

Concurrent Validity of the Woodcock Johnson Third Edition Tests of Cognitive  
Ability and the Differential Ability Scales for Young Children with Speech and  
Language Delays

by

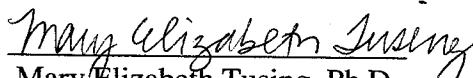
Jennifer Anne Salava


A Research Paper


Submitted in Partial Fulfillment of the  
Requirements for the  
Education Specialist degree  
in

School Psychology

Approved: 6 Semester Credits

  
Mary Elizabeth Tusing, Ph.D.

  
Scott Orme, Ph. D.

  
Amy Gillett, Ph.D.

The Graduate School

University of Wisconsin-Stout

November 2004

The Graduate School  
University of Wisconsin Stout  
Menomonie, WI 54751

ABSTRACT

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

Concurrent Validity of the Woodcock Johnson Third Edition Tests of Cognitive Ability  
and the Differential Ability Scales for Young Children with Speech and Language Delays  
(Title)

|                   |                      |             |              |
|-------------------|----------------------|-------------|--------------|
| School Psychology | Dr. Mary Beth Tusing | 11/2004     | 109          |
| (Graduate Major)  | (Research Advisor)   | (Mon. /Yr.) | (# of Pages) |

|   |
|---|
| <u>American Psychological Association (APA) Publication Manual (5<sup>th</sup> ed.)</u> |
| (Name of Style Manual Used in this Study)   |

The purpose of this paper is to present the findings from an investigation of the concurrent validity between the Woodcock-Johnson III Tests of Cognitive Ability (WJ III) and the Differential Ability Scales (DAS) when used with young children receiving early childhood special education services for speech and language delays. Thirty young children between the ages of 3 years, 6 months and 5 years, 11 months were administered both the WJ III and the DAS. The results were compared through data analysis of scores obtained on each test battery. Consistent with previous research, results implied that the WJ III and DAS measure similar constructs. Further, the results also support that the WJ III measures unique cognitive abilities. In general, examinees' performance on both batteries was below average. Findings suggest that language delays may impact performance on a variety of tasks included in cognitive ability tests. Therefore,

practitioners need to be aware of the differences in the abilities measured by each test battery when making assessment decisions and interpreting test results.

## Acknowledgements

The author would like to thank the special education directors, early childhood special education teachers, and speech language pathologists from Independent School District #196 and Independent School District #833 who assisted in recruiting participants for this study. In addition, a very special thank you is extended to the members of this thesis committee and the faculty of the University of Wisconsin-Stout school psychology program for their dedication and patience with this project.

## TABLE OF CONTENTS

|  |     |
|--|-----|
| ABSTRACT.....  | ii  |
| List of Tables.....  | vii |
| Chapter I: Introduction.....   | 1   |
| <i>Statement of Problem</i> .....  | 4   |
| <i>Purpose of the Study</i> .....  | 4   |
| <i>Research Questions</i> .....  | 5   |
| <i>Benefits of the Study</i> .....   | 6   |
| <i>Assumptions</i> .....   | 6   |
| <i>Definition of Terms</i> .....   | 7   |
| Chapter II: Review of Relevant Literature.....   | 9   |
| <i>Preschool Assessment</i> .....  | 9   |
| <i>Unique Characteristics of Young Children: Implications for</i><br><i>Assessment</i> ..... | 10  |
| <i>Development of Language Skills</i> .....  | 11  |
| <i>Implications of Language Skills in the Assessment of Young Children</i> .....             | 15  |
| <i>Technical Qualities of Assessment Tools for Young Children</i> .....                      | 18  |
| <i>Standardization</i> .....   | 18  |
| <i>Reliability</i> .....   | 19  |
| <i>Test Floor</i> .....  | 20  |
| <i>Item Gradients</i> .....  | 21  |
| <i>Validity</i> .....  | 22  |
| <i>The Woodcock Johnson III Tests of Cognitive Abilities</i> .....                           | 25  |
| <i>Theoretical Background</i> .....  | 25  |
| <i>Preschool Applications</i> .....  | 26  |
| <i>Validity</i> .....  | 29  |
| <i>Correlations of the WJ III with Other Cognitive Measures</i> .....                        | 30  |
| <i>The Differential Ability Scales</i> .....   | 34  |
| <i>Theoretical Background</i> .....  | 34  |
| <i>Preschool Applications</i> .....  | 36  |
| <i>Validity</i> .....  | 39  |
| <i>Correlations of the DAS with Other Cognitive Measures</i> .....                           | 39  |

|  |     |
|--|-----|
| <i>DAS and Preschool Screening Tools</i> .....                       | 43  |
| <i>Critical Analysis</i> .....                                       | 45  |
| Chapter III: Methodology.....  | 48  |
| <i>Participants</i> .....  | 49  |
| <i>Procedures</i> .....  | 52  |
| <i>Instrumentation</i> .....   | 53  |
| <i>The Woodcock Johnson III Tests of Cognitive Abilities – Early</i> |     |
| <i>Development Scales</i> .....                                      | 53  |
| <i>The Differential Ability Scales</i> .....                         | 59  |
| <i>Analyses</i> .....  | 63  |
| Chapter IV: Results.....   | 64  |
| <i>Descriptive Analyses</i> .....                                    | 64  |
| <i>Research Question One</i> .....                                   | 68  |
| <i>Research Question Two</i> .....                                   | 68  |
| <i>Research Question Three</i> .....                                 | 70  |
| Chapter V: Discussion.....   | 72  |
| <i>Summary</i> .....   | 73  |
| <i>Research Question One</i> .....                                   | 73  |
| <i>Research Question Two</i> .....                                   | 75  |
| <i>Verbal Ability</i> .....  | 75  |
| <i>Nonverbal Ability</i> .....                                       | 78  |
| <i>Research Question Three</i> .....                                 | 80  |
| <i>Conclusions and Implications for Future Study</i> .....           | 83  |
| <i>Limitations of the Study</i> .....                                | 85  |
| References.....  | 88  |
| Appendix A: Consent Form.....  | 96  |
| Appendix B: Demographic Data Form.....                               | 97  |
| Appendix C: Correlation Matrices.....                                | 98  |
| Appendix D: Subtest Mean Scores and Standard Deviations.....         | 101 |

## List of Tables

|   |    |
|---|----|
| Table 2.1: <i>Milestones for Speech and Language Development</i> .....  | 12 |
| Table 2.2: <i>Correlations of the WJ III with Other Measures of Cognitive Ability</i> .....                                     | 31 |
| Table 2.3: <i>Correlations of the DAS with Other Measures of Cognitive Ability</i> .....  | 41 |
| Table 3.1: <i>Demographics of the Sample</i> .....  | 51 |
| Table 3.2: <i>WJ III Tests</i> .....  | 55 |
| Table 3.3: <i>DAS Upper Preschool Level Battery Subtests</i> .....  | 60 |
| Table 4.1: <i>Means, Standard Deviations, and Skewness values for the WJ III Broad and Cluster Scores</i> .....                 | 66 |
| Table 4.2: <i>Means, Standard Deviations, and Skewness Values for the DAS Broad and Cluster Scores</i> .....                    | 67 |
| Table 4.3: <i>Correlations Between Cluster Scores of the WJ III and Cluster Scores and Diagnostic Subtests of the DAS</i> ..... | 69 |

## Chapter I

### *Introduction*

Traditionally, intellectual assessment has been a major part of the school psychologist's role in the educational community, as the Individuals with Disabilities Education Act (IDEA, 2003) 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 to developmental, learning, and behavior problems in subsequent years (Harrison, 1990).

In response to research supporting the efficacy of early intervention services, the 1986 revision, as well as subsequent revisions (1990, 1997, 2003) of special education law mandates that special education services also be offered to infants, toddlers, and preschool-age children (Demers & Fiorello, 1999). As such, preschool-age children must be given the same opportunities for 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, 1999).

Because of this legislation, school psychologists have experienced an increased role in the assessment of young children in the past decades. 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 (Flanagan & Alfonso, 1995). 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 contributing to the diagnostic decision-making 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 to survey findings indicating that most school psychologists 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; Alfonso & Flanagan, 1999) for use with young children were the Differential Ability Scales (Elliott, 1990a) and the Woodcock-Johnson Revised Tests of Cognitive Ability (Woodcock & Johnson, 1989).

In addition to assessing the quantitative characteristics of each test battery, Alfonso and Flanagan (1999) also examined the qualitative properties of the instruments, in particular the level of language complexity in subtest directions. Complexity of test

directions, including the level of basic concepts included in the wording, is one of the most important qualitative characteristics of a test instrument. The validity of a score can be greatly compromised if a child does not earn credit for a particular set of items due to the fact that they did not understand test directions, rather than because they have less than average ability in the area being assessed. Again, the authors rated the DAS and WJ-R favorably within this construct. However, despite being reviewed favorably, the authors called for more research on the use of each test battery with young children, particularly young children with suspected disabilities. Specifically, further investigation related to the use of intelligence instruments with young children with language delays is warranted.

Recent discussions on the Early Childhood listserv published on the National Association of School Psychologists (NASP) website ([www.nasponline.org](http://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 (1995), the Differential Ability Scales appears to have gained increased popularity among practitioners working with young children (NASP, 2001) and the WJ-R has been revised and republished as the Woodcock-Johnson Third Edition Tests of Cognitive Ability (WJ III; Woodcock, McGrew, & Mather, 2001) and Diagnostic Supplement to the Tests of Cognitive Abilities (WJ III DS; Woodcock, McGrew, Mather, & Schrank, 2003). Improvements in the WJ III's test construction and assessment technology suggest that it may be an even more adequate tool for use with young children than the WJ-R. For one, the WJ III provides additional cluster scores that were not components of the WJ-R (McGrew & Woodcock, 2001). Further, the Diagnostic

Supplement includes subtests specifically designed for use with young children, that when combined with tests from the WJ III COG allows for an Early Development GIA (EDev GIA) score. The EDev GIA provides a measure of overall cognitive ability specifically for children as young as two years of age.

### *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 with 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 and WJ III Diagnostic Supplement have 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 for preschool-age children, and no published research has examined the tool's use with special populations of young children, including children with speech and language delays. Because of the high linguistic demand of many traditional assessment tools, and the fact the complexity of subtest directions may compromise the validity of test scores, further investigation related to this population of children is necessary.

### *Purpose of the Study*

The purpose of this study is to examine the concurrent validity of two intellectual assessment instruments designed for use with young children, the Differential Ability Scales (DAS) and select portions of the Woodcock Johnson-Third Edition Tests of Cognitive Abilities (WJ III) and WJ III Diagnostic Supplement. The study will examine the use of these assessment tools with a special population of young children, specifically

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 topic addressed in this study will be the relationship between the broad scores of the DAS and the WJ III. The General Conceptual Ability (GCA) score of the DAS will be compared with the WJ III Early Development General Intellectual Ability (EDev GIA) score. Mean composite scores across the two test batteries will also be examined.
2. The second area addressed in the study will be the level of concurrent validity between specific 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 measures of ability will be examined within each respective battery. Specific cluster scores from the WJ III that will be analyzed include: Comprehension Knowledge (*Gc*), Long Term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv / Gv3*), Phonemic Awareness, and Associative Memory.
3. The study will also examine the relationship between the diagnostic subtests of the DAS and select 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.

### *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 test batteries.

### *Assumptions*

In order for the proposed study to adequately address the research questions, the following assumptions were made. First, it was assumed that all administrations of the DAS and WJ III followed standardized testing procedures and tests were scored

appropriately. All assessments were administered by individuals specifically trained in administering the DAS and WJ III to young children. It was also assumed that the instruments investigated were reliable measures of the cognitive abilities that they purport to measure. This is supported by a review of available technical information for each test, which is reported in Chapter Two.

### *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 test 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 the 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.

### *Linguistic Demand*

The linguistic demand of an intelligence test is defined as the level of complexity of the language contained in the wording of subtest directions. When this concept is considered for tests designed to be used with young children, the prevalence of basic concepts (knowledge of wording such as *first*, *next*, *over*, *under*, etc.) is of particular importance.

### *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 (2003). Children in this age range are also often referred to as preschoolers.

## Chapter II

### *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 with speech and language delays. 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. Specific emphasis is given to a discussion of children's language skills as they relate to intellectual assessment. Current research concerning the concurrent validity of the two instruments is also 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. 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). 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.

Young children 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. 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). Additionally, Romero (1999) 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 shoulder or a silent nonresponse. Due to the unique challenges of the assessment of young children, it is important that examiners take these factors into account when interpreting the results of an evaluation. In addition, experience in working with young children, as well as a thorough understanding of assessment instruments, allows the examiner to make appropriate inferences about a child's behaviors (Romero, 1999).

Another important aspect of development that can potentially impact the cognitive assessment of young children is their language ability. Children with developmental delays in speech and language may not 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 and its relationship to the assessment of cognitive ability 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       | <p>Crying serves as the earliest form of communication</p> <p>Babbling stage at 4 to 6 months, begin experimenting with sounds</p>  |
| 10 to 12 months      | <p>Begin to use nonverbal gestures, such as pointing</p> <p>Emergence of first true words, initial vocabulary of 10 words</p> <p>Understand more words than produced</p>  |
| 18 months to 2 years | <p>Begin to produce two-word combinations</p> <p>Vocabulary of approximately 50 words, mainly nouns</p> <p>Frequently over-extend and under-extend the meaning of words</p> <p>Often produce simplified and modified versions of words</p>  |
| 2 to 3 years         | <p>Begin using key grammatical forms (e.g. plurals, "ing")</p> <p>Begin developing use of pronouns (e.g. "I," "mine," "me")</p> <p>Begin to use first negatives (e.g. "no," "not," "can't")</p> <p>Begin to use early helping verbs (e.g. "do," "be")</p> <p>Begin asking "what," "where," yes/no questions</p> |

Table 2.1

*Milestones for Speech and Language Development (cont.)*

| Age          | Developmental Milestone   |
|--------------|---|
| 3 to 4 years | <p>Begin to master grammatical rules of language</p> <p>Consistently producing the following sounds “n,” “m,” “p,” “b,” “h,” “k,” “g,” “f,” “w,” “y,” “t,” and “d”</p> <p>Able to produce words with more adult-like pronunciation</p>  |
| 4 to 5 years | <p>Begin to use more complex negatives, questions, pronouns,</p> <p>Average sentence length 4 to 5 words, more complex appearing</p> <p>Starting to master more complex sounds such as “s,” “z,” “l,” “r,” “sh,” and “ch,” however, errors still present</p> <p>Beginning to master conventional rules of speech</p> <p>Begin to adapt language to meet the needs of their listener</p> |

*Note.* Adapted from *Assessing and Screening Preschoolers: Psychological and Educational Dimensions* (pp. 218-239), Nuttal, E.V., Romero, I., & Kalesnik, J, 1999, Boston: Allyn & 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, children with language delays many also have difficulty classifying objects and recognizing similarities and differences (Dumtschin, 1988). As many cognitive assessment tools for young children include verbal subtests related to object naming and identification of similarities, as well as non-language based tasks with verbal directions, a language delay may impact a child's performance on a standardized assessment in multiple ways (Flanagan, Alfonso, Kaminer & Rader, 1995).

Under IDEA (2003), 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, 2003).

### *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). If a child with a language delay has difficulty comprehending language, they may not understand what is required of them during test administration. For example, 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). When related to subtest directions, excessive linguistic demand can impact

language based and non-language based abilities, such as memory and visual-spatial tasks, alike.

The administration of many tests of cognitive ability also relies heavily on verbal responses from the child. This is problematic in two ways. First, children with expressive language disorders may have difficulty demonstrating their knowledge of the domain being assessed. Secondly, a child's articulation difficulties may impede the examiner's ability to understand the child. This is particularly evident in the preschool age range given that children are not expected to have mastered the entire phonological system until as late as seven years of age (Wyatt & Seymour, 1999). While most tests do not penalize a child's response due to articulation problems, children with significant articulation disorders can be challenging to understand and therefore assess. 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. Additionally, although one would expect a child with an identified language delay to have difficulty with tasks specifically measuring expressive language, such as a vocabulary subtest, the overall cognitive ability score on many assessment instruments is heavily influenced by a child's performance on the verbal ability cluster. As such, difficulty with expressive language tasks may unfairly impact a child's overall score as well (Flanagan et al., 2000).

Beyond impacting performance on language based tasks, language delays impact performance on additional areas of cognitive ability assessment instruments for young children as well. For instance, if comprehension of a basic concept in test instructions is inextricably linked to the ability to perform a required task, a child's score on the measure, whether language based or not, is dependent upon their capability to understand the directions (Flanagan et al, 1995). For example, if during a visual-spatial task a child is shown a series of pictures and asked to identify which one should come "after" a particular image, difficulty understanding the meaning of the word "after" will impact their performance, in spite of the fact that the task is assessing a nonlanguage based ability. Similarly, being asked to place a picture "next" to one of several pictures that it is most like requires that a child have a thorough understanding of the meaning of the word "next" in order to adequately perform the task.

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, particularly young children with speech and language delays. Additional technical qualities to be considered when evaluating assessment tools to be used with all young children are discussed below.

### *Technical Qualities of Assessment Tools for Young Children*

Flanagan and Alfonso (1995) reviewed the commonly used preschool cognitive assessment tools in the mid-1990's (Wechsler Preschool and Primary Scales of Intelligence- Revised (Wechsler, 1989), the Differential Ability Scales (Elliott, 1990a), the Stanford-Binet Intelligence Scale: Fourth Edition (Thorndike, Hagen, & Sattler, 1986), the Woodcock-Johnson Psycho-Educational Battery-Revised: Tests of Cognitive Ability (Woodcock & Johnson, 1989), and the Bayley Scales of Infant Development-II (Bayley, 1993) according to a number of technical characteristics; namely, standardization characteristics, reliability, test floor, item gradients, and validity. 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). Descriptions of the domains and findings are provided below.

#### *Standardization.*

Intelligence tests have no overall 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 standardization process a test be 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. In Flanagan and Alfonso's (1995) review, all major intelligence test batteries for preschool age children had 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). 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 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. 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). However, 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 tests reviewed remain 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. Inadequate test floors may result in an overestimation of ability for children with delayed functioning, 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 on a subtest. The better item gradients a subtest 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 standard deviation. 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 measure, 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, NCME, 1999; 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 claim to 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. Good concurrent validity of an instrument is suggested if a measure shows high correlations between similar constructs and low correlations between dissimilar constructs (AERA, APA, NCME, 1999; 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 specify the conditions under which an instrument is considered valid (Bracken, 1987; Flanagan & Alfonso, 1995; Bracken & Walker, 1997). According to Bracken and Walker (1997), 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 various technical characteristics reported above, Flanagan and Alfonso (1995) determined that, in terms of psychometric properties, the DAS and the WJ-R were two of the most appropriate intelligence tests for use with young children. However, they also cautioned that continued study of the validity of both tests in their use with young children was needed. Since Flanagan and Alfonso's 1995 review, the WJ-R has undergone major revisions and has been published

as the WJ III. To further the work of Flanagan and Alfonso, appropriate versions of the DAS and WJ III for use with young children were reviewed for this 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 higher order factor measures 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.

The WJ III also provides CHC factor scores that represent 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).

*Preschool Applications.*

The 2001 revision of the WJ III emphasized the creation of a battery of tests with greater generalizability at the cluster score level (Tusing, Maricle, & Ford, 2003). Subtest revision expand portions of the WJ III to include early development items appropriate for children as young as the age of two. However, as this revision impacted subtest selection for each cluster, the WJ III does not, by itself, provide an overall ability score for young children (Tusing et. al, 2003). As such, the recently published Diagnostic Supplement (Schrank, Mather, McGrew, & Woodcock, 2003) contains several subtests that, when combined with portions of the WJ III, yield an Early Development General Intellectual Ability (EDev GIA) score. This measure is a six-cluster score, and includes a measure of each CHC factor, with the exception of fluid reasoning (*Gf*) (Tusing et al., 2003).

Administration of select tests from the WJ III and WJ III Diagnostic Supplement to a young child yields test scores within the areas of Comprehension-Knowledge (*Gc*), Long-Term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv*), Auditory Processing (*Ga*), Short-Term Memory (*Gsm*), and Processing Speed (*Gs*). An understanding of cognitive strengths and areas of difficulty among these abilities can be important in developing early intervention programs for young children (Ford, 2003).

The subtests included in the WJ III EDev GIA have many merits for use with young children. For one, the standardization of the WJ III allows for flexibility in the order of subtest administration. Therefore, assessment can begin with nonthreatening, engaging tasks to encourage children to have a pleasant testing experience. 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. In addition, the format of the WJ III subtest easily allows natural breaks to be taken during the assessment session, which is of particular importance when working with young children. The WJ III test materials contain colorful, child friendly stimulus items and several subtests utilize a tape recorder, which tends to further engage children in the test activities (Tusing et al., 2003). In addition, a chapter of the Examiner's Manual (Mather & Woodcock, 2001) and a chapter of the Diagnostic Supplement (Schrank et al., 2003) provide additional administration considerations for young children.

The WJ III also has some limitations in its use with young children. For one, the test kit does not contain any manipulatives to be used during testing. As such, it can be difficult to hold a young child's attention during testing. Additionally, WJ III 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.

While the WJ-R was identified as one of the strongest tools for use with young children according to Flanagan and Alfonso (1995), the WJ III Test of Cognitive Abilities and the WJ III Diagnostic Supplement have undergone substantial revisions. Tusing, Maricle, and Ford (2003) evaluated the WJ III according to Flanagan and Alfonso's (1995) criteria.

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.

Analysis of the internal consistency of the cluster scores of the WJ III COG indicates that scores range from inadequate to good across the preschool range. As such, information gathered by the current author as well as other researchers (Gelb & Alfonso, 2002) suggests that the WJ III COG does not function well as a stand alone assessment battery for children of all ability levels until the age of 4:0. Prior to this age range, several core subtests contributing to the EDev GIA do not demonstrate adequate internal consistency and / or adequate test floor until age 4:0. As a result, use of the Diagnostic Supplement, which contains more developmentally appropriate tasks, greatly improves

the test's adequacy. Further, the majority of the tests contributing to the EDev GIA demonstrate adequate floor by the age of 3:6. This allows for EDev GIA scores at least 2 standard deviations below average for children as young as 2:0 (Tusing et al., 2003).

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 administered the test within a span of two years. The remaining 69 children were tested within a period between 3 to 10 years (McGrew & Woodcock, 2001). 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, as findings are only reported for children between the ages of four and seven (Tusing et al., 2003).

#### *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 additional 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.*

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 months 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*, by K.S. McGrew and R.W. Woodcock, 2001, Itasca, IL: Riverside Publishing. GIA = General Intellectual Ability; 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 discriminant 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).

As the WJ III contains measures of specific abilities (*Gv*, *Gf*) not common in other instruments for young children, analysis of validity scores were restricted in some areas. 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 (McGrew & Woodcock, 2001). For example, the *Gv* cluster is most strongly related to the Performance scale of Wechsler assessment instruments.

Additionally, 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).

At this time, limited information is available regarding the concurrent validity of the Diagnostic Supplement of the WJ III with other cognitive instruments, further

confirming the importance of the current study. Information provided in the Diagnostic Supplement to the WJ III Test of Cognitive Abilities (Schrank et al., 2003) outlines a number of comparisons between the Diagnostic Supplement cluster scores and test scores from other cognitive batteries. Of these studies, two emphasized examining the use of the WJ III Diagnostic Supplement with young children.

One study consisted of a sample of young children ages 1 year, 9 months to 6 years, 3 months being administered several WJ III tests and the WPPSI-R and DAS. A second study involved administering all preschool appropriate subtests of the WJ III and Diagnostic Supplement and the SB: IV to a group of young children ages 3 years, 0 months to 5 years, 10 months (Schrank et al., 2003)..

Overall results of the studies indicated that the WJ III EDev GIA demonstrated strong correlations with overall ability scores of the DAS and WPPSI-R ( $r = .68$  and  $.72$ , respectively). According to Mather and Woodcock (2001) correlations of this magnitude are similar to those reported between global composite scores from major intelligence batteries, as well as those reported in the WJ III Technical Manual. Examination of the correlations of the expanded Diagnostic Supplement *Gv3* score indicated that this cluster correlated higher with other measures of visual or nonverbal processing, namely a  $r = .49$  correlation with the DAS Nonverbal Ability score in one preschool sample discussed in the technical manual.

Overall results obtained from studies conducted during the norming of the WJ III Diagnostic Supplement indicates that the Diagnostic tests and clusters possess the same levels of reliability and score precision as other WJ III tests and clusters. Diagnostic Supplement cluster scores were found to have strong correlations with composite scores

of other intelligence batteries. Further, narrow or specialized Diagnostic Supplement scores were found to correlate with similar measures in other batteries, yet still maintained uniqueness in the abilities they measure (Mather & Woodcock, 2001).

Many other cognitive ability instruments do not include composite scores of the CHC ability measured by specific WJ III Diagnostic Supplement clusters. In addition, the comparison composite scores of other batteries are typically mixed measures of CHC abilities. However, when factorially clean and valid measures of a CHC ability were available, the related Diagnostic Supplement clusters were observed to have high correlations with external criteria (Schrank et al., 2003). In terms of the studies involving young children, the *Visualization* cluster was found to correlate most highly with the DAS Nonverbal Ability composite ( $r = .65$ ) and the SB-IV Abstract/Visual Reasoning composite ( $r = .57$ ).

However, outside of the Diagnostic Supplement manual, no published literature is available regarding the relationship between this test and other measures of cognitive ability. In particular, no analysis of the use of the instrument with a special population, such as young children with language delays, is available. Further research in this area, such as that conducted in the current study, is needed to assist practitioners in making sound assessment decisions and in guiding test interpretation.

### *The Differential Ability Scales*

#### *Theoretical Background.*

Based upon its European predecessor, the British Ability Scales (BAS; Elliott, Murray, & Pearson, 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 from finding 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, which 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 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.

#### *Preschool Applications.*

The DAS cognitive battery is divided into two levels, one specifically designed for preschool-aged children. The two levels of the battery were deliberately designed to be developmentally appropriate and engaging for preschool age and school age children, respectively (Elliott, 1990b). The subtests included in the preschool range employ many manipulative objects and brightly colored pictures designed to appeal to young children and to hold their interest (Elliott, Daniel & Guiton, 1991). Additionally, a major focus in the development of DAS subtests was to ensure as far as possible that children being assessed understand what they are supposed to be doing. As such, most of the DAS subtests have demonstration and teaching items built into administration (Elliott, 1990b).

The DAS demonstrates additional strong qualitative characteristics for use with young children as well (Alfonso & Flanagan, 1999). 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 current study, as the children participating had speech and language delays and many were not 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). As previously discussed, the use of these basic concepts in subtest directions may inadvertently impact a young child’s score if they have difficulty understanding what is required of them during a particular task. Although the DAS does contain relatively complex subtest directions, it also offers the opportunity for a Special Nonverbal Composite score (SNC) to be calculated in place of the GCA. Directions for the subtest in the Special Nonverbal scale can be conveyed through gestures and the child’s responses require only pointing, drawing, or manipulation of objects. As such, this scale may be useful when working with shy preschoolers, children with delayed language or speech problems, elective mutes, children whose primary language is not English, or those with suspected hearing loss who are recommended for evaluation (Elliott et al., 1991).

Assessment by Flanagan and Alfonso (1999) suggested that the DAS appeared to strike a balance between good quantitative and qualitative characteristics, and as such had wide utility in the cognitive assessment of a variety of populations of young children. Specifically, Flanagan and Alfonso (1995; Alfonso & Flanagan, 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. According to Flanagan and Alfonso (1995), the standardization of the DAS is good and well representative of preschool age children.

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 measures what it intends to measure, 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 and colleagues (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 present study as it includes a special population of children receiving special education services, and therefore lower scores across subtests may emerge.

#### *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*, by C.D. Elliott, 1990, 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). 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, 1990b). 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 higher level cognitive 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, the Wechsler Preschool and Primary Scale of Intelligence – Third Edition (WPPSI-III; Wechsler, 2002), the Kaufmann Ability Scale for Children – Second Edition (KABC- II; Kaufman & Kaufman, 2003), the Stanford Binet Intelligence Test - Fifth Edition (SB5; Roid, 2003), 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). Being that the current study focuses on a sample of young children with speech and language delays, it is helpful to examine the relationship between the DAS and screening tools which contain measures of language and verbal ability.

Two studies conducted by McIntosh and colleagues (McIntosh, Wayland, et. al, 1995; McIntosh, Brown, et. al, 1995) examined the relationship between the DAS and the Bracken Basic Concepts Scale (BBCS; Bracken 1984). The BBCS provides a measure of receptive language, basic concepts, and school readiness for young children. One study (McIntosh, Wayland et al., 1995) utilized a sample of typically developing young children, while the second (McIntosh, Brown, et al., 1995) focused on a group of children identified as lacking basic school readiness skills and labeled “at-risk.” In both studies, a strong level of correlation was found between the DAS GCA score and the BBCS Total score, suggesting that the two measures assess similar skills (McIntosh, Wayland, et. al, 1995; McIntosh, Brown, et al., 1995).

McIntosh and colleagues (2000) also 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, further research related to the concurrent validity of the DAS with special populations is warranted.

### *Critical Analysis*

Controversy related to the validity of norm-referenced assessment with young children remains (Appl, 2000; Bagnato & Neisworth, 1994; Flanagan & Alfonso, 1995). 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. Following the review of the literature presented in this chapter, specific research needs are discussed below as related to the assessment of intelligence with young children and specifically, as related to the assessment of intelligence with young children with language delays.

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. Further, recent publication of the revised version of the WJ-R, the WJ III and Diagnostic Supplement to the Tests of Cognitive Abilities (Woodcock, McGrew, Mather, & Schrank, 2003) 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.

Finally, an important facet of the concern related to the utility of standardized assessment instruments with young children 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).

Difficulty understanding test expectations due to the excessive linguistic demand of an instrument may result in poor performance and thus impact a child's overall score. Doing so impacts the validity of a score in providing an accurate picture of true ability (Flanagan et al., 1995). Because overall test scores are often utilized to determine program placement and eligibility (Bagnato & Neisworth, 1994), further research related to the understanding of the linguistic demand of assessment instruments is necessary. Of particular importance is an examination of the use of assessment tools with children who have been identified as having speech and language delays, as difficulty in this area may greatly impact test performance and thus eligibility decisions.

## Chapter III

### *Methodology*

The purpose of this chapter is to outline a research study based upon key issues highlighted in the preceding review of the literature. Specific research questions addressed in the study are defined. In addition, information about the participants, procedures, instrumentation, and data analyses conducted in the study are outlined.

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). This study examined the relationship between the WJ III and DAS 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 was examined in order to determine the level of concurrent validity between each test battery. The following specific questions were examined:

1. What is the relationship between the Early Development General Intellectual Ability (EDev GIA) 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 test battery for a special population?
2. What is the relationship between the Comprehension-Knowledge (*Gc*), Visual-Spatial Thinking (*Gv* / *Gv3*), Auditory Processing (*Ga*), and Long-Term Retrieval (*Glr*), cluster 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 cluster scores of the WJ III and the diagnostic subtests of the DAS for a special population of young children?

### *Participants*

Children aged 3 years, 6 months (3:6) to 5 years, 11 (5:11) months were targeted for participation in the study. This age range was 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. In addition, this age range coincides with the Upper Preschool Level of the DAS test battery. The children targeted for inclusion in this study were receiving early childhood special education services for speech and language delays as defined by the eligibility criteria for the states of Minnesota and Wisconsin. All children who participated in the study were receiving special education services in a classroom setting, typically three to four days a week, for half a day. Approximately half of the children in the sample were in classrooms designed for children identified as having delays exclusively within the area of speech and language, including articulation and expressive and receptive language delays. The other half of the sample were in classrooms designed for children identified as having overall developmental delays. All of these children who participated in the study were receiving direct support within the area of speech and language within this classroom setting.

The sample consisted of 31 children, including 18 males and 13 females. However, one female participant was removed from the analysis due to incomplete data. This child was unable to complete the second test administered, which resulted in an invalid profile.

Therefore, all subsequent analyses were performed with the final sample of thirty children.

The mean age of participants was 4 years, 6 months ( $M = 54.53$  months,  $SD = 7.82$  months). Five children were between the ages of 3:6-3:11, 11 between 4:0-4:5, five between 4:6-4:11, six between 5:0-5:5, and three between 5:6-5:11. Seventy-three percent of the participants were Caucasian, 7% African-American, 10% Hispanic, 7% Asian and 3% indicated an "other" ethnic descent. Data related to parent education levels was also collected; however information in this area was not obtained for all participants. Available data indicated that the majority of the parents / guardians had some education beyond high school. A summary of the sample demographics in this study is included in Table 3.1.

Table 3.1

*Demographics of the sample (n = 30)*

| Demographic      | n  | Percentage |
|------------------|----|------------|
| Sex:             |    |            |
| Male             | 18 | 60         |
| Female           | 12 | 40         |
| Age:             |    |            |
| 3:6-3:11         | 5  | 17         |
| 4:0-4:5          | 11 | 37         |
| 4:6-4:11         | 5  | 17         |
| 5:0-5:5          | 6  | 20         |
| 5:6-5:11         | 3  | 10         |
| Race:            |    |            |
| Caucasian        | 22 | 73         |
| African-American | 2  | 7          |
| Hispanic         | 3  | 10         |
| Asian            | 2  | 7          |
| Other race       | 1  | 3          |

Table 3.1

*Demographics of the sample (n = 30; cont.)*

| Demographic                                  | n  | Percentage |
|--|----|------------|
| <b>Mother's level of education (n = 29):</b> |    |            |
| High school / GED                            | 10 | 33         |
| Some college                                 | 10 | 33         |
| Technical school                             | 2  | 7          |
| Bachelor's degree                            | 3  | 10         |
| Graduate school                              | 4  | 13         |
| <b>Father's level of education (n = 27):</b> |    |            |
| Less than high school                        | 2  | 7          |
| High school / GED                            | 7  | 23         |
| Some college                                 | 6  | 20         |
| Technical school                             | 4  | 13         |
| Bachelor's degree                            | 2  | 7          |
| Graduate school                              | 6  | 20         |

*Procedures*

Two school districts in the Minneapolis / St. Paul area of Minnesota were contacted for permission to solicit subjects from their early childhood special education programs. Permission to recruit subjects was also obtained from teachers of each program

approached. Once permission from program directors was granted, a brief description of the study and a letter of permission was sent to the parents/guardians of children who were eligible to participate. Upon returning consent forms, parents interested in having their child participate were contacted by the principal investigator regarding the specific procedures involved in the study. Several other children who fit the parameters of the study and were referred to the author by colleagues were also included in the sample. Parents signed statements of informed consent indicating that they understood the procedures of the study and aspects of their child's participation.

After parent permission was obtained, each child was administered the DAS and WJ III following standardized testing procedures. Tests were administered by licensed school psychologists. The majority of children were tested in private rooms, within their school building or at the University of Wisconsin-Stout. Some children were assessed in private rooms within the child's home. Test batteries were administered in counterbalanced order to avoid practice effects. Each child was assigned a code number when tested in order to maintain confidentiality. Only the researcher and her faculty advisor had access to the demographic data associated with each code number. Each child received a small toy as a reward for participation. Parents had the option of receiving a brief summary of their child's test performance upon request.

### *Instrumentation*

*The Woodcock Johnson Third Edition Tests of Cognitive Abilities – Early Development Scales.*

The Woodcock Johnson Third Edition Tests of Cognitive Abilities (WJ III; Woodcock et al., 2001) is the cognitive component of the Woodcock Johnson

Psychoeducational Battery – Third Edition. This battery was designed to assess cognitive abilities of individuals aged two to adult. Although the WJ III contains several tests that are designed for use with young children, it does not, by itself, provide an overall GIA score specifically for young children as the WJ-R did (Tusing et al., 2003). The recently published Diagnostic Supplement to the WJ III (Woodcock et al., 2003) expands the usefulness of this assessment tool with young children, by providing three additional tests for preschool-age children. When combined with tests from the WJ III, an Early Development General Intellectual Ability score (EDev GIA) is available. The six-test cluster provides one measure of each of the CHC broad factors, with the exception of Fluid Reasoning (*Gf*). Additional WJ III subtests can be administered to young children, based upon age parameters outlined in the test manual, in order to gather further information about specific cognitive abilities.

In the present study, the six tests required for the EDev GIA were administered to all participants. In addition, several other tests from the WJ III COG were administered in order to obtain other cluster scores. Test selection was determined based upon the subtest floor and internal consistency of each measure, as discussed in the Tusing, Maricle, and Ford (2003) book chapter outlined in Chapter II of this paper. The tests of the WJ III used in the present study are described in Table 3.2.

Table 3.2

*WJ III Tests*

| Test/Cluster Score  | Description   |
|---|---|
| Standard Battery:   |   |
| Verbal Comprehension<br>GIA-EDev, <i>Gc</i>                 | Measure of acquired knowledge, requires examinees to identify synonyms, antonyms, and complete verbal analogies   |
| Incomplete Words<br>GIA-EDev, <i>Ga</i>                     | 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        |
| Visual Matching<br>GIA-EDev, <i>Gs</i>                      | Measures processing speed, examinees are required to match two identical numbers in a row of six numbers  |
| Visual-Auditory Learning<br><i>Glr</i> , Associative Memory | 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 |

Table 3.2

*WJ III Tests (cont.)*

| Test/Cluster Score                   | Description  |
|--------------------------------------|--|
| Spatial Relations<br><i>Gv</i>       | Measures spatial ability, examinees must identify pieces that form a complete shape  |
| Sound Blending<br>Phonemic Awareness | Assesses the ability to synthesize phonemes, examinees listen to an audio recording and, blend sounds presented into a word                                    |
| General Information<br><i>Gc</i>     | Measures depth of general knowledge, examinees must orally identify where objects are found and what you typically do with objects                             |
| Retrieval Fluency<br><i>Glr</i>      | 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 |

Table 3.2

*WJ III Tests (cont.)*

| Test/Cluster Score  | Description  |
|---|--|
| Picture Recognition<br><i>Gv</i>                                | Assesses visual memory, the examinee must identify a set of previously presented pictures within a group of distracters  |
| Diagnostic Supplement:  |  |
| Memory for Names<br><i>GIA-EDev, Glr,</i><br>Associative Memory | Measures long-term retrieval, examinees are shown pictures of spaces creatures and told their names, they then must point to the appropriate picture when prompted |
| Visual Closure<br><i>GIA-EDev, Gv</i>                           | Measures, examinees must use visual closure abilities to identify increasingly abstract pictures   |
| Memory for Sentences<br><i>GIA-EDev, Gsm</i>                    | Measures short-term memory, examinees must repeat back increasingly complex phrases and sentences  |

*Note.* Adapted from *Woodcock Johnson III Tests of Cognitive Abilities Examiner's Manual*, by N. Mather & R.W. Woodcock, 2001. Itasca, IL: Riverside Publishing and *Woodcock Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities* by F.A. Schrank et. al., 2003, Itasca, IL: Riverside Publishing.

Administration of the tests outlined in Table 3.1 to children between the ages of 3:6 and 5:11 yields an overall cognitive ability score, the Early Development General Intellectual Ability (EDev GIA) score, as well as various cluster scores to measure specific cognitive abilities. The overall EDev GIA is derived from six test scores: Verbal Comprehension, Incomplete Words, Visual Matching, Memory for Names, Visual Closure, and Memory for Sentences. Cluster scores are derived from a combination of tests described below.

The Verbal Comprehension and General Information tests contribute to the Comprehension-Knowledge (*Gc*) cluster score. This primarily language-based measure indicates the breadth and depth of one's acquired knowledge, the ability to communicate knowledge, and the ability to reason using previously learned information. The Long-Term Retrieval (*Glr*) cluster consists of the Memory for Names, Visual-Auditory Learning, and Retrieval Fluency tests. This cluster provides a measure of one's ability to store information and fluently retrieve it later in the process of thinking. The visual-spatial (*Gv / Gv3*) clusters are derived from the Visual Closure, Spatial Relations, and Picture Recognition tests and provide a measure of one's ability to perceive, analyze, synthesize, and think about and remember visual patterns. The Memory for Sentences test provides a measure of short-term memory (*Gsm*), which is the ability to take in and hold information in immediate awareness and then use it within a few seconds (Mather & Woodcock, 2001; Schrank et al., 2003).

In addition to the cluster scores based upon CHC theory, portions of the WJ III provide information within other areas pertinent to thinking and reasoning. The Sound Blending test provides a measure of Phonemic Awareness, which is knowledge and skills

related to analyzing and synthesizing speech sounds. The Associative Memory cluster, derived from the Visual-Auditory Learning and Memory for Names tests, provides additional information about the ability to store and retrieve associations (Mather & Woodcock, 2001; Schrank et al., 2003).

*The Differential Ability Scales.*

The Differential Ability Scales (DAS) is an individually administered battery of cognitive and academic achievement tests for children and adolescents between the ages of 2:6 and 17:11. 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 six core subtests are those most strongly related to “g” and are used to compute the General Conceptual Ability (GCA) composite score. All core subtests weigh equally into the GCA score. The Upper Preschool Level 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 (Elliott, 1990b).

In the present study all participants were administered the core subtests of the Upper Preschool Battery of the DAS. In addition, all diagnostic subtests appropriate for each subject, based upon chronological age, were also administered. The subtests of the Upper Preschool Battery of the DAS are summarized in Table 3.3.

Table 3.3

*DAS Upper Preschool Level Battery Subtests*

| Subtest / Age-range / Cluster                             | Description  |
|---|--|
| Core Subtests:  |  |
| Copying – 3:6-5:11<br>GCA, Nonverbal Ability              | Measures visual-perceptual matching and fine motor coordination, examinees must reproduce line drawings                                  |
| Early Number Concepts - 3:6-5:11<br>GCA                   | Measures knowledge of pre-numeric and numerical concepts, examinees must answer questions by using colored chips or pointing to drawings |
| Naming Vocabulary - 2:6-5:11*<br>GCA, Verbal Ability      | Measure of expressive language and knowledge of vocabulary, examinees are shown an object or picture and asked to name it                |
| Pattern Construction - 3:6-5:11<br>GCA, Nonverbal Ability | Measure of nonverbal reasoning and spatial visualization, examinees must reproduce visually presented designs using colored blocks       |

Table 3.3

*DAS Upper Preschool Level Battery Subtests (cont.)*

| Subtest / Age-range / Cluster                              | Description  |
|--|--|
| Verbal Comprehension - 2:6-5:11*<br>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               |
| Picture Similarities - 2:6-5:11*<br>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                         |
| Diagnostic Subtests:                                       |  |
| Matching Letter-Like Forms 4:6-5:11                        | Measures visual discrimination ability, examinees must find the identical match to a visually presented 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  |

Table 3.3

*DAS Upper Preschool Level Battery Subtests(cont.)*

| Subtest / Age-range / Cluster       | Description   |
|-------------------------------------|---|
| Recall of Objects - 4:0-5:11        | Measures short-term and intermediate-term recall of verbal and pictorial information, examinees are visually and orally presented 20 pictures with names for a specified amount of time, and must orally 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*, by C.D. Elliott, 1990, San Antonio, TX: Psychological Corporation. GCA = General Conceptual Ability.

\*indicates tests of the Lower Preschool Battery.

Elliott (1997) defines the Verbal cluster score as a measure of verbal ability, both expressive and receptive, which provides a good measure of complex processing in the auditory-verbal information processing system. In the Upper Preschool Level of the DAS, the Naming Vocabulary and Verbal Comprehension subtests contribute to this cluster. The Nonverbal cluster is defined as a measure of fluid intelligence, as well as a measure of the integrative functioning between the two information processing systems

associated with visual and auditory modalities. This cluster is determined by the Copying, Pattern Construction, and Picture Similarities subtests of the DAS. The GCA score is a combination of all the subtests included in the Verbal and Nonverbal Ability clusters, as well as the Early Number Concepts subtest. The DAS diagnostic subtests provide measures of other specific abilities including visual and auditory memory and visual discrimination.

### *Analyses*

Pearson product-moment correlations were calculated to examine the relationship between the broad and cluster scores of each test battery and between DAS diagnostic subtests and select WJ III cluster scores. Means, standard deviations, and range of scores were calculated for the broad scores of each test battery to determine if the performance outcomes of each test battery differed significantly.

## Chapter IV

### *Results*

This study examined the concurrent validity between the Woodcock Johnson III Tests of Cognitive Ability (WJ III; Woodcock, McGrew, & Mather, 2001; Woodcock, McGrew, Mather, & Schrank, 2003), and the Differential Abilities Scales (DAS; Elliott, 1990a). This chapter presents the empirical results for the examination of the data with regard to the three research questions detailed in Chapter III.

Mean scores, standard deviations, Pearson product-moment correlation coefficients, and t-test techniques were utilized to examine the overall performance of participants in the sample and to examine the relationships between the DAS and WJ III. Descriptive analyses are discussed first in this chapter, including distributions of scores in terms of means, standard deviations, and skewness values. Next, results from correlation analyses are summarized according to each respective research question. Further discussion and implications of the findings, including implications for future research, are presented in Chapter V.

#### *Descriptive Analyses*

A summary of means, standard deviations, and skewness values for the WJ III and DAS subtest and cluster scores are reported in Table 4.1 and Table 4.2, respectively. Skewness values of the overall ability scores indicate that the sample approximated a normal distribution. The mean WJ III cluster scores were generally within the average range. However, the mean EDev GIA score for a sample of young children receiving early childhood special education services was in the low average range ( $M = 85.2$ ). The standard deviation for the EDev GIA score ( $SD = 18.10$ ) was somewhat higher than that

which is reported in the Examiner's Manual (Mather & Woodcock, 2001), indicating that there was greater variability in subjects' EDev GIA scores for this sample than was found in the WJ III standardization sample. Standard deviations of the cluster scores of the WJ III were comparable to that reported in the Technical Manual (McGrew & Woodcock, 2001). An exception was the Phonemic Awareness cluster (SD = 7.64), which was somewhat restricted. Similar to the WJ III GIA EDev, the mean DAS GCA score also fell within the low average range (M = 89.17; SD = 14.06). The mean value for the Verbal Ability cluster was also within in the low average range (M = 86.57; SD = 16.05); whereas the Nonverbal Reasoning Ability cluster was in the average range (M = 92.8; SD = 14.92).

Table 4.1

*Means, Standard Deviations, and Skewness Values for the WJ III Broad and Cluster Scores (n = 30)*

| Test / Cluster   | M     | SD    | Skewness |
|--|-------|-------|----------|
| WJ III Early Development – General Intellectual Ability (EDev GIA) | 85.20 | 18.10 | -.189    |
| Verbal Ability – Extended (VA ext)                                 | 92.63 | 14.30 | .093     |
| Comprehension-Knowledge (Gc)                                       | 92.63 | 14.30 | .093     |
| Long-Term Retrieval (Glr)  | 90.46 | 13.68 | .336     |
| Visual-Spatial Thinking (Gv)                                       | 89.52 | 14.94 | .606     |
| Visual-Spatial Thinking 3 (Gv3)                                    | 95.04 | 14.02 | .574     |
| Phonemic Awareness   | 96.84 | 7.64  | .632     |
| Associative Memory   | 95.25 | 12.50 | -.671    |

Table 4.2

*Means, Standard Deviations, and Skewness Values for the DAS Broad and Cluster Scores (n = 30)*

| Test / Cluster                     | M     | SD    | Skewness |
|------------------------------------|-------|-------|----------|
| General Conceptual Ability (GCA)   | 89.17 | 14.06 | .631     |
| Verbal Ability                     | 86.57 | 16.05 | -.067    |
| Nonverbal Reasoning Ability        | 92.08 | 14.92 | -.175    |
| <i>Diagnostic Subtests</i>         |       |       |          |
| Recall of Digits (n=28)            | 42.08 | 14.33 | 2.713    |
| Recall of Objects-Immediate (n=22) | 41.55 | 10.85 | .431     |
| Recall of Objects-Delayed (n=22)   | 42.55 | 9.30  | 1.114    |
| Recognition of Pictures (n=28)     | 46.54 | 9.22  | .621     |
| Matching Letter-Like Forms (n=15)  | 47.07 | 12.56 | .882     |

### *Research Question One*

*What is the relationship between the Early Development General Intellectual Ability (EDev GIA) score 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?*

Pearson product-moment correlation coefficients were calculated between the broad ability scores of the WJ III and the DAS. The correlation between the WJ III EDev GIA and the DAS GCA scores ( $r = .73$ ) was significant and suggests a strong relationship exists between the overall scores from each test. The correlation size suggests that 53% of the variance for both measures is shared. Comparisons of the overall mean scores of the WJ III EDev GIA ( $M = 85.20$ ) and the DAS GCA ( $M = 89.17$ ) indicated that, on average, participants in the sample scored 3.97 points higher on the DAS. A two tailed pair-wise t-test between the overall scores of the WJ III and the DAS indicated that the difference in the mean scores was not statistically significant ( $t = 1.76, p > .05$ ). A similar comparison between the DAS Verbal Ability cluster and the WJ III *Gc* cluster indicated that the sample's overall performance on each cluster was statistically significantly different ( $t = -3.213, p < .01$ ).

### *Research Question Two*

*What is the relationship between the Comprehension Knowledge (Gc), Long-Term Retrieval (Glr), Visual Spatial Thinking (Gv / Gv3), Phonemic Awareness, and Associative Memory 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?*

Pearson product-moment correlation coefficients were calculated between the cluster scores of the WJ III and the cluster scores of the DAS. With the exception of the relationship between the WJ III Phonological Awareness Cluster and DAS Verbal Ability cluster, and the WJ III Associative Memory Cluster and the DAS Nonverbal Ability Cluster, all correlations between scores from both batteries were significant at the  $p < .05$  level. For a better understanding of the relationship between the two tests, correlation findings for the study are discussed in terms of the size of the correlation coefficients and shared variance between the two measures, rather than in significance values alone. Correlation coefficients between the cluster scores of the WJ III and the cluster scores and diagnostic subtests of the DAS are reported in Table 4.3. Subtest correlation matrices are found in Appendix C.

Table 4.3

*Correlations Between Cluster Scores of the WJ III and Cluster Scores and Diagnostic Subtests of the DAS (n = 30)*

| DAS                         | WJ III |       |       |       |      |       |
|-----------------------------|--------|-------|-------|-------|------|-------|
|                             | Gc     | Glr   | Gv    | Gv3   | PhAw | AsM   |
| Verbal Ability              | .77**  | .59** | .64** | .66** | .36* | .49** |
| Nonverbal Reasoning Ability | .38**  | .49   | .78** | .74** | .42* | .24   |
| Recall of Digits            | .25    | .09   | .36   | .37   | .39  | -.12  |
| Recall of Objects-Immediate | .79**  | .68** | .52*  | .60** | .28  | .61** |
| Recall of Objects-Delayed   | .27    | .35   | .29   | .44*  | -.01 | .33   |
| Recognition of Pictures     | .58**  | .50*  | .62** | .56** | .35  | .22   |
| Matching Letter-Like Forms  | .49    | .44   | .62*  | .58*  | .47  | .47   |

\*\* Correlation is significant at the .01 level (2-tailed test)

\* Correlation is significant at the .05 level (2-tailed test)

A comparison of the mean scores of the DAS Verbal Ability composite and the WJ III *Gc* cluster indicates that participants scored significantly higher on the WJ III. The mean Verbal Ability score was in the low average range ( $M = 86.57$ ), while the mean *Gc* score fell within the average range ( $M = 92.63$ ). A two tailed pair-wise t-test between the scores indicated that the difference in the mean scores was statistically significant ( $t = -3.21, p < .01$ ). The Verbal Ability cluster of the DAS and the *Gc* cluster of the WJ III were the most strongly related ( $r = .77$ ; 59% shared variance). Moderate correlations also existed between the DAS Verbal Ability composite and the *Glr*, *Gv*, *Gv3*, and Associative Memory cluster scores of the WJ III. A weak correlation was found between the DAS Verbal Ability score and the WJ III Phonemic Awareness cluster.

The Nonverbal Reasoning Ability cluster of the DAS demonstrated a similarly strong relationship to the *Gv* and *Gv3* cluster scores of the WJ III ( $r = .78$ ; 61% shared variance and  $r = .74$ ; 55% shared variance, respectively). In contrast, weak correlations were found between the DAS Nonverbal Reasoning Ability cluster and the *Gc* and Associative Memory clusters of the WJ III.

### *Research Question Three*

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

In comparison with cluster score correlations, the DAS diagnostic subtests did not evidence convergent and discriminant validity as distinctly. Additionally, the amount of shared variance between scores was generally smaller than that evidenced between cluster scores. The DAS Recall of Objects-Immediate subtest evidenced moderate to

strong correlations with several WJ III clusters, including *Gc*, *Glr*, *Gv*, *Gv3*, and Associative Memory. However, the DAS Recall of Objects-Delayed subtest did not correlate strongly with any of the WJ III clusters. Its highest correlation was a moderate correlation with the *Gv3* cluster, and this evidenced only 19% shared variance ( $r = .44$ ). The Recognition of Pictures subtest demonstrated moderate correlations with the *Gc*, *Glr*, *Gv*, and *Gv3* clusters. Weak correlations were found between this subtest and the Phonemic Awareness and Associative Memory clusters of the WJ III. The Matching Letter-Like Forms subtest of the DAS correlated moderately with all clusters of the WJ III, with its strongest correlation with the *Gv* cluster ( $r = .62$ ; 38% shared variance). Conversely, the DAS Recall of Digits diagnostic subtest did not demonstrate significant correlation with any of the cluster scores of the WJ III.

## Chapter V

### *Discussion*

Prior to the current study, no published research outside the WJ III Diagnostic Supplement manual (Schrank, Mather, McGrew, & Woodcock, 2003) has examined the relationship between the Woodcock Johnson III Tests of Cognitive Abilities (WJ III; Woodcock, McGrew, & Mather, 2001) Early Development GIA score and the Differential Abilities Scales (DAS; Elliott, 1990a) for use with young children. Further, no information is available regarding the use of these assessment instruments with a special population of young children, specifically those receiving early childhood special education services for speech and language delays. As such, further research related to the use of the WJ III and DAS with young children with speech and language delays was necessary. The present study adds to the body of knowledge available on the relationship between the WJ III and other cognitive assessment instruments. It also provides pertinent information related to the utility of standardized intellectual assessment instruments with young children with speech and language delays.

The purpose of this study was to examine the concurrent validity of the WJ III and WJ III Diagnostic Supplement, and the DAS when used with a special population of young children. Specifically, the study examined the concurrent validity between the broad ability scores of the WJ III and the DAS, as well as the relationship between the CHC clusters of the WJ III Early Development Scale with the composite and diagnostic subtest scores of the DAS. The following chapter discusses the findings presented in Chapter IV with regard to each research question. In addition, limitations of the present study and implications for future research are presented.

### *Summary*

Thirty children from the Twin Cities metropolitan area of Minnesota and Western Wisconsin participated in this study. Children who participated were between the ages of 3 years, 6 months and 5 years, 11 months at the time they were administered the WJ III and DAS. Children involved in the study came from five different ethnic backgrounds, although the majority of the participants were of Caucasian descent. Parent education levels of the participants ranged from less than a high school education to advanced graduate degrees. Overall, approximately an equal number of males and females participated in the study.

Participants were administered the WJ III tests contributing to the EDev GIA as well as those pertaining to other CHC clusters of interest, and the DAS core and diagnostic subtests of the Upper Preschool Level. Children were administered both tests batteries within a two-week time frame in counterbalanced order to avoid practice effects. Correlation coefficients of the broad scores of each test were examined to determine the similarities across abilities measured by both test batteries. Mean overall scores were compared to determine if participants obtained similar overall scores on each test battery. Correlation coefficients of the CHC cluster scores of the WJ III and the cluster and diagnostic subtests of the DAS were also compared in order to determine whether similar abilities are being measured with respective cluster areas.

### *Research Question One*

Overall mean scores obtained by participants, all young children receiving early childhood special education services for speech and language delays, on both the WJ III and DAS were in the low average range (WJ III EDev GIA  $M = 85.2$ , DAS GCA  $M =$

89.17). Although participants scored slightly higher on the DAS, the difference between the mean scores on the test batteries was not statistically significant. Thus, when the WJ III and WJ III Diagnostic Supplement are used to derive an overall Early Development GIA score, the test battery appears to yield a measurement of intelligence that is similar to the Differential Abilities Scale. This includes samples of children identified as having speech and language delays.

Previous research with typically developing children provides support for the General Intellectual Ability (GIA) score of the WJ III and the General Conceptual Ability (GCA) score of the DAS as measures of general intelligence, as they have strong positive correlations with broad ability scores from other commonly used intelligence test batteries (McGrew & Woodcock, 2001). Further, research comparing the DAS and the previous version of the Woodcock Johnson Early Development Scale suggest that the two batteries have moderate correlations at the overall ability score level (Tusing, 1998).

Results of the present study are consistent with previous research indicating a positive relationship between the WJ III GIA Early Development score and the DAS GCA score. The overall correlation of the broad ability scores in this study ( $r = .73$ ) suggests 66% shared variance across the two tests. These results extend findings reported in the technical manual and indicate that when the GIA Early Development score is used with young children, it too provides an overall ability score that is strongly related to the DAS. Given that significant data exists to support the relationship between the DAS Upper Preschool Level and other measures of cognitive ability (Elliott, 1990a; Wechsler, 2002) this is a positive finding for the WJ III GIA Early Development score in its use with young children.

A unique difference exists between the current WJ III Early Development Scale and previous WJ-R, which further supports the contention that the WJ III is an improvement over previous versions of the Woodcock Johnson in its use with young children. Specifically, the amount of shared variance found in this study is larger than that discovered in similar research comparing the DAS and WJ-R Early Development Scale (Tusing, 1998). The difference in the size of the correlation may be attributed to the fact that the WJ III EDev GIA is derived differently than previous overall ability scores on the WJ for young children. The EDev GIA score is obtained from six test scores, each weighted as determined by their overall contribution to *g*; whereas, the WJ R Early Development Scale overall score was derived from an average of five tests.

Additionally, substantial revisions were made to some of the subtests contributing to the GIA score. For one, the Verbal Comprehension subtest, which contributes to the *Gc* cluster as well as strongly to the EDev GIA score, was revised to include more reasoning tasks involving synonyms, antonyms, and analogies in addition to the previous picture naming tasks. Further, the revisions of the WJ III have provided a means for deriving a Short-Term Memory (*Gsm*) and Auditory Memory Span cluster, through the administration of the Memory for Sentences and Memory for Words subtests. These clusters are particularly useful for young children in that they provide a measure of auditory or phonological, memory ability, which may provide insight into early reading skills (Tusing et al., 2003).

#### *Research Question Two*

Results of this study showed a strong relationship between the *Gc* cluster of the WJ III and the Verbal Ability cluster of the DAS, which was expected given the nature of

the abilities measured by each cluster. Similarly, a strong relationship was found in the relationship between the WJ III *Gv* / *Gv3* clusters and the Nonverbal Ability cluster of the DAS, which was also consistent with expectations. However, analyses of the differences between cluster scores on both test batteries indicated statistically significant differences in performance across the sample in some areas. For example, the Verbal Ability and Nonverbal Ability clusters of the DAS were statistically significantly different ( $t = 2.233$ ,  $p < .05$ ), with the mean Nonverbal Ability cluster within the average range ( $M = 92.8$ ) and the mean Verbal Ability cluster in the low average range ( $M = 85.67$ ). No statistically significant differences were found between the WJ III *Gc* cluster and any other WJ III cluster score. Instead, the mean scores of all WJ III clusters were fairly similar, with the exception of the *Gv* and *Gv3* clusters, with the mean *Gv3* cluster score being significantly higher than the *Gv* score. Implications for interpretation of the verbal and nonverbal ability domains for each test battery are discussed below.

#### *Verbal Ability.*

Concurrent validity for the WJ III *Gc* cluster was established by a substantial amount of overlap ( $r = .77$ ; 59% shared variability) with the Verbal Ability cluster of the DAS. In fact, the *Gc*/Verbal Ability relationship was one of the strongest across cluster scores from both test batteries. These findings are similar to those reported by McGrew and Woodcock (2001). The size of this relationship suggests that each test provides a similar measure for abilities related to crystallized intelligence, which is defined by Carroll (1993) as including the depth and breadth of acquired knowledge, the ability to communicate knowledge verbally, and the ability to reason using previously learned experiences or procedures.

Despite the strong positive correlation, a statistically significant difference existed across the overall performance of children on the DAS Verbal Ability cluster versus the WJ III *Gc* cluster, with children scoring significantly higher on the WJ III *Gc* measure. This difference warranted closer examination. Post hoc analyses of the participants' performance at the subtest level suggest that children scored significantly lower on the DAS receptive language task, Verbal Comprehension, than they did on the expressive language task, Naming Vocabulary. The overall mean performance on Verbal Comprehension was in the below average range. A similar difference did not exist between receptive and expressive language tasks on the WJ III, where participants scored in the average range on both tests. An understanding of these differences for the current sample may be found in an examination of the task requirements for subtests contributing to each cluster. The Verbal Ability cluster of the DAS is derived from two subtests, one a measure of expressive vocabulary (naming pictured objects/Naming Vocabulary) and the other a measure of receptive language (following verbal directions using manipulatives/Verbal Comprehension). The WJ III *Gc* cluster is derived from a test of expressive language (Verbal Compression) that includes measures of picture naming ability, word knowledge (synonyms/antonyms), word reasoning, and a task considered to be a measure of receptive language (General Information), which also assesses overall depth of general verbal knowledge. While the WJ III General Information test requires receptive language abilities in the form of understanding content-based questions (e.g., where do you usually find a window; what do you do with a toothbrush), it may actually function more as a measure of overall acquired knowledge. As such, the WJ III *Gc* tests may not be overly effective in distinguishing children with language delays. Further

investigation of the utility of this aspect of the tool for children with language delays is needed.

*Nonverbal Ability.*

Results of the study also support a strong relationship between the WJ III *Gv* / *Gv3* clusters and the Nonverbal Ability cluster of the DAS. The WJ III *Gv* cluster consists of a measure of visual-spatial thinking requiring the examinee to identify two to three pieces that form a complete target shape (Spatial Relations) as well as a task of visual memory for objects and pictures (Picture Recognition). The *Gv3* cluster is unique to the Early Development Scales of the WJ III. It includes both of the *Gv* subtests in addition to a task of visual-spatial thinking that consists of identifying a picture or a drawing that has been altered (Visual Closure). In comparison, the DAS Nonverbal Ability cluster consists of a task requiring the reproduction of line drawings (Copying), reproducing designs with colored blocks (Pattern Construction) and a nonverbal reasoning task (Picture Similarities). Despite the difference in the task requirements, the relationship between the nonverbal cluster scores of the WJ III EDev and DAS suggests that similar reasoning abilities, related to visual-spatial skills and less linked to language, are being assessed by each test.

The size of the relationships for these comparisons is similar to those reported by McGrew and Woodcock (2001). However, in the current study the WJ III *Gv* score is more strongly related to the DAS Nonverbal Ability cluster than found in previous research with typically developing preschool aged children. Further, the overall performance on the *Gv* cluster is somewhat lower than that achieved by young children in previous studies. This lower performance on nonverbal tasks could potentially be related

to the language requirements in the directions of the *Gv* subtests. In particular, the Spatial Relations subtest may have been impacted as the directions require an understanding of such concepts as “together,” “whole,” and “flip.”

In addition, post hoc examination of the nonverbal tasks of the WJ III indicates that performance was significantly stronger on the *Gv3* cluster, as compared to the *Gv* cluster. This appears to be attributed to performance on the Visual Closure subtest, which was the subtest with the highest mean performance on the WJ III ( $M = 102.9$ ). Visual Closure is a subtest specifically designed to be used with young children that was added to the test battery during the publication of the Diagnostic Supplement. The specific design and structure of this subtest may have contributed to the strong performance of participants; however, further research in this area is necessary before drawing any firm conclusions.

Although the DAS Nonverbal Ability cluster was most strongly related to the WJ III *Gv* / *Gv3* clusters, moderate correlations with other WJ III clusters, particularly those less verbal in nature, were also discovered. These relationships can be attributed to the fact that while the DAS Nonverbal cluster groups visual-spatial and reasoning tasks together, the WJ III attempts to provide individual measures of specific abilities. In general, the Nonverbal Ability score of the DAS and the *Gv* score of the WJ III evidenced discriminant validity with dissimilar measures. Specifically, although still within the moderate range, the DAS Nonverbal Ability score was less strongly related to the EDev GIA and *Gc* scores than had been previously discovered in studies with typical preschool aged children. Further, the *Gv* cluster was also less strongly related to Verbal Ability than in previous research (McGrew & Woodcock, 2001). Such relationships

suggest that this portion of each assessment tool may be of particular use with children with language delays, as they appear to assess unique abilities less tied to verbal skills.

### *Research Question Three*

It was expected that the diagnostic subtests of the DAS would correlate most strongly with the WJ III cluster scores to which they were most similar. That is, it was expected that the DAS diagnostic subtests related to memory (Recall of Objects, Recall of Objects – Delayed, Recall of Digits, and Recognition of Pictures) would evidence the strongest correlations with the WJ III Long-Term Retrieval (*Glr*) and Associative Memory clusters.

Not to be confused with long-term memory, the *Glr* cluster measures the processing efficiency through which information is initially stored and later recalled. As such, the DAS Recall of Objects and Recall of Objects-Delayed subtests were expected to correlate most strongly with this measure. In contrast, the Associative Memory cluster is a narrow ability of *Glr* related to the ability to form and retrieve associations. As the DAS does not specifically assess such abilities, moderate correlations were expected with all diagnostic subtests related to memory with this cluster. It should be noted that the WJ III test administered during this study did not provide a measure of short-term memory (*Gsm*) at the cluster level. As such, relationships related to short-term memory skills at the subtest level are highlighted and discussed in the following section.

Across all memory-related WJ III clusters, the Recall of Objects subtest demonstrated its strongest correlation with the *Glr* and Associative Memory clusters ( $r = .68$  and  $.61$  respectively). However, it is notable that this subtest also demonstrated a strong relationship with the WJ III *Gc* cluster ( $r = .79$ ) and a moderate relationship with

the *Gv3* cluster ( $r = .59$ ) suggesting that it related to a variety of CHC abilities. The results of this study are consistent with previous factor analytic research studies that have determined that Recall of Objects subtests have proved difficult subtests to classify according to CHC theory (McGrew & Flanagan, 1998; Elliott, 1990a, Tusing, 1998). Elliott (1990a) describes the Recall of Objects subtest as a measure of short-term verbal memory. As such, further research including the *Gsm* cluster is necessary to assist in differentiating the skills assessed by this measure.

In contrast, the Recall of Objects-Delayed subtest did not demonstrate strong relationships with any cluster scores, with its strongest relationship being with the *Glr* and *Gv3* clusters ( $r = .35, .44$ ). Recall of Objects-Delayed is intended to measure retrieval of information after an intervening task inhibits the immediate recall of the visually and orally presented stimuli. That being said, a stronger relationship with the *Glr* cluster was expected.

In addition to the hypothesis that the Recall of Objects and Recall of Objects-Delayed subtests assess multiple narrow CHC abilities, the lack of close alignment with any of the WJ III clusters examined in this study may also be explained by the nature of task presentation. The subtest directions for the Recall of Objects subtest requires an understanding of several basic concepts (ex. same, all, before, some, as many) that are best understood by children ages five and above (Flanagan et al., 1995). That knowledge, in addition to the fact that the children assessed had identified language delays, suggests that there is a strong possibility that participants did not have a clear understanding of the requirements of these subtests.

The Recognition of Pictures subtest demonstrated a weak correlation with the Associative Memory cluster. However, it was moderately correlated with several WJ III clusters, including *Gc*, *Glr*, and *Gv / Gv3*, suggesting that it is somewhat related to tasks assessing overall thinking skills and long-term retrieval. The WJ III subtest most similar to this task, Picture Recognition, does not contribute to the *Glr* or Associative Memory cluster. Rather, it is a component of *Gv / Gv3*, which assists in explaining the relationships discovered with these clusters. At the test level, Recognition of Pictures and Picture Recognition were found to be moderately correlated. The fact that this relationship is not stronger may be due to variations in the format of the subtests. Although each task requires the respondent to pick a previously viewed picture or object out of a group after having been exposed to it briefly, in general the DAS items have more distracter items present than similar items on the WJ III.

As expected, the Recall of Digits subtest did not correlate strongly with any WJ III clusters, as no WJ III clusters assessing short-term memory were derived for this sample. At the subtest level, Recall of Digits was most strongly correlated with the WJ III Memory for Sentences subtest, which is also a verbally presented short-term memory task. The lack of a *Gsm* cluster for comparison with similar tasks is a limitation of this study. This is an area in which further research is needed in order to aid in the understanding of memory tasks and related skills assessed by each test battery.

The DAS diagnostic subtest Matching Letter-Like Forms is designed to measure visual discrimination ability. As such, it was expected to correlate most strongly with the WJ III cluster assessing similar skills, *Gv / Gv3*. Results indicated that the Matching

Letter Like Forms subtest demonstrated its strongest relationships with these cluster scores, suggesting that they are assessing similar abilities.

Overall, the Phonemic Awareness and Associative Memory clusters of the WJ III did not correlate strongly with any DAS measures (with the exception of the Recall of Objects-Delayed subtest). This suggests that the WJ III provides measurement of unique abilities, specifically abilities important in the assessment of young children. Research has suggested that better performance on tasks of short-term verbal and phonologic memory are associated with higher reading performance, while deficits in short-term memory have been identified as precursors to reading difficulties (Tusing et al., 2003). In light of such findings, information provided by the WJ III in the areas of memory and phonological awareness may assist practitioners in program planning and early intervention for young children.

It is interesting to note that despite having identified speech / language delays, participants in the study did quite well of the Phonemic Awareness cluster, with scores within the average range on both contributing subtests, Sound Blending and Incomplete Words. This performance suggests that the children generally demonstrated adequate skills related to analyzing and synthesizing speech sounds. This information also assists in understanding the significantly stronger performance on the DAS expressive language task (Naming Vocabulary) as compared to the receptive language (Verbal Comprehension) task.

#### *Conclusions and Implications for Further Research*

The results of this study indicate that the WJ III and the DAS are measuring similar broad constructs from a theoretical standpoint. Further, the two measures yielded

fairly similar overall scores. For this sample, the overall mean scores obtained on both the WJ III and DAS were in the low average to below average range. Since the children in this study were all identified in special education as having language delays, it is likely that this impacted their performance on portions of the assessment.

It was discovered that a statistically significant difference existed across overall performance on the DAS Verbal Ability cluster versus the WJ III *Gc* cluster, with significantly higher performance on the WJ III tasks. Further, performance on the WJ III *Gc* tasks was generally similar, and within the average range, while performance on the DAS expressive language tasks was significantly stronger than performance on the receptive language task. These relationships suggest that the WJ III *Gc* tasks may be less impacted by receptive language skills and more a measure of overall fund of general knowledge.

Despite the fact that participants scored significantly lower on the DAS Verbal Ability cluster, overall differences between the broad ability scores of the two test batteries was not statistically significant. It is important to note that certain cluster scores on the WJ III contain subtests which do not contribute to the overall ability score, which is not the case on the DAS. Further, tests contributing to the EDev GIA are weighted according to their overall contribution to *g*. For children between the ages of three and six, the Memory for Sentences, Memory for Names, and Verbal Comprehension tasks weigh most heavily into the overall ability score. The increased sophistication of the WJ III in terms of test construction impacts the interpretive meaning of the test, as well as examining its relationship with other instruments. Further research examining cluster

scores with both language delayed and typically developing young children is needed in order to broaden understanding of the utility of this tool with such populations.

Additional research might also want to analyze the tests qualitatively to examine the format of subtest directions, the incidence of basic concepts, and the impact that language-based tasks have upon overall ability scores. When considering the WJ III and DAS, information related to the utility of the DAS Special Nonverbal Composite (SNC) score and its relationship to the overall ability score of other intelligence batteries is warranted. Further information in this area will assist practitioners in making sound assessment decisions and also provide insight into the role that standardized assessments should have in placement and programming decisions within the field of early childhood education.

The results of this study further indicate that not all of the DAS diagnostic subtests are strongly related to the corresponding CHC clusters of the WJ III included in this study. This suggests that the diagnostic subtests may be mixed measures of CHC abilities. Additional research in this area may provide further insight as to the nature of the abilities measured within these subtests, particularly when used with young children. Specifically, examination of the relationship between short-term memory (*Gsm*) and reasoning measures (*Gf*) measures with the diagnostic DAS subtests is warranted. This information would aid practitioners in determining whether these subtests provide useful information for academic planning.

#### *Limitations of the Study*

One limitation of this study is that children participating were from a narrow geographic region. Only children from the Twin Cities metropolitan area of Minnesota

and Western Wisconsin were recruited to participate. Similarly, although the sample consisted of children from a range of ethnic backgrounds, the majority of the participants were Caucasian. Therefore, study findings may not generalize to children from other demographic backgrounds. Further, only children aged 3 years, 6 months to 5 years, 11 months were included in the sample. Therefore, results of the study may not generalize to other age groups. A larger sample more representative of the overall population would provide results that may potentially generalize to a greater population of children.

In addition, children currently receiving early childhood special education services in the area of speech and language were targeted for inclusion in the study. With this population, the distribution of scores tended to be in the low average to average range of ability. Although the nature of the participants provides specific information about the use of the WJ III and DAS with children with speech and language delays, the results of the study may not generalize to children who have not been involved in special education within this area. A larger sample with greater representation of ability would assist in providing greater interpretation of results that are not limited by this.

The current sample consisted of 30 participants. However, due to the age of some children and missing data due to absences and other logistical factors, some cluster scores and diagnostic subtests were based upon a sample of less than 30, which is fewer than typically considered desirable for a study of this nature. Further research with a larger sample would provide more reliable, generalizable data related to the use of the WJ III and DAS with young children.

A final limitation of this study is that not all CHC factors were present in the analysis of the WJ III. This is partially due to the lack of a WJ III *Gf* cluster for young

children, but is also due to an oversight on the authors' part in not including all subtests contributing to the *Gsm* cluster. Further research with young children including these clusters would provide important information about such skills in young children, particularly in relation to early reading skills, as well as implications for assessment and program planning.

*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.
- American Education Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7<sup>th</sup> ed.). Upper Saddle River, NJ: Prentice Hall.
- Appl, D.J. (2000). Clarifying the preschool assessment process: Traditional practices and alternative approaches. *Early Childhood Education Journal*, 27(4), 219-225.
- 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<sup>th</sup> 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*. (2001). 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., Daniel, M.H., & Guiton, G.W. (1991). Preschool cognitive assessment with the Differential Ability Scales. In B.A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (pp.133-153). Boston: Allyn & Bacon.
- Elliott, C.D., Murray, D.J., & Pearson, L.S. (1979). *British Ability Scales*. Windsor, 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., Alfonso, V.C., Kaminer, T., & Rader, D.E. (1995). Incidence of basic concepts in the directions of new and recently revised American intelligence tests for preschool children. *School Psychology International*, 16, 345-364.

- 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.
- Ford, L. (2003). Assessment of young children. In F.A. Schrank, N. Mather, K.S. McGrew & R.W. Woodcock (Eds.), *Diagnostic supplement to the Tests of Cognitive Ability manual* (pp. 37-46). Itasca, IL: Riverside Publishing.
- Gelb, M. & Alfonso, V.C. (2002). *Utility of the Woodcock Johnson III tests of cognitive abilities with preschoolers with developmental delays*. Paper presented at the Train the Trainers Meeting, Itasca, IL.
- 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* (5<sup>th</sup> ed.). Boston: McGraw-Hill College.
- Horn, J.L., & Cattell, R.B. (1967). Age difference in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107-129.
- Individuals with Disabilities Education Act (IDEA). H.R. 5, Fed. Reg. 10-476. (1990, 1997, 2003).
- Kaufman, A.S., & Kaufman, N.L. (2003). *Kaufman Intelligence Battery for Children – Second Edition*. Circle Pines, MN: AGS Publishing
- Kazdin, A.E. (1992). *Research design in clinical psychology* (2<sup>nd</sup> ed.). Boston: Allyn & Bacon.
- Keith, T.Z. (1990). Confirmatory and hierarchical confirmatory analysis of the Differential Ability Scales. *Journal of Psychoeducational Assessment*, 8, 391-405.

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

- Nuttall, E.V., Romero, I., & Kalesnik, J. (Eds.). (1999). *Assessing and screening preschoolers: Psychological and educational dimensions*. 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.
- Roid, G.H. (2003). *Stanford Binet Intelligence Scale – Fifth Edition*. Itasca, IL: Riverside Publishing
- Romero, I. (1999). Individual assessment procedures with preschool children. In E.V. Nuttall, 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., Mather, N., McGrew, K.S., & Woodcock, R.W. (2003). Manual. *Woodcock-Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.
- Thorndike, R.L., Hagen, E.P., & Sattler, J.S. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition*. Itasca, IL: Riverside Publishing.

- Tusing, M.E. (1998). *Validation studies with the Woodcock-Johnson Psycho-Educational Battery – Revised Tests of Cognitive Ability: Early Development Scales and the Differential Abilities Scale: Preschool Level*. Unpublished doctoral dissertation, University of South Carolina.
- Tusing, M.E., Maricle, D.E., & Ford, L. (2003). Assessment with the Woodcock-Johnson III and young children. In F.A. Schrank & D. Flanagan (Eds.). *WJ III clinical use and interpretation*, (pp. 243-283). New York: Academic Press.
- Wechsler, D. (1989). *Wechsler Preschool and Primary Intelligence Scale- Revised*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2002). *Wechsler Preschool and Primary Intelligence Scale – 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.
- Woodcock, R.W., McGrew, K.S., Mather, N., Schrank, F.A. (2003). *Woodcock-Johnson III Diagnostic Supplement to the 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

**Please complete and return with permission form**

**WJ-III /DAS 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

## Appendix C

## Subtest Correlation Matrices

| WJ III                           |                |                 |                      |                   |                    |
|----------------------------------|----------------|-----------------|----------------------|-------------------|--------------------|
| DAS                              | Verbal<br>Comp | V-A<br>Learning | Spatial<br>Relations | Sound<br>Blending | Visual<br>Matching |
| Verbal<br>Comp                   | .64**          | .51**           | .59**                | .21               | .55**              |
| Picture<br>Similarities          | .47**          | .10             | .43*                 | .31               | .50**              |
| Naming<br>Vocabulary             | .59**          | .23             | .47*                 | .35               | .44*               |
| Pattern<br>Construction          | .45*           | .23             | .62**                | .14               | .72**              |
| Early Number<br>Concepts         | .35            | .17             | .55*                 | .38               | .56**              |
| Copying                          | -.02           | .03             | .36                  | .27               | .54**              |
| Matching<br>Letter-Like<br>Forms | .53*           | .38             | .78**                | .06               | .70**              |
| Recall of<br>Digits              | .20            | -.18            | .28                  | .15               | .28                |
| Recall of<br>Objects             | .76**          | .56**           | .57**                | .11               | .37                |
| Recall of<br>Objects-Delay       | .19            | .21             | .40                  | -.24              | .53*               |
| Recognition<br>of Pictures       | .45*           | .25             | .50*                 | .37               | .62**              |

\*\* Correlation is significant at the .01 level (2-tailed)

\* Correlation is significant at the .05 level (2-tailed)

## Appendix C

## Subtest Correlation Matrices (cont.)

| WJ III                           |                      |                     |                        |                      |                        |
|----------------------------------|----------------------|---------------------|------------------------|----------------------|------------------------|
| DAS                              | Visual<br>Matching 2 | Incomplete<br>Words | General<br>Information | Retrieval<br>Fluency | Picture<br>Recognition |
| Verbal<br>Comp                   | .93                  | .29                 | .61**                  | .42**                | .48**                  |
| Picture<br>Similarities          | .18                  | -.06                | .33                    | .33                  | .29                    |
| Naming<br>Vocabulary             | .74                  | .09                 | .54**                  | .14                  | .51**                  |
| Pattern<br>Construction          | .98                  | .25                 | .50**                  | .52**                | .59**                  |
| Early Number<br>Concepts         | .44                  | .47*                | .32                    | .56**                | .59**                  |
| Copying                          | .85                  | .47**               | -.01                   | .51*                 | .45*                   |
| Matching<br>Letter-Like<br>Forms | .93                  | .50                 | .40                    | .15                  | .34                    |
| Recall of<br>Digits              | .95                  | .34                 | .16                    | .52**                | .35                    |
| Recall of<br>Objects             | .52                  | .27                 | .70**                  | .42                  | .39                    |
| Recall of<br>Objects-Delay       | .00                  | .21                 | .17                    | .26                  | .13                    |
| Recognition<br>of Pictures       | -.42                 | .08                 | .59**                  | .50*                 | .49*                   |

\*\* Correlation is significant at the .01 level (2-tailed)

\* Correlation is significant at the .05 level (2-tailed)

## Appendix C

## Subtest Correlation Matrices (cont.)

| WJ III                           |                     |                   |                         |
|----------------------------------|---------------------|-------------------|-------------------------|
| DAS                              | Memory<br>for Names | Visual<br>Closure | Memory<br>for Sentences |
| Verbal<br>Comp                   | .39**               | .45**             | .63**                   |
| Picture<br>Similarities          | .26                 | .27               | .11                     |
| Naming<br>Vocabulary             | .27                 | .33               | .42*                    |
| Pattern<br>Construction          | .41*                | .31               | .35                     |
| Early Number<br>Concepts         | .37*                | .35               | .37*                    |
| Copying                          | .02                 | .29               | .08                     |
| Matching<br>Letter-Like<br>Forms | .53*                | .29               | .38                     |
| Recall of<br>Digits              | .09                 | .23               | .52**                   |
| Recall of<br>Objects             | .59**               | .47*              | .69**                   |
| Recall of<br>Objects-Delay       | .38                 | .66**             | .45*                    |
| Recognition<br>of Pictures       | .21                 | .24               | .31                     |

\*\* Correlation is significant at the .01 level (2-tailed)

\* Correlation is significant at the .05 level (2-tailed)

*Appendix D*

## Subtest Mean Scores and Standard Deviations

|                                 | M      | SD    |
|---------------------------------|--------|-------|
| WJ III Subtests                 |        |       |
| Verbal Comprehension (n=30)     | 94.83  | 13.48 |
| Visual-Auditory Learning (n=26) | 97.69  | 16.03 |
| Spatial Relations (n=25)        | 93.64  | 13.60 |
| Sound Blending (n=25)           | 94.16  | 7.89  |
| Visual Matching 1 (n=29)        | 95.00  | 13.46 |
| Visual Matching 2 (n=3)         | 98.67  | 6.03  |
| Incomplete Words (n=29)         | 102.52 | 8.80  |
| General Information (n=28)      | 93.39  | 13.46 |
| Retrieval Fluency (n=24)        | 88.79  | 11.03 |
| Picture Recognition (n=27)      | 91.00  | 14.04 |
| Memory for Names (n=30)         | 98.17  | 11.80 |
| Visual Closure (n=30)           | 102.87 | 13.91 |
| Memory for Sentences (n=30)     | 94.23  | 13.88 |
| DAS Subtests                    |        |       |
| Verbal Comprehension (n=30)     | 38.93  | 9.62  |
| Picture Similarities (n=30)     | 46.10  | 11.64 |
| Naming Vocabulary (n=30)        | 45.70  | 10.57 |
| Pattern Construction (n=30)     | 48.50  | 8.50  |

*Appendix D*

## Subtest Mean Scores and Standard Deviations (cont.)

|                                    | M     | SD    |
|------------------------------------|-------|-------|
| Early Number Concepts (n=30)       | 43.03 | 9.35  |
| Copying (n=30)                     | 44.63 | 9.30  |
| Matching Letter-Like Forms (n=15)  | 47.07 | 12.56 |
| Recall of Digits (n=28)            | 42.68 | 14.33 |
| Recall of Objects (n=22)           | 41.55 | 10.85 |
| Recall of Objects – Delayed (n=22) | 42.55 | 9.30  |
| Recognition of Pictures (n=28)     | 46.54 | 9.22  |

Note. WJ III subtests M = 100, SD = 15. DAS subtests M=50, SD=10.