliott et al., 1991). Further factor analyses conducted by Keith (1990) supported the underlying factor structure of the DAS and determined that the constructs measured by the DAS are consistent across overlapping age levels of the test.

Correlations of the DAS with Other Cognitive Measures.

A number of studies reported in the DAS technical manual were conducted to determine the similarities and differences between the DAS and other measures of cognitive ability. The majority of these studies occurred during the standardization of the DAS, and thus did not target special populations of children. However, children receiving special education services were represented in the normative sample of the DAS. The goal of these studies was to determine if the DAS appears to be measuring constructs similar to those measured by other cognitive batteries (Elliott, 1990b).

Three studies reported in the technical manual included samples of young children. One study examined the relationship between the WPPSI-R and DAS for a sample of children ages 4 years, 6 months to 5 years, 11 months. A similar study examined the relationship between the SB:IV and DAS for children ages 4 years to 5 years, 11 months. Finally, the DAS was compared to the McCarthy Scales of Children's Abilities (MSCA; McCarthy, 1972) for a group of children aged 3 years, 4 months to 3 years, 7 months. Results of these studies are summarized in Table 2.3.

Table 2.3

Correlations of the DAS with other Measures of Cognitive Ability

DAS	WPPSI-R FSIQ	SB: IV Composite	MSCA GCI
Verbal Ability	.77	.74	.84
Nonverbal Ability	.72	.69	.55
GCA (6 subtests)*	.89	.77	.82
GCA (4 subtests)*			.76
Special Nonverbal			.34

Note. Adapted from *Differential Ability Scales Introductory and technical manual.* (1990) By C.D. Elliott. San Antonio, TX: Psychological Corporation. FSIQ = Full Scale IQ; GCI = General Cognitive Index.
*In the MSCA study data were analyzed two ways; first for all children ages 3:4-3:7 based on the 4 subtest GCA, then for children ages 3:6-3:7 based on the 6 subtest GCA

Overall, the DAS GCA score consistently had strong correlations with the overall composite scores of other cognitive batteries, with $r = .76$ or higher. The GCA score appears to be most similar to the Full Scale score of the WPPSI-R with a correlation of $r = .89$. Thus, although the GCA score includes only the subtests that are strong measures of *g*, whereas the composite scores of most other batteries contain more subtests with a wider range of *g* loadings, it appears as though the various test batteries are measuring a similar construct (Elliott, 1990b). Therefore, the DAS can be regarded as a valid measure of overall cognitive ability. The DAS correlations were stronger than those found between the WJ-III and other cognitive measures. This indicates that although the various batteries are measuring a similar overall construct, the WJ-III may be measuring unique aspects of intellectual functioning relative to other batteries.

The DAS Verbal Ability score consistently had stronger correlations with measures of verbal ability than any other measure across test batteries. The correlations with the WPPSI-R Verbal IQ score and SB:IV Verbal Reasoning score were quite similar, $r = .74$ and .72, respectively. The DAS Verbal Ability score correlated most strongly with the MSCA Verbal score, with $r = .79$. Overall, it appears that the DAS Verbal Ability score is assessing a construct that requires a high degree of verbal skill (Elliott, 1990b) as similarly measured by other intelligence tests.

The Nonverbal Ability cluster score of the DAS followed a similar pattern of relationships. It correlated most strongly with measures of nonverbal reasoning and conceptual ability composites from other test batteries. The DAS Nonverbal Ability score was most strongly related to the WPPSI-R Performance IQ score, $r = .75$. A slightly lower correlation was found with the SB:IV Abstract-Visual Reasoning score, at $r = .64$. Overall, the Nonverbal Ability score had weaker correlations than the DAS Verbal Ability score with all MSCA clusters. Even so, it was related most strongly to the Perceptual-Performance composite score, $r = .66$ (Elliott, 1990b).

The core subtests of the DAS showed patterns of convergent and discriminant validity similar to the test's cluster scores, as would be expected given that the cluster scores are comprised only of core subtests. In contrast, the diagnostic subtests of the DAS typically appeared to be fairly independent of the composite scores of other batteries. Three of the four diagnostic subtests correlated below $r = .40$ with the WPPSI-R Full Scale IQ score. The same was true for correlations with the SB:IV Composite score (Elliott, 1990). In both cases the subtest that had higher correlations with overall composite scores was Matching Letter-Like Forms. According to Elliott (1990b), this suggests that this subtest is related to a range of higher level abilities in addition to perceptual matching, which it is intended to measure.

Overall, the DAS appears to have much in common with other intelligence instruments. The GCA score has been found to correlate strongly with composite scores of other test batteries. Additionally, the core subtests and cluster scores of the DAS have shown expected patterns of convergent and discriminant validity over a range of studies. However, because the correlation studies reported in the manual were conducted during the standardization of the DAS there is a lack of information available about the similarities and differences of the DAS with newer tests used with young children, such as the WJ-III and The Leiter International Performance Scale-Revised (Leiter-R; Roid & Miller, 1997). Further research is needed to extend these findings to newer assessment tools.

DAS and Preschool Screening Tools.

A number of independent studies have been conducted to determine the relationship between the DAS and preschool screening instruments, as sound screening instruments should demonstrate reasonable high correlations with intelligence measures (McIntosh et al., 2000). McIntosh, Wayland, Gridley, and Barnes (1995) examined the relationship between the DAS and the Bracken Basic Concept Scale (BBCS; Bracken, 1984) for a sample of typically developing preschool-age children. The BBCS provides a measure of receptive language, basic concepts, and school readiness for young children. A strong relationship was found between the DAS GCA score and the BBCS Total Test Score, suggesting that the two measures assess similar skills. The correlations between the DAS core subtests and the BBCS subtests were in the moderate range, an indication that the subtests share some common characteristics, while still assessing unique abilities. As the BBCS has similar correlations with both the DAS Verbal Ability and Nonverbal Ability cluster scores it appears as though it measures both verbal and nonverbal-related abilities (McIntosh, Wayland, et al., 1995).

McIntosh, Brown, and Ross (1995) conducted a similar study of the DAS and BBCS with an at-risk sample of young children. At risk was defined as those children lacking school readiness skills because of learning difficulties that stemmed from being economically disadvantaged or having physical or emotional disabilities. Again, a strong correlation between the DAS GCA score and the BBCS Total Test Score existed. However, the BBCS Total Test score correlated only moderately with both the Verbal and Nonverbal Cluster scores of the DAS, which was slightly lower than that obtained for the sample of typically developing children (McIntosh, Wayland, et al., 1995). The authors suggested that this probably reflected a restriction of range in the test scores obtained, as only at-risk children and a smaller sample size were involved in this study (McIntosh, Brown, et al., 1995).

McIntosh and colleagues (2000) examined the relationship between the DAS and Early Screening Profiles (ESP; Harrison, 1990) for a sample of at-risk young children ages 3 years, 6 months to 5 years, 11 months. Children were identified as at-risk because of their enrollment at an at-risk preschool or Head Start program. A strong positive correlation was obtained between the ESP Total Screening score and the DAS GCA score, which suggests that these two measures assess similar constructs. The Verbal Ability score correlated most strongly with the Cognitive/Language Profile (r = .69) while displaying a weaker correlation (*r* = .38) with the Motor profile. Similarly, the DAS Nonverbal Ability score had its strongest correlation with the Motor Profile (*r* = .63) and its weakest correlation (*r* = .46) with the Language Profile. Overall, the results of this study indicate that while the ESP screens areas of development not typically assessed on an intelligence scale (e.g. Daily Living Skills and Socialization), the two instruments are assessing similar constructs at the general score level. Based upon their findings, the authors suggest that the ESP Total Screening score could be a relatively good predictor of a child's DAS GCA score (McIntosh et al., 2000).

In general, it has been found that the DAS is positively related to commonly used preschool screening instruments for samples of typically developing children and children at-risk for learning difficulties. Based upon these findings it appears that the DAS Verbal Ability score relates most strongly to other language based measures, indicating that this domain assess skills in the areas of language and acquired knowledge. In contrast, the DAS Nonverbal Ability scores relates most strongly to nonverbal measures, such as motor ability, supporting the notion that it provides assessment of skills less tied to language ability. Given these findings, the authors of each study indicated that further investigation of the concurrent validity of the DAS and preschool screening tools is warranted.

Critical Analysis

A review of relevant literature has indicated that there is much controversy related to the assessment of young children. Questions of the validity of standardized measures of intelligence have caused concern over how accurate cognitive assessment is with young children (Bagnato & Neisworth, 1992; Flanagan & Alfonso, 1995; Bracken & Walker, 1997).

Despite the controversy surrounding this issue, norm-referenced assessments are often a required component of determining eligibility for special education services for young children. Thus, standardized intelligence tests continue to be used regularly with this population (Flanagan & Alfonso, 1995; NASP listserv, 2001). Due to concern regarding to the utility of these instruments with young children it is crucial that professionals are provided with information about the strengths and weaknesses of various instruments in order to make appropriate assessment and evaluation decisions. Continued evaluation of assessment tools for use with young children is needed to provide practitioners with information to utilize in choosing appropriate assessment tools for practice.

In Flanagan and Alfonso's (1995) review of intelligence tests for use with young children the two instruments that emerged as the most technically sound were the DAS and WJ-R. Thus, it follows that continued research should focus on these two instruments. However, recent publication of the revised version of the WJ-R, the WJ-III calls for additional analysis of this test's utility with young children. In addition, little in-depth analysis of the relationship between the DAS diagnostic subtests and cluster scores of other batteries has been conducted. Given the greater breadth of young children's abilities discussed by the WJ-III it follows that a better understanding of the DAS diagnostic subtests could be determined by relating them to cluster scores on the WJ-III. In addition, the WJ-III now offers an intellectual screener in the BIA score. However, there is little information available regarding the use of the WJ-III BIA score as a screener for young children

Finally, an important facet of this concern is related to the linguistic demand of traditional intelligence tests. Many tests designed to be used with young children require a great deal of verbal responding (expressive language skills) as well as an understanding of many relational concepts (receptive language skills) that may be challenging for young children, especially those with speech or language delays (Flanagan et al., 2000). In order to better understand the linguistic demand of traditional intelligence tests, further research in this area is needed. Of particular importance is an examination of the use of assessment tools with children who have been identified as having speech and language delays.

Chapter Three

Methodology

The purpose of this chapter is to propose a research study based upon key issues highlighted in the preceding review of the literature. Specific research questions to be addressed in the proposed study will be defined. In addition, information about the participants, procedures, instrumentation, and data analyses to be conducted in the proposed study will be outlined. Finally, statements about expected findings, which are based on previous research, will be included.

The previous review of the literature provides support for the need to examine the concurrent validity of the Woodcock-Johnson-Third Edition Tests of Cognitive Abilities (WJ-III; Woodcock, McGrew, & Mather, 2001), and the Differential Ability Scales (DAS; Elliott, 1990a) for a sample of young children receiving special education services in the area of speech and language. The relationship between the broad and cluster scores of the two batteries will be examined in order to determine the level of concurrent validity between each battery. The following specific questions will be examined:

1. What is the relationship between the General Intellectual Ability-Standard (GIA-Std) and General Intellectual Ability-Extended (GIA-Ext) scores of the WJ-III and the General Conceptual Ability (GCA) score of the DAS for a special population of young children? How comparable are the mean scores obtained on each battery for a special population?
2. What is the relationship between the Comprehension-Knowledge (*Gc*), Fluid Reasoning (*Gf*), Visual-Spatial Thinking (*Gv*), Processing Speed (*Gs*), Auditory Processing (*Ga*), Long-Term Retrieval (*Glr*), and Short-

Term Memory (*Gsm*) cluster scores and the Verbal Ability, Thinking Ability, and Cognitive Efficiency broad scores of the WJ-III and the Verbal Ability and Nonverbal Ability cluster scores of the DAS for a special population of young children?

3. What is the relationship between the CHC cluster scores of the WJ-III and the diagnostic subtests of the DAS for a special population of young children?

4. What is the relationship of the WJ-III Brief Intellectual Ability (BIA) score and the DAS GCA score for a special population of young children?

Participants

Children aged 3 years, 6 months (3:6) to 5 years, 11 (5:11) months will be targeted for participation in the study. This age range has been chosen because both the WJ-III and DAS have been found to have better subtest floor for children in this age range, and therefore provide more accurate measures of ability. The children targeted for inclusion in this study will be receiving special education services for speech and language delays as defined by the eligibility criteria for the state of Wisconsin. The sample will consist of approximately 30 children. An equal number of males and females will be targeted to participate. Additionally, the sample will be distributed equally across the age range of 3:6 to 5:11. An effort to include equal number of males and females at each three-month interval will be made.

Procedures

School districts in western Wisconsin will be contacted for permission to solicit subjects from early childhood special education programs within each district.

Permission to recruit subjects will also be obtained from teachers of each program approached. Once permission has been granted, a brief description of the study and a letter of permission will be sent to the parents/guardians of children who are eligible to participate. Parents interested in having their child participate will be contacted regarding the specific procedures involved in the study. They will be required to sign statements of informed consent indicating that they understand the procedures of the study and their child's participation.

After parent permission has been obtained, each child will be administered the DAS and WJ-III. Standardized testing procedures will be followed for each administration. Test administrators will be graduate students in school psychology who have been specifically trained in administering both the DAS and WJ-III to young children. Children will be tested in private rooms, either at the University of Wisconsin-Stout or within their school building. Test batteries will be administered in counterbalanced order to avoid practice effects. Each child will be assigned a code number when tested in order to maintain confidentiality. Only the researcher and faculty advisor will have access to the demographic data associated with each code number. Each child will receive a small toy as a reward for participation. Parents will have the option of receiving a brief summary of their child's test performance upon their request.

Instrumentation

The Differential Ability Scales.

The Differential Ability Scales (DAS) is an individually administered battery of 17 cognitive and three achievement tests for children and adolescents between the ages of two years, six months (2:6) and 17 years 11 months (17:11). Children ages 2:6 to 3:5 are typically administered the Lower Preschool Battery, which consists of six subtests; and children ages 3:6 to 5:11 are administered the Upper Preschool Level that consists of ten subtests (Dumont, Willis & Sattler, 2001). The cognitive subtests of the DAS are divided into core subtests and diagnostic subtests. The core subtests are those most strongly related to *g* and used to compute the General Conceptual Ability (GCA) composite score. All core subtests weigh equally into the GCA score. At the Lower Preschool Level four core subtests combine to yield the GCA; at the Upper Preschool Level the GCA is computed based upon six subtests.

The Upper Preschool Level also provides two cluster scores, Verbal Ability and Nonverbal Ability that are also based on core subtests. The four diagnostic subtests of the Upper Preschool Level measure additional specific abilities. They are useful in examining an examinee's strengths and weaknesses, across specific domains (e.g. short-term auditory memory, visual discrimination, short-term and intermediate-term verbal recall, and short-term visual recognition) but do not contribute to the composite scores (Dumont et al., 2001). The subtests of the Preschool Battery of the DAS are summarized in Table 3.1.

Table 3.1

DAS Preschool Level Subtests

Subtest - Age Ranges/Cluster Score	Description
Core Subtests:	
Block Building - 2:6-3:5* GCA	Measures visual-perceptual matching of spatial orientation, examinees are required to copy two and three-dimensional block patterns
Copying – 3:6-5:11 GCA, Nonverbal Ability	Measures visual-perceptual matching and fine motor coordination, examinees must reproduce line drawings
Early Number Concepts - 3:6-5:11 GCA	Measures knowledge of prenumeric and numerical concepts, examinees must answer questions by using colored chips or pointing to drawings
Naming Vocabulary - 2:6-5:11* GCA, Verbal Ability	Measure of expressive language and knowledge of vocabulary, examinees are shown an object or picture and asked to say its name
Pattern Construction - 3:6-5:11 GCA, Nonverbal Ability	Measure of nonverbal reasoning and spatial visualization, examinees must reproduce designs using colored blocks
Picture Similarities - 2:6-5:11* GCA, Nonverbal Ability	Measures nonverbal reasoning, when shown a group of four pictures examinees must place a fifth picture with the card that has a common element
Verbal Comprehension - 2:6-5:11* GCA, Verbal Ability	Measure of receptive language, examinees are required to point to pictures or manipulate objects in response to oral directions involving basic concepts

Table 3.1

DAS Preschool Level Subtests (cont.)

Subtest – Age Range/Cluster Score	Description
Diagnostic Subtests:	
Matching Letter-Like Forms - 4:6-5:11	Measures visual discrimination ability, examinees must find the identical match to an abstract figure among distracters that are rotated or reversed
Recall of Digits - 3:0-5:11*	Measures short-term auditory memory, examinees must repeat a sequence of numbers presented orally
Recall of Objects - 4:0-5:11	Measures short-term and intermediate-term recall of verbal and pictorial information, examinees view a card with pictures of 20 objects for a specified amount of time, and must recall as many objects as possible after the card is removed
Recognition of Pictures - 3:0-5:11*	Measures short-term nonverbal visual memory, examinees are shown a picture of one or more objects for 5 to 10 seconds and then must identify the previously viewed objects among various distracters

Note. Adapted from *Differential Ability Scales introductory and technical handbook*. (1990). C.D. Elliott. San Antonio, TX: Psychological Corporation.
*indicates tests of the Lower Preschool Battery

Flanagan and Alfonso (1995, 1999) evaluated the DAS in terms of standardization, reliability, test floors, item gradient, and validity as well as qualitative characteristics. The results of their findings are summarized in the following pages. The DAS was standardized on a group of 3,475 children between the ages of 2:6 and 17:11. The sample was representative of the United States population and included children with various disabilities. According to Flanagan and Alfonso (1995), the standardization of the DAS is good when evaluated against the criterion described in Chapter Two.

The overall reliability of the DAS is considered adequate based upon the criteria established by Flanagan and Alfonso (1995). The test-retest sample and interval was appropriate for generalization to the preschool age range, and results yielded a test-retest reliability coefficient of r = .90. Internal consistency reliability coefficients were $r = .90$ or higher across the preschool age range, with the exception of ages 3:0 and 4:6, when the coefficients were slightly lower ($r = .89$). This data suggests that the DAS generally measures what it intends to measure reliably, and yields consistent estimates of ability over time (Alfonso & Flanagan, 1999).

All DAS composite scores have adequate floors across both the preschool levels. However, most of the subtests, with the exception of Verbal Comprehension, have inadequate floors at the beginning of both the Upper and Lower Preschool Levels (ages 2:6 and 3:6, respectively; Alfonso & Flanagan, 1999). As discussed by Elliott, Daniel, and Guiton (1991) the floor of overall composite scores are of greater importance than individual subtest floors, because composite scores are more likely to be utilized in making educational decisions. However, poor floor at the individual subtest level can be problematic. A lack of enough easy items on a particular subtest will produce an inflated score, providing the examiner with misleading information about the child's ability (Bracken & Walker, 1997).

The DAS generally has adequate to good item gradients for the subtests to be used with young children. Because of this the DAS is able to detect differences across the average, low average, borderline, and mild mental retardation ranges of ability in young children. That is, the subtests of the DAS are sensitive to minor differences in ability because a child's success or failure on a single item does not substantially change their standard score on that measure (Alfonso & Flanagan, 1999). This is an important consideration for the proposed study as it will include a special population of children receiving special education services and therefore lower scores across subtests may emerge.

The DAS also demonstrates several strong qualitative characteristics when used with young children (Alfonso & Flanagan, 1999). For example, most DAS subtests include sample and teaching items, second trials, and examiner demonstrations in order to ensure that the child has understood the task. The administration time of the DAS rarely exceeds one hour and many of the subtests require minimal expressive language skills. This quality is of particular importance for the proposed study, as the children participating will have speech and language delays and may not be proficient in expressing themselves verbally. An additional benefit of DAS administration is the inclusion of alternate stopping rules. According to Elliott (1990b), these were designed to help prevent "over-testing" and minimize a child's frustration level.

The most significant limitation of the DAS in terms of qualitative structure is the complexity of subtest directions. Although the subtest directions tend to be short and simply structured, they contain several concepts that may not be well understood by some young children (Alfonso & Flanagan, 1999). Several of the subtests' directions contain relational concepts (e.g. "on," "under," "right," "down") that young children typically do not attain understanding of until the later preschool years (Flanagan et al., 2000).

The Woodcock-Johnson-Third Edition Tests of Cognitive Abilities.

The Woodcock Johnson Third Edition Tests of Cognitive Abilities (WJ-III COG) is the cognitive component of the Woodcock Johnson Psychoeducational Battery – Third Edition. This battery was designed to assess cognitive abilities, scholastic aptitude, and achievement of individuals aged two to adult. The WJ-III COG scale is divided into two sections: the Standard Battery and the Extended Battery, which yields a variety of composite and cluster scores. Each battery consists of 10 individual tests. These tests are used to compute the General Intellectual Ability (GIA) score. Seven of the subtests of the Standard Battery combine to provide the GIA-Std score; the GIA-Ext score is computed with the addition of 7 subtests from the Extended Battery.

The WJ-III subtests also yield Cognitive Performance Cluster scores in the areas of Verbal Ability, Thinking Ability, and Cognitive Efficiency. Additionally, CHC Factor Cluster scores are obtained in the areas of Comprehension-Knowledge (*Gc*), Long-Term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv*), Auditory Processing (*Ga*), Fluid Reasoning (*Gf*), Processing Speed (*Gs*), and Short-Term Memory (*Gsm*; McGrew & Woodcock, 2001). The subtests of the WJ-III are described in Table 3.2.

Table 3.2

WJ-III Subtests

Test/Cluster Score	Description
Standard Battery:	
Verbal Comprehension GIA-Std, GIA-Ext, BIA, Verbal Ability-Std, Verbal Ability-Ext, *Gc*	Measure of acquired knowledge, requires examinees to identify synonyms, antonyms, and complete verbal analogies
Visual-Auditory Learning GIA-Std, GIA-Ext, Thinking Ability-Std, Thinking Ability-Ext, *Glr*	Assesses ability to learn and recall pictographic representations of words, requires examinees to associate a name with a symbol, then read each symbol in a story format
Spatial Relations GIA-Std, GIA-Ext, Thinking Ability-Std, Thinking Ability-Ext, *Gv*	Measures spatial ability, examinees must identify pieces that form a complete shape
Sound Blending GIA-Std, GIA-Ext, Thinking Ability-Std, Thinking Ability-Ext, *Ga*	Assesses the ability to synthesize phonemes, examinees listen to an audio recording and blend sounds presented into a word
Concept Formation GIA-Std, GIA-Ext, BIA, Thinking Ability-Std, Thinking Ability-Ext, *Gf*	Measures the ability to derive, categorize, and identify rules from complex stimulus; examinees must determine rules relating to the categorization of visual stimuli
Visual Matching GIA-Std, GIA-Ext, Cognitive Efficiency-Std, Cognitive Efficiency-Ext, *Gs*	Measures processing speed, examinees are required to match two identical numbers in a row of six numbers
Numbers Reversed GIA-Std, GIA-Ext, Cognitive Efficiency-Std, Cognitive Efficiency-Ext, *Gsm*	Measures short-term memory, specifically working memory, the examinee must hold numbers presented orally in immediate memory and repeat them in reverse order

Table 3.2

WJ-III Subtests (cont.)

Test/Cluster Score	Description
Incomplete Words	Assesses auditory analysis and closure abilities, examinees are required to identify a word after hearing it via audio recording with one or more phonemes missing
Auditory Working Memory	Measures short-term memory, examinees listen to a series of digits and letters and then repeat the letters first, then the digits in their sequential order
Visual-Auditory Learning-Delayed	Provides an index of long-term retrieval abilities, examinees are required to relearn the stimulus items presented in the Visual Auditory Learning subtest
Extended Battery:	
General Information GIA-Ext, Verbal Ability-Ext, *Gc*	Measures depth of general knowledge, examinees must orally identify where objects are found and what you typically do with objects
Retrieval Fluency GIA-Ext, Thinking Ability-Ext, *Glr*	Assesses fluency of retrieval of stored knowledge, requires an examinee to list as many items as possible from a given category within a one-minute time limit
Picture Recognition GIA-Ext, Thinking Ability-Ext, *Gv*	Assesses visual memory, the examinee must identify a set of previously presented pictures within a group of distractors
Auditory Attention GIA-Ext, Thinking Ability-Ext, *Ga*	Measures auditory discrimination, examinees listen to a word amid increasing background noise and must point to the corresponding correct picture

Table 3.2

WJ-III Subtests (cont.)

Test/Cluster Score	Description
Analysis-Synthesis GIA-Ext, Thinking Ability-Ext, *Gf*	Measures ability to draw conclusions based upon a given set of conditions, examinees are required to analyze puzzles to determine missing components
Decision Speed GIA-Ext, Cognitive Efficiency-Ext, *Gs*	Measures speed of processing simple concepts, examinees are required to locate and circle similar pictures in a set of stimuli
Memory for Words GIA-Ext, Cognitive Efficiency-Ext, *Gsm*	Measures short-term auditory memory, the examinee must repeat a list of unfamiliar words in correct sequence
Rapid Picture Naming	Assess the ability to quickly name common pictures within a two-minute time limit
Planning	Measures the ability to use forethought in problem solving, examinees must trace a pattern without retracing any lines or removing the pencil from the paper
Pair Cancellation	Measures the ability to stay on task under time restraints, examinees must identify and circle instances of a repeated pattern during a three-minute time limit

Note. Adapted from *Woodcock-Johnson III Tests of Cognitive Abilities Technical Manual* (2001). N Mather & R.W. Woodcock. Itasca, IL: Riverside Publishing Company. GIA-Std = General Intellectual Ability-Standard; GIA-Ext = General Intellectual Ability-Extended; *Gc* = Comprehension-Knowledge; *Glr* = Long-Term Retrieval; *Gv* = Visual-Spatial Thinking; *Ga* = Auditory Processing; *Gf* = Fluid Reasoning; *Gs* = Processing Speed; *Gsm* = Short-Term Memory.

Analysis of the WJ-R Early Development Scale conducted by Flanagan and Alfonso (1995) confirmed that the test battery was a valid and reliable tool appropriate for use with young children. However, because the WJ-III is a recently published instrument, there is little published research available regarding its use other than that provided in the technical manual. Therefore, the standardization, reliability, test floor, item gradient, and validity, as well as the qualitative characteristics of the WJ-III were reviewed by the present author for the purpose of the proposed investigation and are discussed below. The analysis is based upon the criteria established by Flanagan and Alfonso (1995).

The WJ-III was standardized on a group of 8,818 subjects from over 100 diverse United States communities. The norming sample was selected to be representative of the current United States population in terms of gender, geographic region, race/ethnicity, community size and socioeconomic status. The sample consisted of 1,143 preschool aged children (ages 2 to 5, and not enrolled in kindergarten; McGrew & Woodcock, 2001). Thus, the size and representation of the WJ-III standardization sample is considered good and appropriate for young children based upon Flanagan and Alfonso's (1995) criterion.

The WJ-III also has overall adequate reliability based upon Flanagan and Alfonso's (1995) standards. The internal consistency reliability coefficients of the WJ-III COG cluster scores range from adequate to good across the preschool range with the exception of the Fluid Reasoning (*Gf*) and Long-Term Retrieval (*Glr*) score at the two year-old range. In contrast, the reliability of the subtest scores of the WJ-III range from inadequate to good across the preschool age range. The following subtests have inadequate reliability at the age ranges indicated; Spatial Relations (two years), Concept Formation (two years), Retrieval Fluency (two to three years, six years) Picture Recognition (two, five to six years), Decision Speed (three to four years), and Planning (four to six years). The adequate reliability of cluster scores for young children suggests that these scores can be interpreted as accurate representations of the abilities they purport to measure. However, the WJ-III was designed to provide information about specific abilities through individual subtest scores. The inadequate reliability of several individual subtests indicates that these measures cannot necessarily be interpreted for young children.

Overall test-retest reliability coefficients of the WJ-III range from inadequate to good. This suggests that some abilities measured by the WJ-III may change over time. Slightly higher reliability coefficients were found for Acquired Knowledge subtests than those subtests that measure traits that are less stable over time, such as those included in the Thinking Ability and Cognitive Efficiency domains. Typically abilities such as acquired knowledge remain more stable over time while others, such as motoric problem solving, are more susceptible to change due to practice effects and ability changes (McGrew & Woodcock, 2001).

The WJ-III test-retest sample consisted of 1,196 participants and was representative of the United States population. The sample contained 235 children between the ages of two and seven. However, the technical manual does not state the distribution of the children over this age range. One hundred sixty six of the children in the test-retest sample were tested within a span of two years. The remaining 69 children were tested within a period of 3 to 10 years. Because of the limited information that is provided about the test-retest sample it is difficult to conclude if the information gained as a result of this study can be generalized to all young children. Based upon Flanagan and Alfonso's (1995) criterion, the standardization of the WJ-III is adequate.

Using the WJ-III Compuscore (Schrank & Woodcock, 2001), the floor of WJ-III cluster scores was analyzed by the present author. Adequate floor was defined as a standard score at least two standard deviations below the normative mean when a raw score of one was entered into the scoring program. Findings indicate that while the test included children as young as two years of age in the norming process, adequate floor for the composite GIA scores is not demonstrated until 4:2 (GIA Standard) and 3:9 (GIA Extended). The BIA score demonstrates adequate floor at 3:11. Thus, the WJ-III COG is best used with children at least 3:9 when the ability to obtain scores that differentiate across the lower levels of functioning is important. The youngest age at which adequate floor is demonstrated across the remaining cluster scores is 3:2 (Visual Processing, *Gv*). Half of the cluster scores demonstrated adequate floor by four years of age, and three fourths of them by 4:6.

It is the author's opinion that the WJ-III COG has several qualitative assets and some limitations in terms of its use with young children. The technical and administration manuals of the test are comprehensive and user-friendly. Each subtest begins with sample items that can be repeated and taught to the subject if they appear to be having difficulty grasping the concept presented. Some subtests also allow the examinee to point to their responses, which can be beneficial for young children with limited language abilities. Additionally, test administration can be organized to administer tasks that do not require expressive language first, allowing the child to become more comfortable prior to being required to verbalize responses.

A limitation of the WJ-III COG is that it does not contain any manipulatives to be used during testing. As such, it may be difficult to hold a young child's attention during testing. Additionally, WJ-III COG subtests do not have alternative stopping points, which serve to discontinue administration of a subtest before a child becomes overly frustrated with many difficult items. Rather, subtest administration ends when a subject has reached a predetermined ceiling. As a result, test administration may become lengthy if a child must be administered many items prior to establishing a ceiling. Despite the fact that certain subtests require motoric responses, such as pointing or circling an answer, the majority of the WJ-III COG relies on oral responses. This may be problematic for very young children with limited expressive vocabulary skills.

Analyses

Pearson product-moment correlations will be calculated to examine the relationship between the broad and cluster scores of each battery, the relationship between the WJ-III BIA score and the DAS GCA score, and the relationship between the DAS diagnostic subtests and the WJ-III cluster scores. Means, standard deviations, and range of scores will be calculated for the broad scores of each battery to determine if the performance outcomes of each battery differ significantly.

Expected Findings

Based upon previous research, as well as the findings of a study examining the concurrent validity of the DAS and WJ-III COG with young children reported in the WJ-III COG manual, the following findings are expected from the proposed study.

Strong correlations are expected between the GCA score of the DAS and the GIA-Std and GIA-Ext scores of the WJ-III COG. Mean scores between the two batteries are expected to be similar as well.

Moderate to strong correlations are expected between the Verbal Ability cluster score of the DAS and the Verbal Ability-Std and Verbal Ability-Ext scores of the WJ-III COG. In addition, a moderate to strong correlation is expected between the WJ-III COG

Comprehension-Knowledge (*Gc*) score and the DAS Verbal Ability score. Weak correlations are expected between other cluster scores of each battery with these clusters.

Moderate correlations are expected between the DAS Nonverbal Ability score and the Thinking Ability-Std and Thinking Ability-Ext scores of the WJ-III COG. Similarly, moderate correlations are expected between the Fluid Reasoning (*Gf*) and Visual-Spatial Thinking (*Gv*) cluster scores of the WJ-III COG and the DAS Nonverbal Ability cluster. Weak correlations are expected between other cluster scores of each battery with these clusters.

Because of a lack of comparable measures between the two batteries weak correlations are expected between the Cognitive Efficiency-Std and Cognitive Efficiency-Ext and the cluster scores of the DAS. It is expected that the Cognitive Efficiency-Std and Cognitive Efficiency-Ext scores will correlate most strongly with the DAS GCA score. Similarly, weak correlations between the Processing Speed (*Gs*) and Short-Term Memory *(Gsm*), and Auditory Processing (*Ga*) clusters of the WJ-III COG and the clusters of the DAS are expected.

Given that Elliott (1990b) describes the diagnostic subtests as distinct measures of specific cognitive abilities, moderate to strong correlations of DAS subtests and WJ-III *Gf-Gc* clusters are expected. Specifically, the WJ-III Long-Term Retrieval (*Glr*) cluster score is expected to demonstrate a moderate to strong correlation with the DAS Recall of Objects diagnostic subtest. Similarly, the DAS Recall of Digits diagnostic subtests is anticipated to correlate strongly with the WJ-III COG Short-Term Memory *(Gsm*) cluster score. The Recognition of Pictures and Matching Letter-Like Forms diagnostic subtests of the DAS are expected to correlate moderately with the Visual-Spatial Reasoning (*Gv*) cluster score of the WJ-III COG.

The WJ-III COG BIA score is expected to have moderate to strong correlations with the DAS GCA score.

References

Alfonso, V.C., & Flanagan, D.P. (1999). Assessment of cognitive functioning in preschoolers. In E.V. Nuttal, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers: Psychological and educational dimensions* (pp. 186-217). Boston: Allyn and Bacon.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Bagnato, S.J., & Neisworth, J.T. (1994). A national study of the social and treatment "invalidity" of intelligence testing for early intervention. *School Psychology Quarterly, 9*(2), 81-102.

Bayley, N. (1993). *Bayley Scales of Infant Development-II.* San Antonio, TX: Psychological Corporation.

Bracken, B.A. (1984). *Bracken Basic Concept Scale*. San Antonio, TX: Psychological Corporation.

Bracken, B.A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 4*(1), 313-326.

Bracken, B.A. (1994). Advocating for effective preschool assessment practices: A comment on Bagnato and Neisworth. *School Psychology Quarterly, 9*(2), 103-108.

Bracken, B., & Walker, K.C. (1997). The utility of intelligence tests for preschool children. In D.F Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 484-502). New York: Guilford Press.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge Press.

Demers, S.T., & Fiorello, C. (1999). Legal and ethical issues in preschool assessment and screening. In E.V. Nuttal, I. Romero, & J. Kalesnik, (Eds.), *Assessing and screening preschoolers: Psychological and educational dimensions* (pp. 50-58). Boston: Allyn and Bacon.

Dumont, R., Willis, J., & Sattler, J.M. (2001). Differential Ability Scales. In J.M. Sattler (Ed.), *Assessment of children: Cognitive applications* (4^{th} ed.) (pp.507-545). Le Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Dumont, R., Willis, J.O., Farr, L.P., McCarthy, T., & Price, L. (2000). The relationship between the Differential Ability Scales (DAS) and the Woodcock-Johnson Tests of Cognitive Ability-Revised (WJ-R COG) for students referred for special education evaluations. *Journal of Psychoeducational Assessment, 18*, 27-38.

Dumtschin, J.U. (1988). Recognizing language development and delay in early childhood. *Young Children, 42*(3), 16-24.

Early Childhood Special Education. (n.d.). Retrieved July 18, 2001, from Southwest Missouri State University Web site: http://www.courses.smsu.edu/dg734f/

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

Elliott, C.D. (1990b). *Differential Ability Scales: Introductory and technical handbook.* San Antonio, TX: Psychological Corporation.

Elliott, C.D. (1997). The Differential Ability Scales. In D.F. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp.183-208). New York: Guilford Press.

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

Flanagan, D.P., & Alfonso, V.C. (1995). A critical review of the technical characteristics of new and recently revised intelligence tests for preschool children. *Journal of Psychoeducational Assessment, 13*, 66-90.

Flanagan, D.P., Mascolo, J., & Genshaft, J.L. (2000). A conceptual framework for interpreting preschool intelligence tests. In B.A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (pp.428-473). Boston: Allyn & Bacon.

Feldman, R.S. (1997). *Development across the life span*. Upper Saddle River, NJ: Prentice Hall.

Harrison, P.L. (1990). *Early Screening Profiles.* Circle Pines, MN: American Guidance Service.

Hetherington, E.M., & Parke, R.D. (1999). *Child psychology: A contemporary viewpoint* (5th ed.). Boston: McGraw-Hill College.

Horn, J.L., & Cattell, R.B. (1967). Age difference in fluid and crystallized intelligence. *Acta Pyschologica, 26*, 107-129.

Individuals with Disabilities Education Act (IDEA). H.R. 5, Fed. Reg. 10-476. (1990, 1997).

Kazdin, A.E. (1992). *Research design in clinical psychology* (2nd ed.). Boston: Allyn & Bacon.

McCarthy, D. (1972). *McCarthy Scale of Children's Abilities*. New York: Psychological Corporation.

McGrew, K.S. & Flanagan, D.P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Boston: Allyn & Bacon.

McGrew, K.S., & Woodcock, R.W. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities technical manual*. Itasca, IL: Riverside Publishing.

McIntosh, D.E., Brown, M.L., & Ross, S.L. (1995). Relationship between the Bracken Basic Concepts Scale and the Differential Ability Scales with an at-risk sample of preschoolers. *Psychological Reports, 76*, 219-224.

McIntosh, D.E., Gibney, L., Quinn, K., & Kundert, D. (2000). Concurrent validity of the Early Screening Profiles and the Differential Ability Scales with an at-risk preschool sample. *Psychology in the Schools, 37*(3), 201-207.

McIntosh, D.E., Wayland, S.J., Gridley, B., & Barnes, L.B. (1995). The relationship between the Bracken Basic Concepts scale and the Differential Ability Scales with a preschool sample. *Journal of Psychoeducational Assessment, 13*, 39-48.

National Association of School Psychologists (NASP) Early Childhood listserve (2001). Retrieved July 18, 2001 from http://groups.yahoo.com/group/NASP-IG-EarlyChild/

Nagle, R.J. (2000). Issues in preschool assessment. In B.A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (pp.19-32). Boston: Allyn & Bacon.

Reschly, D.J., & Grimes, J. P. (1995). Best practices in intellectual assessment. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 763-773). Washington, DC: The National Association of School Psychologists.

Roid, G.H., Miller, J.L. (1997). *Leiter International Performance Scale-Revised.* Wood Dale, IL: Stoelting Corporation.

Romero, I. (1999). Individual assessment procedures with preschool children. In E.V. Nuttal, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers: Psychological and educational dimensions* (pp. 59-71). Boston: Allyn and Bacon.

Sattler, J. (1992). *Assessment of children: Revised and updated third edition.* San Diego, CA: Jerome Sattler, Publisher, Inc.

Schrank, F. A., & Woodcock, R. W. (2001). WJ III Compuscore (Version1.1b) [Computer software]. Itasca, IL: Riverside Publishing.

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

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: Psychological Corporation.

Woodcock, R.W., & Johnson, M.B. (1977). *Woodcock-Johnson Psychoeducational Battery*. Itasca, IL: Riverside Publishing.

Woodcock, R.W., & Johnson, M.B. (1989). *Woodcock-Johnson-Revised Tests of Cognitive Ability*. Itasca, IL: Riverside Publishing.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.

Wyatt, T.A., & Seymour, H.N. (1999). Assessing the speech and language skills of preschool children. In E.V. Nuttal, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers: Psychological and educational dimensions* (pp. 218-239). Boston: Allyn and Bacon.

Appendix A
Consent Form

Dear Parent/Guardian:

I am a graduate student in the School Psychology program at the University of Wisconsin-Stout. I am currently obtaining data for my thesis. The purpose of my study is to examine two instruments that assess cognitive abilities of preschool age children. This information is important for professionals who work with children so they may provide appropriate educational services based upon a child's academic abilities.

I would like to ask for your permission for your child to participate in this study. Participation would involve administering two intellectual assessments to your child, the Differential Ability Scales and the Woodcock-Johnson Tests of Cognitive Abilities – Third Edition. Administration of each assessment will take approximately 60 to 90 minutes. The tests will be administered on separate occasions within the span of 2 to 3 weeks.

Your child will gain valuable test-taking experience by participating. The tests are generally interesting and challenging. Children will receive a small toy as a reward for their participation. You may receive a brief summary of the results of your child's test performance upon request. Each child will receive a code number so that the results of the assessments will be kept completely confidential. Your child will be free to discontinue testing at any point if they do not feel comfortable completing the assessment.

If you have any questions about the study please feel free to contact Jennifer Salava or Mary Beth Tusing at the University of Wisconsin-Stout -- (715) 232-2211.

Thank you,

Jennifer Salava
University of Wisconsin-Stout

I give consent for my child to participate in this study.

Signature ____

Child's name_____________________________________ Male ______ Female_____

Telephone number _________________________________ Child's Birthdate ___________

The best time to reach me is:

____ morning
____ afternoon
____ evening
____ other (please fill in)

I would like a brief summary of my child's performance _____ yes ______ no

I would be willing to transport my child to the University of Wisconsin-Stout for testing on two occasions _____ yes _______ no (if this is not possible alternate accommodations will be made)

Appendix B
Demographic Data Form

WJ-III /DAS PRESCHOOL STUDY
DEMOGRAPHIC DATA

Identification Number _______________

Birthdate ____________ Age______________

School _________________________________ Grade____________

Gender: ______ Male
______ Female

Ethnicity: ______ Caucasian
______ African American
______ Hispanic
______ Asian
______ Native American
______ Other (please specify ____________________)

What is the highest educational level of the child's **mother / female guardian**?
______ less than high school (please specify highest grade ______)
______ high school graduate or GED
______ some college
______ bachelor's degree
______ some graduate school
______ graduate degree (please specify degree ____________________)
______ technical school

What is the highest educational level of the child's **father/ male guardian**?
______ less than high school (please specify highest grade ______)
______ high school graduate or GED
______ some college
______ bachelor's degree
______ some graduate school
______ graduate degree (please specify degree ____________________)
______ technical school