5-1-2002

Test-Relief Reliability of the Universal Nonverbal Intelligence Test with Children Diagnosed with Attention Deficit Hyperactivity Disorder

Julia Pendley
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses
Part of the Education Commons, and the Psychology Commons

Recommended Citation
Pendley, Julia, "Test-Relief Reliability of the Universal Nonverbal Intelligence Test with Children Diagnosed with Attention Deficit Hyperactivity Disorder" (2002). Masters Theses & Specialist Projects. Paper 627.
http://digitalcommons.wku.edu/theses/627

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
TEST-RETEST RELIABILITY OF THE UNIVERSAL NONVERBAL INTELLIGENCE TEST WITH CHILDREN DIAGNOSED WITH ATTENTION DEFICIT HYPERACTIVITY DISORDER

A Thesis
Presented to
the Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
of the Requirements for the Degree
Education Specialist in School Psychology

by
Julia Dawn Pendley
May 2002
TEST-RETEST RELIABILITY OF THE UNIVERSAL NONVERBAL INTELLIGENCE TEST WITH CHILDREN DIAGNOSED WITH ATTENTION DEFICIT HYPERACTIVITY DISORDER

Date Recommended 2-1-02

Carl L. Myers
Director of Thesis

Elizabeth L. Jones

Dean, Graduate Studies and Research 4/1/02
Date
Acknowledgements

Many important people in my life deserve a thank you from me for the completion of this project. First, I would like to thank Dr. Carl Myers for all of his help. As my thesis chairperson, he guided me through this endeavor from beginning to end and never gave up hope that I would succeed on this project. A note of thanks also goes to Dr. Reagan Brown and Dr. Elizabeth Jones, thesis committee members, for their advice and encouragement throughout this project. Special recognition goes to the Pennyroyal Mental Health Clinic for allowing me access to clientele, which made this thesis possible. Special recognition also goes to my friend, Terri Owens, for all of the encouragement and support she gave me throughout the years of graduate school we shared together.

Lastly, I would like to recognize my family. Thanks to my husband, Jonathan, and my parents, Tony and Diane Fortney, for continually reminding me that I must complete this project to graduate! Their constant reminders served as strong motivators for me, mostly when I felt like giving up. Thanks to my sister, Katie, and brother, Joey, for continuing to care for me even though my stress level made my presence unbearable at times. Without my family, I would not have been able to achieve the success that I am so lucky to possess today.
Table of Contents

List of Tables ........................................................................................................... v
Abstract ...................................................................................................................... vi
Introduction .............................................................................................................. 1

Literature Review ...................................................................................................... 5
  Universal Nonverbal Intelligence Test ................................................................. 6
  Subtest Analysis ..................................................................................................... 11
  Attention Deficit Hyperactivity Disorder ............................................................ 14
  Purpose ................................................................................................................... 18

Method .................................................................................................................... 21
  Participants .......................................................................................................... 21
  Instrument ............................................................................................................. 21
  Procedure ............................................................................................................. 22

Results ..................................................................................................................... 25

Discussion ............................................................................................................... 34
  Limitations of the Study ...................................................................................... 36
  Future Research ................................................................................................... 37

References .............................................................................................................. 38

Appendix A – Human Subjects Review Board Approval ....................................... 41
Appendix B – Letter to Parents to Describe Study ................................................ 43
Appendix C – Consent Form ................................................................................... 45
List of Tables

Table 1: Subtests Composing Each Index of the Universal Nonverbal Intelligence Test.................................................................9

Table 2: Test-Retest Mean Scores and Standard Deviations for ADHD Sample......26

Table 3: Test-Retest Mean Scores and Standard Deviations for the Standardization Sample.................................................................27

Table 4: Test-Retest Correlations Obtained for ADHD Sample and Standardization Sample.................................................................29

Table 5: Consistency of Strengths and Weaknesses for Each Participant from Time 1 to Time 2.................................................................32
The Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998) [UNIT]) is a relatively new intelligence test that is administered in an entirely nonverbal way. Research supports the use of this test with special populations such as those with learning disabilities, those who are intellectually gifted, as well as with those who have speech/language impairments (Bracken & McCallum, 1998). One population not accounted for in the test’s standardization sample are children diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). This study investigates the test-retest reliability of the UNIT with children diagnosed with ADHD. Another main focus of this study involves determining the appropriateness of utilizing a test interpretation method, known as ipsative analysis, with the UNIT and children with ADHD. The results of this study support the notion that the UNIT is a reliable test to use with children diagnosed with ADHD. Obtained test-retest correlation coefficients are very similar to those found in standardization sample studies. The results of this study do not support the use of ipsative analysis of the UNIT involving children with ADHD.
Introduction

Historical trends, court rulings, and legislation in education have drawn attention to the education of students with special needs. With the implementation of the Education for All Handicapped Children Act (PL 94-142) in 1975 came drastic changes and further development of America’s special educational system. This law and subsequent revisions state that all children, regardless of disability, have the right to attend public schools. As a result, schools are serving larger numbers of children with special needs.

Students with special needs require additional attention from teachers, special educators, and school psychologists. As part of the evaluation process that enables these students to receive special education services, school psychologists are often called upon to assess students’ abilities, strengths and weaknesses. Assessment procedures followed by school psychologists are guided by federal law and ethical principles. One consideration is that the assessor must use a testing device that is fair for a particular examinee (National Association of School Psychologists, 2000). For example, the Wechsler Intelligence Scale for Children-Third Edition (Wechsler, 1991 [WISC-III]) may not be the best IQ test to use with a student who has significant expressive language delays because many of the subtests on the WISC-III use verbal instructions and require the examinee to verbally respond to test items. The scores may not provide a representative profile of the child’s true abilities. Using a test that does not require verbal responses and measures intelligence by other means may be more suitable.
Bracken and McCallum developed an intelligence test to address the issues related to unfair assessment described above. Their test, the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998 [UNIT]), has been used with students with a wide range of exceptionalities such as hearing/language impairments, autism, and severe emotional disturbances. Previous internal consistency studies on the UNIT demonstrate that the UNIT is a reliable measure for those with learning disabilities ($r = .95$), mental retardation ($r = .95$), intellectual giftedness ($r = .94$), and speech and language impairments ($r = .97$) (Bracken & McCallum, 1998). The overall test-retest reliability coefficient for the Full Scale score was .81 (Bracken & McCallum, 1998).

However, a special population not accounted for in the UNIT’s standardization sample is children diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). The UNIT has not been shown to provide an appropriate assessment of intelligence for those diagnosed with ADHD. Children with ADHD typically score between 7 and 15 points lower on verbally loaded IQ tests than those who do not possess this diagnosis (Barkley, 1990). The reason for this score discrepancy is not fully understood; however, some attribute this phenomenon to the finding that children with ADHD also have difficulties expressing themselves verbally (Barkley, 1996). The UNIT may minimize this verbal factor, since it is administered in an entirely nonverbal way. On the other hand, nonverbal tests such as the UNIT may actually require more attentiveness from students since they must visually attend to all nonverbal cues and task stimuli. By definition, children with ADHD have short attention spans and display impulsive behaviors (American Psychiatric Association [APA], 1994). The professional literature
does not appear to have addressed the issue of whether a nonverbal test would be more or less appropriate for a child with ADHD.

The researcher will attempt to provide additional research data on the UNIT and also provide additional information on the fair assessment of those diagnosed with ADHD. Would a nonverbal intelligence test, such as the UNIT, be a reliable measure of intelligence for a student with ADHD? This research question will be answered, in part, by determining the test-retest reliability coefficient for a group of students with ADHD. High test-retest correlations have already been demonstrated with a number of populations (Bracken & McCallum, 1998). A high test-retest correlation coefficient in this study would provide evidence to suggest that a nonverbal measure (i.e., the UNIT) is also a reliable test to use with children diagnosed with ADHD. However, a test-retest correlation coefficient lower than those found with other populations might imply that a student’s inattentiveness interferes with the requirements of a nonverbal measure (i.e., the UNIT).

Additionally, this research will examine the relative strengths and weaknesses of each examinee’s subtest scores using procedures described in the UNIT manual (Bracken & McCallum, 1998). This type of ipsative analysis is seen as controversial to many researchers. Kaufman (1995) claims that this type of analysis gives an in-depth look at an individual’s abilities. On the other hand, some researchers do not support ipsative analysis and describe many ways in which this type of profile analysis is problematic (Burcham & DeMers, 1995; McDermott, Fantuzzo, & Glutting, 1990; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992). In this study, the ipsative analysis issue will be evaluated by determining if children with ADHD exhibit consistent
strengths and weaknesses over test administrations. A consistent pattern of strengths and weaknesses would provide additional support for the use of the UNIT with children diagnosed with ADHD. An inconsistent pattern would provide support to the critics of ipsative analysis; that is, ipsative analysis is an unreliable method of interpreting students’ scores.
Literature Review

According to Federal Laws and State regulations, school professionals are required to assess the abilities of students to determine placement and educational services for those suspected of having special needs. As part of such an evaluation, school psychologists and school staff administer a battery of tests that typically include an intelligence test, an achievement test, and measures of adaptive behaviors. The WISC-III (Wechsler, 1991) is one IQ test often used to gain a picture of the student’s cognitive abilities. The WISC-III is an IQ test that, like most other traditional norm-referenced test, is verbally laden. This test requires the student to use expressive and receptive language abilities to answer items. Consequently, this test is not always appropriate for use with specific populations, especially for students with hearing/language impairments. Students with Attention Deficit Hyperactivity Disorder (ADHD) are also thought to be disadvantaged by traditional norm-referenced tests (Barkley, 1990; 1996). These children often have difficulty expressing themselves appropriately through verbal means and may be better assessed by a test that does not require so much language interaction.

Unfair assessment issues have prompted school psychologists and other evaluators to search for measures of intelligence that better suit individual student’s needs. Nonverbal IQ tests such as the Test of Nonverbal Intelligence-Third Edition (Brown, Sherbenor, & Johnsen, 1997) and the Leiter International Performance Scale-Revised (Roid & Miller, 1997) have been developed in order to give evaluators options to
traditional language-loaded IQ measures. However, these tests are not administered entirely nonverbally and the content itself may be verbally based (Bracken, 1999). Therefore, still existing is the problem of locating a test that is most suitable for assessing the skills of one who possesses personal characteristics that could negatively impact test performance. To combat this existing problem, Bracken and McCallum (1998) developed the Universal Nonverbal Intelligence Test, which requires absolutely no verbal communication between the examiner and the examinee.

Universal Nonverbal Intelligence Test

The Universal Nonverbal Intelligence Test (UNIT) is a newly developed intelligence test that is administered in an entirely nonverbal fashion (Bracken & McCallum, 1998). Common gestures, such as head nodding, head shaking, and shrugging, are used to communicate directions and expectations to the examinee throughout the testing session. Because the UNIT is structured in this way, the authors of the test state that this assessment procedure is ideal for those who have language and/or hearing impairments or for those who speak English as a second language.

The UNIT claims to provide a fair and representative way to assess the intellectual abilities of children between the ages of 5 and 17. The construction of the UNIT is hierarchically based with an overall level of general intelligence at the apex, called the Full Scale Intelligence Quotient (FSIQ). The authors define intelligence as "the ability to solve problems using memory and reasoning" (Bracken & McCallum, 1998, p. 12). Therefore, memory and reasoning are the primary aspects of intelligence assessed by the UNIT. The tasks that tap into these two aspects of intelligence yield scores for scales titled Memory Quotient and Reasoning Quotient.
In addition, the UNIT investigates the organizational strategies utilized by the examinee. Two secondary scales, the Symbolic Quotient and the Nonsymbolic Quotient, provide this information. The Symbolic Quotient is composed of subtests that include problem solving situations where the examinee needs to incorporate concrete and abstract stimuli to successfully complete the problem set. The symbols are usually conducive to language mediation; the more the individual is able to mediate the stimuli, the better the chance of getting the items correct. For example, in the Symbolic Memory subtest, the examinee is presented with patterns of green and black figures resembling people. To complete the item, the examinee must be able to remember the original pattern including color and type of person (baby, girl, boy, woman or man) in order to recreate the pattern without using the stimuli as a guide. The patterns increase in length after each successive response. An examinee who is able to view the stimuli, attach word meanings to the pictures, and remember the pattern for a short period of time is more likely to gain credit on the test.

On the other hand, the Nonsymbolic Quotient is composed of subtests that include problem-solving situations where the examinee must perceive and make judgments about the presented stimuli. Symbolic or verbal mediation is not necessary to complete the items. For example, in the Cube Design subtest, the examinee is presented a picture of prearranged blocks and is expected to replicate the exact colored pattern with the stimulus cubes. Each block pattern is novel in that the pictures do not represent any universal pattern; they are not constructed to resemble any common structure. The examinee must rely on his or her ability to manipulate the shapes and make decisions rather than on constructing the blocks in order to represent some known pattern.
The authors of the UNIT propose that this method of intellectual assessment will benefit those individuals who are normally restricted and unfairly assessed by traditional intelligence tests. For example, a significant number of students in America’s schools are recent immigrants and are unable to speak the English language fluently. In fact, approximately 1.4 million children in California speak English as a second language (Puente, 1998). The 1990 U.S. Census Bureau estimated that approximately 31.8 million Americans spoke a language other than English in the home setting. This statistic indicates an increase from just ten years earlier when 23.1 million Americans did not use English as their primary language. Due to the historical trends, the logical assumption can be made that more people than ever primarily use a language other than English in the United States. This type of communication hindrance could adversely affect the potential success that these students could reach on traditionally language-loaded IQ tests.

Also, many students have speech and/or language impairments that leave them unable to adequately respond to or understand verbal test questions. Therefore, an intelligence test that uses no expressive language either by the examiner or by the examinee would be ideal in these testing situations. The UNIT claims to implement this type of assessment procedure.

Three administration batteries exist for the Universal Nonverbal Intelligence Test. The Abbreviated Battery is composed of the first two subtests, Symbolic Memory and Cube Design, and is used mainly for screening purposes. The Standard Battery includes the first four subtests, Symbolic Memory, Cube Design, Spatial Memory, and Analogic Reasoning, and is primarily used for making eligibility decisions. The most extensive battery that is available is the Extended Battery, which contains all six subtests, Symbolic
Memory, Cube Design, Spatial Memory, Analogic Reasoning, Object Memory, and Mazes. This battery is recommended for both diagnosing disorders and making eligibility decisions.

The UNIT contains five scales (also called indexes), each scale addressing a different component of the overall intellectual framework measured by the test. The five indexes are Memory Quotient, Reasoning Quotient, Symbolic Quotient, Nonsymbolic Quotient, and the Full Scale Intelligence Quotient. Each index has a mean of 100 and a standard deviation of 15. These standard scores are considered to be more reliable and encompass a greater level of information than the scaled scores for the individual subtests, which have a mean of 10 and standard deviation of 3. Table 1 illustrates the five indexes and the subtests that comprise each scale.

Table 1

Subtests Composing Each Index of the Universal Nonverbal Intelligence Test

<table>
<thead>
<tr>
<th>Indexes</th>
<th>Full Scale</th>
<th>Memory</th>
<th>Reasoning</th>
<th>Symbolic</th>
<th>Nonsymbolic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbolic Memory</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cube Design</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Spatial Memory</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analogic Reasoning</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Object Memory</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Mazes</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
The test manual provides information on the psychometric properties of the UNIT, including reliability information. Reliability refers to the "consistency of measurement" (Sattler, 1992, p. 25). In other words, a test is said to be reliable if the observed scores are relatively consistent over more than one test administration (Sattler, 1992). The reliability of a test is expressed numerically, ranging from -1.00 to 1.00 with 1.00 signifying a perfect positive relationship, 0.00 signifying no relationship, and -1.00 signifying a perfect negative relationship (Howell, 1992). This number is usually referred to as the reliability coefficient. Many types of reliability coefficients exist; however, the two types of reliability coefficients that are most important to determine when using intelligence tests are test-retest and internal consistency. Test-retest reliability gives an estimate of score stability and thus is often termed the coefficient of stability. This procedure requires participants to be administered the exact same test twice, with a predetermined waiting period in-between the two testing sessions. As a general rule, the shorter the time span between the two test sessions, the higher the reliability coefficient will be (Sattler, 1992).

Internal consistency reliability can be determined during one test administration, unlike the test-retest reliability procedure just described. A common method of determining this type of reliability is to divide the test into two halves, with each half measuring the same trait as much as possible. At the end of the test administration, the two halves are correlated to determine the internal consistency reliability coefficient. (Sattler, 1992)

The reliability studies conducted during the standardization process of the UNIT have yielded a "total test internal consistency reliability of .90 or greater when averaged
across the age groups and also a total test-retest reliability coefficient of .90 or greater” (Bracken & McCallum, 1998, p. 98). “For most tests of cognitive and special abilities, a reliability coefficient of .80 or higher is generally considered to be acceptable” (Sattler, 1992, p. 25). Thus, standardization data for the UNIT exceeds acceptable standards in the field.

Score stability is necessary when interpreting assessment data, making recommendations, and implementing interventions. This type of stability is especially important when working with special populations, such as with children who are diagnosed with a learning disability, Attention Deficit Hyperactivity Disorder, or some other disorder that may negatively impact performance on an individually administered diagnostic test. In order for the teachers, parents, and other professionals involved in these children’s lives to make accurate decisions and correct recommendations to help the child succeed, reliable tests results and information must be available.

Subtest Analysis

Test interpreters typically explain assessment results by focusing on the overall scores. In addition to this traditional method of test interpretation, another method of analysis consists of evaluating the strengths and weaknesses of each participant’s test results. This procedure can be done in two ways. One way, interindividual analysis, compares the examinee’s subtest scores to the normative sample. Scores one standard deviation above the mean of the normative sample are considered strengths and scores one standard deviation below the mean are considered weaknesses. A popular approach to looking at an individual’s abilities is achieved by a second method, which is an ipsative or intraindividual analysis (McDermott et al., 1990). Ipsative analysis requires
the examiner to make an evaluation of the examinee’s strengths and weaknesses compared to the examinee’s own performance on the test. This type of analysis is done by comparing the individual’s subtest scores to the individual’s average subtest score thereby creating a profile of strengths and weaknesses for that particular test-taker (McDermott et al., 1992).

The two methods of subtest analysis can be illustrated with the WISC-III (Wechsler, 1991). To interpret the WISC-III, the examiner must compute standard scores and scaled scores. The WISC-III yields three primary standard scores including the Verbal Scale, Performance Scale, and the Full Scale standard scores. In addition, an examinee may have up to thirteen scaled scores from one administration of the WISC-III. These scaled scores are derived from the thirteen subtests that comprise the WISC-III. As mentioned previously, scaled scores have a mean of 10 and standard deviation of 3, which allows a person’s scaled scores to be compared to other children in the standardization sample. A person who attains a scaled score ranging from seven to thirteen on any one of the subtests is said to be average in the ability assessed by the particular subtest.

To illustrate interindividual subtest analysis, suppose an individual attained a scaled score of six on the Picture Completion subtest and fourteen on the Vocabulary subtest. The conclusion can be drawn that the individual has a strength in verbally defining words because the score of fourteen is more than one standard deviation above the mean. However, this individual is below average in his ability to find missing parts in pictures as assessed by the Picture Completion subtest. These conclusions were made by comparing the individual’s scores to the normative sample.
Typically, however, an interpreter uses ipsative analysis to define strengths and weaknesses relative to the examinee's own performance. The mean of the individual's subtest scores would be determined. For example, if the individual's average subtest score is twelve, one could still make the interpretation that the individual with a Picture Completion subtest score of six is weaker in his ability to find missing parts in pictures. However, a scaled score of fourteen on the Vocabulary subtest would not be considered a strength in this type of analysis because fourteen is not more than one standard deviation above the average subtest score of twelve. When this type of interpretation is made, the norm tables are not consulted. The individual's average score serves as the norm of comparison instead.

Kaufman (1995) advocates the use of this ipsative approach with the Wechsler scales. “The ipsative method makes it more likely that both assets and deficits will be identified for a child” (Kaufman, 1995, p. 30). He has made many attempts to verify the limitations of intelligence tests and urges psychologists to look beyond the realm of standard score interpretation and to understand what lies beneath and represents the individual subtests. Discovering such information seems vital to fully understanding a person’s true abilities, to develop learning experiences based on strengths, and to develop remediation strategies for the weaknesses.

Not all researchers advocate for the use of ipsative analysis. McDermott et al. (1990) believe that even though this type score interpretation is the most popular form of intraindividual assessment, its popularity is unwarranted. McDermott et al. (1990) postulate that the ipsatization of scores causes some problems in accordance with data analysis standards, including lower construct validity, lower predictive validity, and
decreased practical utility. Using the Wechsler Intelligence Scale for Children-Revised Edition (Wechsler, 1974 [WISC-R]) standardization sample, McDermott and his colleagues examined the score stability of ipsatized scores. McDermott et al. (1992) hypothesized that the ipsatization of subtest scores on the WISC-R would cause a noticeable reduction in score stability or test-retest reliability. The results of their study supported this hypothesis. They found that the “average short-term reliability across normative ability attributes is .78, dropping to .63 upon ipsitization using all subtests and .62 using separate IQ scales” (McDermott et al., 1992, p. 512). In this case, the researchers defined short-term as being a one-month time span between the two testings. To further explain these findings, a person would have only a 42.5% chance of having the same identified strength over time. Additionally, an individual would have only a 34.5% chance of having the same identified weakness over time (McDermott et al., 1992). Thus, serious concerns have been raised about the appropriateness of an ipsative approach to test interpretation.

Attention Deficit Hyperactivity Disorder

It appears that more attention has been given to Attention Deficit Hyperactivity Disorder (ADHD) in recent years. The symptomatology has been around for centuries; however, the need to deal with the symptoms was not as great in the past as it is today (Barkley, 1996). The increase in the importance of a formal education has caused parents and teachers to become concerned about the progress of children that display hyperactive and inattentive behaviors. A tremendous amount of research has focused on ADHD to determine the cause of this disorder, how to diagnose it, and what associated problems develop as a result.
ADHD is the clinical diagnosis given to children who are “excessively active, are unable to sustain their attention, and are deficient in their impulse control to a degree that is deviant for their developmental level” (APA, 1994, p. 79). For a child to become diagnosed with this disorder, he or she must meet a set of pre-established criteria found in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (APA, 1994 [DSM-IV]). The DSM-IV provides three different diagnosis options for ADHD, which are (a) ADHD, Predominately Inattentive Type, (b) ADHD, Predominately Hyperactive-Impulsive Type, and (c) ADHD, Combined Type. Two sets of criteria are provided, one set for inattentiveness and one set for hyperactive-impulsive behaviors. To meet the classification criteria for either, a child must exhibit at least six of the noted symptoms. Children who exhibit at least six different symptoms in both categories are given the diagnosis of ADHD, Combined Type. The symptoms must be present before the age of seven and be evident in at least two different settings (for example, home and school).

Inattention is viewed as the largest area of difficulty for children diagnosed with ADHD, Predominately Inattentive type. Inattention problems include making careless errors, exhibiting difficulties in sustaining attention during leisure activities, and becoming easily distracted. Children with these weaknesses often fail to complete assignments at school or chores at home. They exhibit difficulty with following through with requests made by teachers and parents and often show signs of extreme forgetfulness (APA, 1994).

Children who are diagnosed with ADHD, Predominately Hyperactive-Impulsive Type, exhibit many hyperactive and impulsive behaviors. These children are known for their elevated physical activity levels, excessive talking, restlessness, and frequent
interruptions of others. Typical behaviors exhibited by children with ADHD include the inability to take turns or to conform to social norms regarding conversation skills (APA, 1994).

According to the DSM-IV (APA, 1994), approximately 3 to 5% of America's school-aged children are diagnosed with some form of ADHD. Males are three to six times more likely to become diagnosed with ADHD as compared with their female counterparts. These percentages represent a significant portion of the school population. Education laws and regulations regarding equal opportunities for education for all students, regardless of disabilities or other impairments, emphasizes the need to determine how to accommodate and modify for students with ADHD. To develop accommodations and modifications to the curriculum, an evaluation is typically warranted. One component of the evaluation process includes standardized testing to be completed with the child, usually including intelligence, achievement, and adaptive behavior measures.

Research indicates that children with the diagnosis of ADHD tend to score lower on IQ tests. In fact, many studies have shown that children with ADHD score an average of 7 to 15 points below other children without the disorder (Faraone et al., 1993; Fischer, Barkley, Edelbrock, & Smallish, 1990; McGee, Williams, Moffitt, & Anderson, 1989; Prior, Leonard, & Wood, 1983; Tarver-Behring, Barkley, & Karlsson, 1985; Werry, Elkind, & Reeves, 1987). The question is raised as to what accounts for this lower ability level. Some claim that children with ADHD are unfairly assessed by standardized tests because most of these tests require some type of verbal response or manipulation by the examinee. Studies have shown time and again that children with ADHD score lower on
the verbal measures and better on the measures that assess other aspects of intelligence unrelated to verbal ability (Barkley, 1996; Halperin & Gittelman, 1982; McGee, Williams, & Feehan, 1992; Sonuga-Barke, Lamparelli, Stevenson, Thompson, & Henry, 1994; Werry et al., 1987). Other cognitive impairments include poor planning abilities, poor organizational skills, working memory deficiencies, and decreased effort and motivation (Barkley, 1996).

Many theories have been proposed to explain the causality and severity of this disorder. Russell Barkley is considered to be one of the leading experts and theorists in the area of ADHD. He proposed a theory to explain the role that behavioral inhibition plays in the regulation of behaviors. His theory rests upon previous work conducted by Schachar and Logan (1990). These two researchers proposed a theoretical framework to explain the cognitive processes of children diagnosed with ADHD. They hypothesized that every stimulus produces two types of responses in the brain: a primary response and an inhibitory response. These two response patterns strive to react to the situation at hand; however, an individual that is diagnosed with ADHD lacks a fully developed inhibitory response pattern. Therefore, the individual reacts to a situation by behaving in the first way they know how (Schachar & Logan, 1990).

Building upon Schachar and Logan's framework, Barkley developed his own theoretical model that included behavioral inhibition as the predominant factor in explaining the inattentive, hyperactive, and impulsive behaviors exhibited by individuals with ADHD. He stated that behavioral inhibition is the gatekeeper of the four executive functions in the brain that control and respond to stimuli. These four executive functions are working memory, self-regulation of affect, internalization of speech, and
reconstitution. Of particular relevance to this thesis is the aspect of working memory. Barkley (1996) defines working memory as “the ability to maintain mental information on-line while acting upon it” (p. 71).

Because children with ADHD demonstrate insufficient signs of behavioral inhibition, the four executive functions previously mentioned are consistently less successfully utilized resulting in the outward expressions of inattention and hyperactive-impulsive behaviors. Instead of processing information with the four executive functions, the individual often produces the first motor or verbal response that is generated in the brain. Due to the fact that the response does not take into account the executive functions, the child engages in behaviors that are inappropriate which then leads to the problematic situations that are so popular for children diagnosed with ADHD.

The deficiencies that Barkley points out regarding behavioral inhibition raises the question of whether or not the UNIT would be a reliable tool to assess the intelligence of a child with ADHD. As mentioned previously, nonverbal tests require the examinee to remain attentive to all cues given to them during the testing situation. In addition, the UNIT assesses working memory, reasoning skills, organizational skills, and planning behaviors. Verbal skills are not assessed directly; however, verbal mediations of test material is positively correlated with test success.

Purpose

The purpose of this study is to add to the existing research data regarding the UNIT. This newly developed intelligence test shows promise as an assessment tool for use with those who have hearing/language impairments, those who are considered to be mentally disabled, as well as those who are determined to be intellectually gifted.
However, it is unknown as to whether the UNIT is an effective instrument to use with children diagnosed with ADHD. Since the UNIT is administered in an entirely nonverbal fashion, children with ADHD might be more successful at responding to the items correctly if they can respond nonverbally. On the other hand, nonverbal tests may actually require a greater level of sustained attention from the examinee than verbally-laden IQ tests. If so, nonverbal IQ tests may actually decrease the likelihood that children with ADHD will correctly answer the items.

An explicit evaluation of whether or not the UNIT requires more sustained attention than verbally-laden IQ measures is beyond the scope of the present study. This study will, however, examine the issue indirectly by evaluating the test-retest reliability of test scores for children diagnosed with ADHD. If children with ADHD have test-retest reliability coefficients as high as those found in the UNIT Examiner’s Manual (Bracken & McCallum, 1998), the results would lend support to the notion that the UNIT is a reliable IQ measure for those with ADHD. On the other hand, if lower test-retest reliability coefficients are found, the results could imply the UNIT is not advantageous for children with ADHD. In addition, the consistency of strengths and weaknesses will be determined through ipsative analysis procedures by comparing each participant’s performance on the initial test with his or her performance on the second test. Consistent patterns of strengths and weaknesses would be another indicator that the UNIT is a reliable IQ measure for those with ADHD. Inconsistent patterns of strengths and weaknesses would be indicative that ipsative analysis of the UNIT is not a reliable analysis for those with ADHD. Inconsistent patterns may also suggest ipsative analysis is
an inappropriate method of test interpretation. Thus, the following two research
questions are addressed in this study:

1) Does the UNIT have test-retest correlations as high as those found with the
standardization sample?

2) How consistent are patterns of strengths and weaknesses on the UNIT over
time with children who have ADHD?
Method

Participants

Participants were recruited from children served by the Pennyroyal Mental Health Clinic (PMHC), a clinic based in Western Kentucky that services an eight county area. A total of 30 children (21 males, 9 females) participated in this study. All participants were between the ages of 5 years, 3 months and 17 years, 0 months. The mean age was 10 years, 6 months (SD=37.4 months). Twenty-eight (93.3%) of the participants were Caucasian, whereas two (6.7%) were African American. All of the participants (100%) were taking prescription medications to relieve symptoms of ADHD. Each participant had only a diagnosis of ADHD; children with co-morbid or other mental health diagnoses were not asked to participate. The study design and procedures were approved by the Western Kentucky University Human Subjects Review Board (see Appendix A).

Instrument

The UNIT (Bracken & McCallum, 1998) was administered to assess the cognitive abilities of each participant during both testing sessions. Since the UNIT is a relatively new intelligence test, little research exists for this particular diagnostic tool. An internet database search on the UNIT yielded only a few published works involving this test. Thus, the UNIT Examiner’s Manual (Bracken & McCallum, 1998) houses the most extensive research information on the UNIT.

Internal consistency reliability studies conducted with the UNIT standardization sample (n = 175) yielded coefficients ranging from .86 to .93 for the Extended Battery
Full Scale and other overall scores, with the Full Scale having the highest coefficient \((r = .93)\). The Full Scale internal consistency reliability coefficient rose even higher when the studies were conducted with special populations. The learning disabled group yielded a coefficient of .94. Studies conducted with the mentally retarded as well as with those with speech/language impairments yielded reliability coefficients of .97.

Test-retest reliability studies conducted with the standardization sample \((n = 197)\) yielded an Extended Battery Full Scale coefficient of stability of .81. All other scales yielded coefficients of stability ranging from .75 to .79. The Cube Design subtest had the highest coefficient of stability \((r = .83)\) while the Mazes subtest yielded the lowest \((r = .57)\). The UNIT Examiner’s Manual does not state test-retest reliabilities for special populations.

Validity studies conducted during the standardization process of the UNIT provided evidence of concurrent and predictive validity. The UNIT correlates well with widely used cognitive ability measures, such as the WISC-III, the Woodcock-Johnson Psycho-Educational Battery – Revised, the Test of Nonverbal Intelligence – Second Edition, and the Kaufman Brief Intelligence Test (Murphy, Plake, & Impara, 2001). The UNIT also correlates well with achievement tests, such as the Woodcock-Johnson Tests of Achievement – Revised, the Wechsler Individual Achievement Test, and the Peabody Individual Achievement Test – Revised (Murphy, Plake, & Impara, 2001).

**Procedure**

A list of potential participants was generated using the PMHC database. The primary diagnosis of ADHD and age were used as parameters to search this database. The Executive Director and staff of the PMHC sent a description of the study by mail to
176 potential participants (see Appendix B). The forty-four parents (25.0%) who agreed to allow their children to participate in the study were asked to complete and sign the consent form (see Appendix C) and also an information release form. The information release form allowed the primary investigator to validate diagnoses and ages by looking into the patients’ records. The parents were asked to return the permission forms to the Pennyroyal Mental Health Center. The primary investigator contacted each parent within one week of obtaining their consent form to schedule the initial testing session and answer any questions from the parents. The parents were given information on administration time frames and were asked to schedule the testing sessions at their convenience. All testing sessions were conducted in the afternoons and evenings during the week and during daytime hours on Saturdays to insure that the children did not miss any school or therapy sessions. Also, the parents were asked to keep medication management a priority during the entire time their children were involved in the study to insure that the administration of medication did not change between test administrations. The primary investigator asked each parent whether there were had changes in medication at the time of the second testing. No participants had changes in their medication between testings.

During the first session, the primary investigator explained the basic procedures for the test administration including necessary hand motions used to communicate during the session. The children were told that some of the tasks would be easy and some would be hard, but to work as best as they could. They were allowed to take breaks during the test, as needed.
Each participant was administered the Extended Battery of the UNIT, which included the following subtests: Symbolic Memory, Cube Design, Spatial Memory, Analogic Reasoning, Object Memory, and Mazes. The examiner administered the items in a sequential order and started with the item deemed appropriate for the age of the participant. The first item on each of the subtests was considered a demonstration item; therefore, all of the first items were both administered and answered by the examiner in order to show the child how they were expected to answer the remaining items. The second item on each of the subtests was a sample item that the participants were expected to answer independently. However, the examiner was allowed to correct the child’s response if he or she answered incorrectly. Featured throughout the rest of the item sets were checkpoint items that the examiner could also correct, if necessary. Each administration lasted between 1 and 1 ½ hours.

A four-week period (mean of 31.0 days, mode of 32 days, and range of 27 to 34 days) elapsed before each participant came in for the second administration of the UNIT. In order to prevent subjects from remembering tasks from the initial testing, a four-week time lapse was chosen. A longer time period between testings might have allowed external factors (e.g., learning, therapy) to influence the results. The exact same procedures that were established during the first testing session were followed during the second testing session. Parents who requested feedback on their child’s performance were contacted by phone to explain the results.
Results

Of the 44 parents who returned consent forms, a final sample of 30 participants were selected. Due to time constraints, the first 30 consent forms that were returned were selected for this present study. The other 14 parents who sent in signed consent forms were contacted and told that their children were not needed for this study. A summary of the results from both administrations of the UNIT for the 30 children with ADHD is provided in Table 2. The mean Full Scale IQ for this sample of ADHD participants was 87.6 which is almost one standard deviation below the standardization mean of 100. The variances for both groups, however, were very similar. On average, participants scored lower on the Memory and Symbolic scales and higher on the Reasoning and Nonsymbolic scales. Average scores on the UNIT’s scales rose two to three points from the first to the second testing, as might be expected due to an increased familiarity with the test.

Table 3 presents the means and standard deviations for the test-retest data as supplied in the UNIT Examiner’s Manual (Bracken & McCallum, 1998). The table was based on 197 participants with approximately equal number of students at each age level from ages five to seventeen. The majority (76.1%) of the participants were Caucasian, while 19.8% were African American. Males comprised 49.2% of the sample. Average scores on the UNIT’s scales rose four to five points from the first to the second testing. “The mean test-retest interval was 20.3 days, with a range from 3 to 42 days” (Bracken & McCallum, 1998, p. 107).
Table 2

Test-Retest Mean Scores and Standard Deviations for ADHD Sample

<table>
<thead>
<tr>
<th>Scales</th>
<th>Time 1</th>
<th></th>
<th>Time 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>87.6</td>
<td>12.6</td>
<td>90.2</td>
<td>14.0</td>
</tr>
<tr>
<td>Memory IQ</td>
<td>84.3</td>
<td>11.9</td>
<td>86.4</td>
<td>12.0</td>
</tr>
<tr>
<td>Reasoning IQ</td>
<td>94.1</td>
<td>14.5</td>
<td>96.6</td>
<td>15.5</td>
</tr>
<tr>
<td>Symbolic IQ</td>
<td>86.8</td>
<td>13.0</td>
<td>88.9</td>
<td>13.2</td>
</tr>
<tr>
<td>Nonsymbolic IQ</td>
<td>91.2</td>
<td>12.9</td>
<td>93.7</td>
<td>14.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subtests</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbolic Memory</td>
<td>7.5</td>
<td>2.7</td>
<td>7.7</td>
<td>2.2</td>
</tr>
<tr>
<td>Cube Design</td>
<td>9.0</td>
<td>2.6</td>
<td>9.6</td>
<td>2.8</td>
</tr>
<tr>
<td>Spatial Memory</td>
<td>7.1</td>
<td>2.3</td>
<td>7.7</td>
<td>2.1</td>
</tr>
<tr>
<td>Analogic Reasoning</td>
<td>8.3</td>
<td>2.8</td>
<td>8.9</td>
<td>2.6</td>
</tr>
<tr>
<td>Object Memory</td>
<td>8.0</td>
<td>2.4</td>
<td>8.3</td>
<td>2.6</td>
</tr>
<tr>
<td>Mazes</td>
<td>10.1</td>
<td>2.7</td>
<td>10.0</td>
<td>2.7</td>
</tr>
</tbody>
</table>
Table 3

Test-Retest Mean Scores and Standard Deviations for the Standardization Sample

<table>
<thead>
<tr>
<th>Scales</th>
<th>Time 1</th>
<th></th>
<th>Time 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>102.1</td>
<td>12.8</td>
<td>107.0</td>
<td>13.8</td>
</tr>
<tr>
<td>Memory IQ</td>
<td>102.0</td>
<td>13.3</td>
<td>106.7</td>
<td>14.2</td>
</tr>
<tr>
<td>Reasoning IQ</td>
<td>100.9</td>
<td>12.8</td>
<td>105.8</td>
<td>14.0</td>
</tr>
<tr>
<td>Symbolic IQ</td>
<td>101.7</td>
<td>13.7</td>
<td>106.6</td>
<td>14.5</td>
</tr>
<tr>
<td>Nonsymbolic IQ</td>
<td>101.1</td>
<td>12.8</td>
<td>106.1</td>
<td>13.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subtests</th>
<th>Time 1</th>
<th></th>
<th>Time 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbolic Memory</td>
<td>10.5</td>
<td>2.7</td>
<td>11.8</td>
<td>2.8</td>
</tr>
<tr>
<td>Cube Design</td>
<td>10.6</td>
<td>2.7</td>
<td>11.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Spatial Memory</td>
<td>10.3</td>
<td>2.6</td>
<td>10.6</td>
<td>2.5</td>
</tr>
<tr>
<td>Analogic Reasoning</td>
<td>10.2</td>
<td>2.6</td>
<td>10.5</td>
<td>2.8</td>
</tr>
<tr>
<td>Object Memory</td>
<td>10.2</td>
<td>2.9</td>
<td>10.9</td>
<td>3.2</td>
</tr>
<tr>
<td>Mazes</td>
<td>9.7</td>
<td>2.9</td>
<td>10.4</td>
<td>3.0</td>
</tr>
</tbody>
</table>

To assess the test-retest stability of the UNIT, reliability coefficients were determined and are reported in Table 4. All test-retest correlations for the ADHD sample were significant. The test-retest correlation coefficients from the ADHD sample were very similar to the correlations from the standardization sample. When comparing the correlation coefficients from the ADHD sample to the standardization sample, the differences for each scale or subtest ranged from .01 to .09 with the largest difference being on the Mazes subtest.
Table 4

Test-Retest Correlations Obtained for ADHD Sample and Standardization Sample

<table>
<thead>
<tr>
<th>Scales</th>
<th>ADHD Sample</th>
<th>Standardization Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Scale IQ</td>
<td>.83*</td>
<td>.81</td>
</tr>
<tr>
<td>Memory IQ</td>
<td>.74*</td>
<td>.75</td>
</tr>
<tr>
<td>Reasoning IQ</td>
<td>.86*</td>
<td>.79</td>
</tr>
<tr>
<td>Symbolic IQ</td>
<td>.82*</td>
<td>.77</td>
</tr>
<tr>
<td>Nonsymbolic IQ</td>
<td>.74*</td>
<td>.77</td>
</tr>
</tbody>
</table>

**Subtests**

<table>
<thead>
<tr>
<th>Subtests</th>
<th>ADHD Sample</th>
<th>Standardization Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbolic Memory</td>
<td>.69*</td>
<td>.68</td>
</tr>
<tr>
<td>Cube Design</td>
<td>.85*</td>
<td>.83</td>
</tr>
<tr>
<td>Spatial Memory</td>
<td>.66*</td>
<td>.63</td>
</tr>
<tr>
<td>Analogic Reasoning</td>
<td>.71*</td>
<td>.67</td>
</tr>
<tr>
<td>Object Memory</td>
<td>.64*</td>
<td>.59</td>
</tr>
<tr>
<td>Mazes</td>
<td>.48*</td>
<td>.57</td>
</tr>
</tbody>
</table>

* The 95% confidence interval for the Full Scale IQ correlation is .67 to .92.


*P < .01.

An ipsative analysis was conducted to determine subtest strength/weakness stability between the two testing sessions. In order to determine strengths and weaknesses, each participant’s subtest performance was compared to his or her overall performance during the corresponding testing session. For example, Participant 1 had the following subtest score profile:
The difference scores for each participant were compared to difference scores provided in the UNIT manual. The difference scores in the manual represent the least amount of difference within each subtest required for statistical significance (Bracken & McCallum, 1998). For the analyses conducted in this study, the following difference scores were used at the .05 level: Symbolic Memory (+/- 2.91), Cube Design (+/- 2.43), Spatial Memory (+/- 3.19), Analogic Memory (+/- 3.33), Object Memory (+/- 3.49), and Mazes (+/- 4.14).

Using the difference scores from the manual, Participant 1 evidenced a weakness on the Symbolic Memory subtest on both Time 1 and Time 2. He possessed a personal strength on the Cube Design subtest on both Time 1 and Time 2. He did not evidence any other strengths or weaknesses on any other part of the test for either testing session. The consistency of strengths and weaknesses from Time 1 to Time 2 is presented in
Table 5. Strengths and weaknesses for each individual were determined using the procedure just described for Participant 1.

As would be expected, most subtest scores (80.6%) were neither strengths nor weaknesses during either test administrations. To avoid an inflated measure of consistency, these subtests were not used in the analysis. When examining only those subtests where a significant strength or weakness occurred, only 14.3% of the participants had a consistent strength or weakness on the same subtest on both Time 1 and Time 2. To calculate this percentage, the number of matched strengths/weaknesses were divided by the number of matched strengths/weaknesses plus the number of mismatches on Time 1 and Time 2 (Y/Y+X or 5/5+30).
Table 5

Consistency of Strengths and Weaknesses for Each Participant from Time 1 to Time 2.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Symbolic Memory</th>
<th>Cube Design</th>
<th>Spatial Memory</th>
<th>Analogic Reasoning</th>
<th>Object Memory</th>
<th>Mazes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>7</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>8</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>9</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>10</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>11</td>
<td>Y</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>12</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>14</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>X</td>
</tr>
<tr>
<td>15</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>16</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>X</td>
</tr>
<tr>
<td>17</td>
<td>N</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>X</td>
</tr>
<tr>
<td>18</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>19</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>X</td>
<td>N</td>
<td>X</td>
</tr>
<tr>
<td>20</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>21</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>22</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

(table continues)
<table>
<thead>
<tr>
<th>Participant</th>
<th>Symbolic Memory</th>
<th>Cube Design</th>
<th>Spatial Memory</th>
<th>Analogic Reasoning</th>
<th>Object Memory</th>
<th>Mazes</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>24</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>25</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>26</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>27</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>28</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>29</td>
<td>X</td>
<td>N</td>
<td>X</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>30</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Note. Y = Participant possessed a consistent strength or weakness on both Time 1 and Time 2; N = Participant did not attain a rating of strength or weakness on either Time 1 or Time 2; X = Participant possessed an inconsistent strength or weakness between Time 1 and Time 2. Total number of Y = 5; Total number of N = 145; Total number of X = 30.
Discussion

The purpose of this study was to determine the test-retest reliability of the UNIT when administered to children diagnosed with ADHD. A secondary purpose was to determine the consistency of subtest strengths and weaknesses over time with children diagnosed with ADHD. According to Sattler (1992), a test must have a reliability coefficient of .80 or greater to be considered acceptable for diagnostic purposes. The reliability coefficient of the Full Scale IQ on the UNIT (.83) exceeded the .80 reliability coefficient criterion, thus deeming the UNIT to have acceptable test-retest reliability for children with ADHD. The Reasoning and Symbolic scales also had test-retest reliability coefficients of .80 or greater (range .82 to .86). The other scales, Memory and Nonsymbolic, did not meet that criterion with coefficients of .74. The test-retest reliability coefficients for the children with ADHD in this sample were found to be as high as those found in the UNIT Examiner’s Manual (Bracken & McCallum, 1998). Thus, it would appear the current results support the notion that the UNIT is a reliable IQ measure for children with ADHD.

The ipsative analysis was demonstrated to be a poor test interpretation method, at least with children with ADHD. Theoretically, if an individual performed well enough or poorly enough to attain a rating of strength or weakness during one testing session, then the person should demonstrate a similar performance on the same test a short time later. This notion was not supported by this study. Only 14.3% of the subtests with a strength or weakness on Time 1 had the same strength or weakness on Time 2. The current results
are much lower than those reported by McDermott et al. (1992) who found a 35-43% chance of having a same identified strength or weakness over time on the WISC-R. There are two possibilities that may explain why the strengths and weaknesses were inconsistent over time with ADHD children. First, ipsative analysis of the UNIT may not be a reliable analysis for those with ADHD. The characteristics of ADHD (e.g., short attention span, distractibility) may interfere with consistent test performance. A second explanation for inconsistent strengths and weaknesses over time is that ipsative analysis is an inappropriate method of test interpretation. Kaufman (1995), however, argued that ipsative analysis is an appropriate interpretation method on a single test administration but that the practice effect of repeating the test confounds the interpretation of strengths and weaknesses. Nevertheless, the current results imply that ipsative analysis is a very unreliable method of test interpretation.

Barkley (1990) reported that children with ADHD typically score between 7 and 15 points lower on verbally loaded IQ tests than children without ADHD. The overall Full Scale IQ for the current sample of ADHD children was 12.4 points below the mean of 100. Thus, the lower IQ scores for children with ADHD do not appear to be related to the verbal content of IQ tests as hypothesized by Barkley (1996).

Participants in this study demonstrated higher mean scores on the Reasoning and Nonsymbolic scales than they did on the Memory and Symbolic scales. Lower scores on the Memory scale are consistent with Barkley’s (1996) theory related to poor working memory in children with ADHD. According to Bracken and McCallum (1998), the Symbolic portion of the test assesses an examinee’s ability to solve problems that involve meaningful stimuli. Test material on this portion of the test is conducive to verbal
mediation and can be solved using organizational strategies, such as labeling and categorizing. Thus, lower Symbolic scores might be expected since, by definition, children with ADHD tend to have poor organizational strategies (APA, 1994).

Limitations

The findings of this present study provide support for the use of the overall Full Scale IQ on the UNIT with children who have ADHD. There are, however, limitations to this study that must be addressed. One problem associated with this study was the small sample size. Only thirty participants from one region of Kentucky were involved in this study. Caution should be exercised before generalizing the performance of the participants in this study to others diagnosed with ADHD. An additional result of the small sample size is that neither age effects nor gender effects were able to be calculated.

A major limitation of this study involves the use of medication. All participants in this study were taking prescribed medication to relieve symptoms of ADHD. The results may have differed had they not been taking medication for ADHD. In addition, the author of this study was the only examiner who performed the testing. Obviously, the author of this study was not blind to the purposes of the research, therefore introducing the possibility of experimenter bias.

Other potential limitations relate to the diagnosis of ADHD. Independent professionals did not verify the diagnosis of ADHD for the current sample of participants; thus, we have no information on the validity of the diagnoses. Additionally, the type of ADHD for this sample of participants was not determined. The DSM-IV (APA, 1994) delineates criteria for three types of ADHD: (a) predominately inattentive,
(b) predominately hyperactive-impulsive, and (c) combined. It is unknown if the type of ADHD would have an effect on test-retest reliability.

Future Research

Future studies regarding the use of the UNIT with children diagnosed with ADHD need to be conducted. Future studies need to incorporate larger, more diverse samples into the methodology so that age and gender effects can be examined and generalizability enhanced. It would be interesting to determine whether or not younger children with ADHD perform significantly different than older children with ADHD on the UNIT. In addition, it would be interesting to determine whether males and females with ADHD perform similarly or differently on the UNIT. Future studies need to be conducted to account for the potential effects of medication on test performance. A test-retest reliability study on children with ADHD who are not on medication needs to be conducted.
References


Appendix A

Human Subjects Review Board Approval
Julia Dawn Beliles Pendley  
2870 State Route 1380  
Central City, KY  42330  

Dear Ms. Pendley:

Your research project, “Test-Retest Reliability of the Universal Nonverbal Intelligence Test with Children Diagnosed with Attention Deficit-Hyperactivity Disorder,” was reviewed by the HSRB and it has been determined that risks to subjects are: (1) minimized and reasonable; and that (2) research procedures are consistent with a sound research design and do not expose the subjects to unnecessary risk. Reviewers determined that: (1) benefits to subjects are considered along with the importance of the topic and that outcomes are reasonable; (2) selection of subjects is equitable; and (3) the purposes of the research and the research setting is amenable to subjects’ welfare and producing desired outcomes; that indications of coercion or prejudice are absent, and that participation is clearly voluntary.

1. In addition, the IRB found that: (1) informed consent will be sought and documented from each prospective subject. (2) Provision is made for collecting, using and storing data in a manner that protects the safety and privacy of the subjects and the confidentiality of the data. (3) Appropriate safeguards are included to protect the rights and welfare of the subjects.

   a. Your research therefore meets the criteria of Full Board Review and is approved.

2. Please note that the institution is not responsible for any actions regarding this protocol before approval. If you expand the project at a later date to use other instruments please re-apply. Copies of your request for human subjects review, your application, and this approval, are maintained in the Office of Sponsored Programs at the above address. Please report any changes to this approved protocol to this office. A Continuing Review protocol will be sent to you in the future to determine the status of the project.

Kindest regards.

Sincerely,

Phillip E. Myers, Ph.D.  
Director, Office of Sponsored Programs and  
Human Subjects Coordinator

[Signature]

Dr. Carl Myers, Department of Psychology  
HSApprovalPendleyHS0138
Appendix B

Letter to Parents to Describe Study
Dear Parent/Guardian,

An employee of the Pennyroyal Center, Julia D. Pendley is a graduate student studying to be a School Psychologist at Western Kentucky University. As part of her graduate work, she is expected to do a special study. Her study, Assessment with the Universal Nonverbal Intelligence Test, is an examination of the usefulness of testing children who are diagnosed with Attention Deficit-Hyperactivity Disorder (ADHD) with the Universal Nonverbal Intelligence Test.

We support her in this work and offer an invitation for your child to participate in the study. We emphasize that your child’s participation is entirely voluntary. If you decide that your child should not participate, there will be no adverse affect as far as services from the Pennyroyal Center are concerned.

If you decide to participate, your child will spend two 1-hour sessions with Ms. Pendley. Your child will be asked to handle objects, such as blocks and cards, in order to complete puzzle-like activities. These activities are found to be enjoyable by most children. The sessions, which are free of charge, will take place at your local Pennyroyal Center clinic at a time that is convenient for you.

You may be assured that your child’s results will be kept strictly confidential. Records will be coded to protect your child’s identity.

Should you have any questions you may contact this office (1-877-4RESPOND – toll free) and we will arrange for Ms. Pendley to call you. We hope that you will agree to participate in this study. If you do, complete the attached consent form and return it to this office in the enclosed envelope. Ms. Pendley will contact you to arrange the initial appointment.

Thank you for your cooperation!

Sincerely yours,

Thomas W. Westerfield, M.A.
Executive Director
Appendix C

Consent Form
Assessment with the Universal Nonverbal Intelligence Test

Consent Form

I have read the information provided about this study and give my consent for my child to participate in two 1-hour sessions with Ms. Pendley. My child’s results will be used for Ms. Pendley’s special study, Assessment with the Universal Nonverbal Intelligence Test. This study will give information about whether using the Universal Nonverbal Intelligence Test is a good test to use with children diagnosed with ADHD.

Parent/Guardian Signature ________________________________
Child’s Name ________________________________
Phone Number ________________________________
Date ________________________________