12-2002

Concurrent and predictive validity of the Universal Nonverbal Intelligence Test and the Leiter International Performance Scale - Revised

V. Scott Hooper
University of Tennessee

To the Graduate Council:

I am submitting herewith a dissertation written by V. Scott Hooper entitled "Concurrent and predictive validity of the Universal Nonverbal Intelligence Test and the Leiter International Performance Scale - Revised." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

R. Steve McCallum, Major Professor

We have read this dissertation and recommend its acceptance:

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Vincent Scott Hooper entitled "Concurrent and Predictive Validity of the Universal Nonverbal Intelligence Test and the Leiter International Performance Scale – Revised." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

R. Steve McCallum, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Vice Provost and Dean of
Graduate Studies

CONCURRENT AND PREDICTIVE VALIDITY OF THE UNIVERSAL NONVERBAL INTELLIGENCE TEST AND THE LEITER INTERNATIONAL PERFORMANCE SCALE – REVISED

A Dissertation
Presented for the
Doctor of Philosophy Degree
The University of Tennessee, Knoxville

V. Scott Hooper
December, 2002

DEDICATION

This dissertation is dedicated to my parents, Edward and Linda Hooper. Your love, support, and encouragement have enabled me to achieve this and other milestones in my life. Thank you for your unconditional emotional and financial support throughout my academic career. Your expectation of excellence from me provided the motivation to achieve the highest level of academic achievement in obtaining my Ph.D.

This dissertation is also dedicated to my wife, Carolyn Wade Hooper. You are my best friend and most respected colleague. I could have never reached this goal without you. You changed my life and taught me the true meaning of unconditional love.

I would also like to make a special dedication to my grandmother, Louise Morrison, and to the memory of my grandfather, Oscar Morrison. From my first tee-ball game to my wedding and beyond, you were always there without fail. You embodied the value of hard work and the concept of marriage as a loving partnership.

Thank you all.... This is only the beginning.

ACKNOWLEDGEMENTS

There are many people who have supported my efforts in completing this project. I would like to thank my committee chair, Dr. Steve McCallum, for his support and wisdom during my graduate school career. I would also like to thank my committee members, Dr. Charles Thompson, Dr. Don Dickinson, and Dr. Richard Saudargas for their important contributions to the completion of this document.

I would like to thank Donna Palmer, Supervisor of Psychological Services for the Hamilton County Schools for her support, encouragement, and allowing me the flexibility I needed to complete this project. I would also like to thank my colleague, Jan Lewis, for her support and encouragement. I have learned so much from you both. I will truly miss you when you decide to retire. I would also like to acknowledge other colleagues for their support including Celeste Hochrein, Pam Blackwell, and Kathy Petros. Thanks also to Irise Chapman, Director of Exceptional Education for the Hamilton County Schools, and Pamela Hudson, Supervisor of Administrative Services for Hamilton County Schools, for your support and flexibility during the completion of this project.

ABSTRACT

The relationship between two tests of cognitive ability measured nonverbally and the relative capability of intellectual constructs measured by those tests to predict academic achievement (operationalized by end of the year group achievement tests) was examined. One hundred elementary and middle school students were administered the Universal Nonverbal Intelligence Test (UNIT) and the Leiter International Performance Scale – Revised in counter-balanced order; reading, math, and language scores from Terra Nova were matched with 37 cases in the sample. Correlation coefficients describing the relationship among global scores within the UNIT and between the UNIT and the Leiter-R were statistically significant ($p < .001$) and ranged from .33 for the UNIT Memory Quotient/Leiter-R Fluid Reasoning scores to .90 for the UNIT Full Scale IQ/UNIT Reasoning Quotient and UNIT Nonsymbolic Quotient/UNIT Reasoning Quotient. The coefficient between the UNIT and Leiter-R Full Scale IQ scores was .72 ($p < .001$). Mean differences between the UNIT and Leiter-R Full Scale were significant with the UNIT Full Scale IQ score being approximately five point higher than the Leiter-R, $t = 4.73$, $p < .001$. Effect size for the *t*-test was modest, .35.

Based on stepwise multiple regression analyses, the UNIT Full Scale IQ predicted all three areas of academic achievement significantly better than the Leiter-R Full Scale IQ score, with the variance accounted for by the UNIT Full Scale IQ score ranging from 39 percent to 55 percent ($p < .01$). The Leiter-R contributed an additional 2 percent of variance. In addition, a number of the UNIT and Leiter-R global scores were statistically significant predictors of achievement (e.g., UNIT Reasoning Quotient, Memory Quotient, Leiter-R Reasoning). Results are consistent with prior research that has found the UNIT and Leiter-R to provide comparable measures of general intelligence. However, this is the first study to suggest that the UNIT may be superior to the Leiter-R in its relative capability to predict academic achievement. School psychologists and administrators will find these results useful in choosing assessment instruments to evaluate the increasingly culturally and linguistically diverse population of students.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Introduction

Purpose

This study has two primary purposes. The first purpose is to examine to the concurrent validity of the *Universal Nonverbal Intelligence Test* (UNIT; Bracken & McCallum, 1998) and the *Leiter International Performance Scale – Revised* (Leiter-R; Roid & Miller, 1997) with a sample of nondisabled, English fluent, students from the majority culture. The second purpose is to examine the extent to which various intellectual subconstructs assessed by these instruments predict reading, math, and language scores on end of the year state-mandated groups achievement tests.

Rationale

Within the last decade, there has been a significant increase in research involving nonverbal measures of intelligence. This has resulted in the publication of several new instruments purporting to assess intelligence utilizing nonverbal techniques (McCallum, Bracken, & Wasserman, 2001). These include the *Universal Nonverbal Intelligence Test* (UNIT; Bracken & McCallum, 1998), the *Leiter International Performance Scale-Revised* (Leiter-R; Roid & Miller, 1997), the *Test of Nonverbal Intelligence-Third Edition* (TONI-III; Brown, Sherbenou, & Johnson, 1997), and the *Comprehensive Test of Nonverbal Intelligence* (CTONI; Hammil, Pearson, & Wiederholt, 1996). All of these instruments use a completely nonverbal administration format as compared to previous attempts to assess intelligence nonverbally using the "nonverbal scales" from traditional intelligence tests such as the *Stanford-Binet-Fourth Edition* (Stanford-Binet-IV), *Wechsler Intelligence Scale for Children-Third Edition (*WISC-III), and the *Differential Ability Scales (*DAS). Because these tests still require the examinee to understand complex verbal directions, they can be at best classified as "language reduced" rather than true nonverbal instruments (McCallum, Bracken, & Wasserman, 2001).

Although various methods of assessing cognitive abilities nonverbally or in a language-reduced format have been used for some time, several new (UNIT, CTONI) or significantly revised tests (Leiter-R) have recently been published. The publication of these instruments has corresponded to a dramatic increase in the number of ethnic/cultural minority students in U.S. public schools. IDEA 1997 projected that one out of every three persons in the U.S. will be a member of an ethnic/cultural minority by the year 2000. This rise in the number of students from diverse linguistic and cultural backgrounds has provided an impetus for improved instruments for the assessment of cognitive abilities of these students. A second factor that has necessitated advances in instruments that minimize the influence of prior cultural and linguistic experiences is the focus on the disproportionate representation of minority children in special education programs. As a result of investigations by the federal Office of Civil Rights and the publication of the National Research Council Panel Report (2002) on minority students in special and gifted education programs, numerous states and local education agencies have developed procedures to conduct comprehensive psychoeducational evaluations in a manner that considers a student's cultural and linguistic background and experiences in determining the presence of a learning disability or mental retardation. The purpose of these procedures is to reduce the overidentification of African-American students with disabilities that are primarily the result of cultural and linguistic differences, and/or a lack of appropriate early reading instruction. A second purpose is to improve the underidentification of minority students for gifted and talented programs.

One method that has been widely employed to address these concerns is to use nonverbal measures of intelligence with this group of students. In order to validate the use of nonverbal measures of intelligence as a primary cognitive measure, it is necessary to determine if these instruments demonstrate validity in measuring general cognitive ability and if they demonstrate a relative capability to predict academic achievement as determined by state-mandated group achievement tests. Additionally, before a test can be determined to be a culturally/linguistically "fair" test, it must first be found to be a good test. Thus, nonverbal measures of intelligence need "to be evaluated in terms of typical psychometric criteria, as well as on the basis of their usefulness with nonmajority groups" (Athanasiou, 2000, p.214).

Historical Context of Intelligence Testing

While the roughly 100-year history of scientifically studying intelligence has generated much excitement among psychologists attempting to unravel exactly what intelligence is, the topic has always been controversial (Brody, 1999). To date, there is no consensus on the definition of intelligence, or certainty regarding its source. From early on, different investigators have emphasized different elements of intelligence in their definitions including the ability to think abstractly, the ability to respond well to questions, or to problem solve. Even today, not all psychologists warmly accept the proposal that it is possible to measure or study intelligence in a meaningful way (Adelson, 1996).

Today's psychologists who do accept the challenge of developing a working model of intelligence often subscribe to one of three prominent research traditions that have evolved largely independently of each other. The psychometric, information processing, and cognitive modifiability approaches have been identified as the most prominent conceptual models for the measurement of intelligence (Taylor, 1993; cited in McGrew & Flanagan, 1998). The psychometric or structural approach has identified that intelligence or "IQ" is not a single entity, but has a complex structure of several fundamental and secondary dimensions. These dimensions or constructs can be measured by psychological tests that yield quantitative scales and are amenable to analyses by correlative and factor-analytic techniques. These analyses allow for the identification of the dimensions that underlie the structure of individual differences in cognitive abilities (McGrew & Flanagan, 1998).

Information processing theories originated within the last 40 years, and are based largely on a computer analogy of humans as information processors. Information processing theories are often considered "limited capacity" theories that focus on how efficiently information is processed in order to solve problems and perform everyday tasks.

Theories of cognitive modifiability have focused primarily on the individual's ability to adapt to the circumstantial demands of various environments. Successful adaptation requires changing behavior in order to cope with new environmental circumstances, but may also require either altering the present surroundings or locating another environment altogether. The major idea in such a model is that intelligence is dynamic, modifiable, and changeable (Lidz, 1991, Feurerstein, Rand, & Hoffman, 1979).

The primary thrust in defining intelligence, then, is to avoid labeling it as strictly cognitive or to overemphasize a general intelligence factor (g) that is often expressed practically in an IQ score. Instead, intelligence might be best viewed as a selective combination of multiple processes intentionally employed for successful functioning across a range of tasks and environments. At the very least, then, intelligence probably comprises a number of mental processes, including elements of sensation, perception, memory, reasoning, learning, and problem solving.

This relatively simple idea has often eluded empirically minded psychologists, explaining why so much of the history of the intelligence field has been plagued with controversy. The dissatisfaction with traditional approaches to constructing theories of intelligence has led to such recent developments as Gardner's (1983, 1993) "multiple intelligence" (MI), Goleman's (1995, 1998, 1999) "emotional intelligence" (EI) theories, and Sternberg's (1997) triarchic theory of intelligence which he now commonly refers to as "practical or successful intelligence". A recent addition to the evolving body of theories of intelligence is the Cattell-Horn-Carroll (CHC) theory of cognitive abilities. CHC theory represents a convergence of two psychometric intelligence taxonomies developed through factor analytic research conducted over the past 50 to 60 years (McGrew & Flanagan, 1998).

The Cattell-Horn Gf-Gc theory traces its origins back over 60 years to Cattell's (1941) initial description of fluid intelligence and crystallized intelligence. Horn's subsequent systematic Gf-Gc research has resulted in the identification of nine broad cognitive abilities (Horn, 1991, Horn & Noll, 1997). CHC theory is also rooted in part in the seminal factor-analytic work of John B. Carroll (1993, 1997). By reanalyzing 460 data sets from previous studies on the structure of intelligence, Carroll posited a three-stratum model of cognitive abilities. In his model, Carroll identified a general cognitive ability factor or *g* at stratum III of the model. This *g* factor subsumes eight broad cognitive abilities within stratum II, which then encompasses approximately 70 narrow abilities at stratum I. Within the past several years, Drs. Horn and Carroll have agreed to refer to the convergence of "modern Gf-Gc theory" (Horn, 1994) and the Gf-Gc based three-stratum theory (Carroll, 1993) as the Cattell-Horn-Carroll Theory of Cognitive Abilities, or simply, CHC Theory (Flanagan & Ortiz, 2001). Even CHC theory is beginning to garner significant criticism as the literature describing it becomes more bountiful. The level of criticism has increased significantly with the publication of the *Woodcock-Johnson Tests of Cognitive Abilities-Third Edition* (WJ-III) (Woodcock, McGrew & Mather, 2001), which is the most concerted attempt yet to operationalize CHC theory in a comprehensive measure of intelligence.

As is often the case with most scientific theories, the assorted views of intelligence have evolved over a succession of major paradigms. These "paradigm shifts" include models within psychometric psychology, cognitive psychology, cognitive-contextual psychology, and physiological science.

Psychometric Theories

Psychometric theorists study the structure of intelligence: its parts and forms. Information for this model of intelligence has traditionally been based on data obtained from paper-and-pencil and individually administered tests of mental abilities, especially classifications, categories, and analogies. Here, intelligence is seen as a composite of abilities as measured by cognitive tests, and is quantified by assuming that test scores mathematically represent underlying mental abilities (Brody, 1999, Carroll, 1993).

One of the first psychometric theorists was Spearman (1927), who devised a statistical technique — factor analysis — to study patterns of individual differences in test scores, as well as provide an explanation for the underlying sources of these differences. Spearman concluded that only two types of factors underlie all individual differences in test scores: the "general factor," or "g" (pervading performance on all intelligence tasks), and a "specific factor" needed to complete each particular test (Jensen, 1998). Understanding the exact nature of "g" has eluded most psychologists, although Spearman labeled it as being akin to "mental energy."

Thurstone (1960) disagreed not only with Spearman's theory but also with his isolation of a single factor of general intelligence. Rather, he proposed several primary mental abilities: verbal comprehension and fluency, memory, spatial visualization, use of numbers, inductive reasoning, and perceptual speed.

Cattell (1971) offered another explanation: that both Spearman and Thurstone were correct in the sense that intellectual abilities are hierarchical, with "g" at the top and increasingly specific abilities below (Jensen, 1998). To that end, Cattell (1971) indicated general ability further subdivides into "fluid" and "crystallized" intelligence. Fluid intelligence includes those abilities that are applied to new learning, including the speed and effectiveness of memorizing, inductive reasoning, and perception of new relationships (Horn, 1982). In contrast, crystallized intelligence comes with experience and education, and involves the ability to use learned strategies to solve new problems, find relationships, and make judgments (Horn, 1982). Fluid ability increases in earlier years and decreases in later ones, while crystallized ability increases over the life span. Through a continued program of rigorous research, Horn (1991) and Horn and Noll (1997) came to build on the work of Cattell to develop what came to be known as contemporary Gf-Gc theory. This theory recognized nine broad cognitive ability factors that include and expand upon the original two factor Gf-Gc model. The 9 broad abilities identified by Horn (1991) and Horn and Noll (1997) include: Fluid Intelligence (Gf), Crystallized Intelligence (Gc), Short-Term Acquisition and Retrieval (Gsm), Visual Intelligence (Gv), Auditory Intelligence (Ga), Long-Term Storage and Retrieval (Glr), Cognitive Processing Speed (Gs), Correct Decision Speed (CDS), and Quantitative Knowledge (Gq).

Most researchers in recent decades have agreed that there are more than Spearman's two kinds of intellectual abilities, but disagree on just how many and what types there are. For example, Guilford (1967) originally suggested 120 abilities, and even later increased that number to 150. What had begun as a single factor "g" had now blossomed into 150 different factors. Such confusion brought problems for the psychometric model, particularly given its lack of scientific parsimony, as well as psychometric theorists' inability to explain the processes underlying intelligence (Jensen, 1998).

Psychometric theories reached an important landmark in 1993 with Carroll's "Three-Stratum Theory of Cognitive Abilities", which has been lauded by a variety of researchers as the most ambitious attempt to develop a complete (no model can truly every be said to be complete) taxonomy of cognitive abilities. In addition to a "general" factor of intelligence *g*, Carroll identified eight broad cognitive abilities that are very similar to those described by Horn. They include Fluid Intelligence (Gf), Crystallized Intelligence (Gc), General Memory and Learning (Gy), Broad Visual Perception, Broad Auditory Perception, (Gu), Broad Retrieval Ability (Gr), Broad Cognitive Speediness (Gs), and Processing Speed/Reaction Time Decision Speed (Gt). These eight broad abilities subsume approximately 70 narrow abilities at stratum I. McGrew and Flanagan (1998) regarded Carrroll's model as a sort of "periodic table of elements" for cognitive abilities that "organizes cognitive ability at three strata that differ as a function of breadth or generalizability of abilities.

The convergence of Cattell-Horn Gf-Gc theory and Carroll's three-stratum theory of cognitive abilities represent the most current and complete theory of cognitive abilities that has evolved from the psychometric tradition. Flanagan and Ortiz (2001) reported, "in general, the CHC theory is based on a more thorough network of validity evidence than are other contemporary multidimensional ability models of intelligence (p. 8). Although similar in their treatment of the factorial structure of intelligence, the Carroll and Cattell-Horn models differ in several aspects. The most prominent difference is in each model's treatment of a general or *g* factor of intelligence at the apex of their respective factorial structure. Carroll (1993) defines the general intelligence factor at the apex of his three-stratum model as being analogous to Spearman's *g*. However, Horn does not identify a general intelligence factor that subsumes the broad Gf-Gc abilities in his contemporary Cattell-Horn model (Horn, 1991; Horn & Noll, 1997).

<u>Cognitive Theories</u>

Without an understanding of the mental processes underlying general intelligence (i.e., including both verbal and nonverbal intelligence), it is possible to come to misleading, if not wrong, conclusions when evaluating assessments of performance and overall test scores. Hence, cognitive psychologists propose that basic to most cognitive approaches to intelligence is the assumption that intelligence comprises a set of mental processes, not separate intelligences, acting upon mental representations. Consequently, a number of cognitive theories of intelligence have evolved that claim basic mental processes — such as the ability to remember names or recall numbers in sequences — might be the <u>building blocks</u> of general intelligence, especially when speed is taken into account (Anderson, 1988, 1992).

Newell pursued a different path in the study of human intelligence, including designing computer models of human cognition (Newell & Bickerton, 1992). Beginning in the late 1950s, he constructed a computer model of human problem solving called the "General Problem Solver," which relied heavily on a heuristic procedure termed "means-ends analysis."

Expanding upon such computer models, other researchers have described a "parallel processing" model of nonverbal and verbal intelligence (Farah & McClelland, 1991; Rumelhart, Widrow, & Lehr, 1994). This "connectionist" view of cognition holds that people are able to process multiple sources of information at the same time. Of course, this view does not imply separate intelligences, even though people are capable of concentrating on more then one task at a time.

Cognitive-Contextual Theories

Cognitive-contextual theories deal with the way that cognitive processes operate in various environmental contexts. Perhaps the best known of these theories is that of Gardner (1983, 1993), who built on the idea of parallel/multiple processing and proposed the theory of "multiple intelligences." Gardner challenged earlier theories that intelligence is comprised of one or multiple general abilities. He argued that there is no such entity as single intelligence. Instead, "intelligence" involves multiple linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, intrapersonal, and possibly other "intelligences." Gardner compiled his listing of intelligences from a variety of sources, including studies of exceptional persons, brain damaged persons, memory and cognitive processing, and cultural differences and similarities.

Not as well-known in popular circles, but more empirically validated than Gardner's work, theories of cognitive modifiability have focused primarily on the individual's ability to adapt to the circumstantial demands of various environments. Rooted in the work of Vygotsky and Feurerstein, theories of cognitive modifiability attempt to explore a "zone of proximal development." According to Lidz (1997), the zone of proximal development is found at the nexus of the “inner mental world of the child and the external influences of the sociocultural environment” (p. 282). The major idea in such a model is that intelligence is dynamic, modifiable, and changeable (Lidz, 1991, 1997; Feurerstein, Rand, & Hoffman, 1979; Feuerstein, Feuerstein, & Gross, 1997). The theory is operationalized in “dynamic assessment” approaches that allow and examiner to evaluate how a person learns novel tasks through what Feurerstein calls a “mediated learning experience”. While there is no specific package of materials that defines dynamic assessment techniques, Feuerstein's *Learning Potential Assessment Device* (LPAD) is one of the most widely used instruments by proponents of the theory of cognitive modifiability and it's assessment.

Physiological Theories

The theories discussed above seek to understand intelligence in terms of underlying hypothetical and rather abstract constructs. On an entirely different front, other researchers have sought to explain intelligence as a function of biochemical processes alone rather than intervening hypothetical constructs. This line of research was pioneered by Donald Hebb who, in opposition to radical behaviorism, sought to understand the process that occurs between stimulus and response (Klein, 1999). Hebb believed psychology to be a biological science and proposed a neuropsychological cell assembly of cognitive processes. His ideas have influenced later research in the areas of cognitive science, neuroscience, and cognitive neuroscience.

While such reductionism may have a philosophical appeal, most psychologists dismiss simplistic explanations for complex phenomena. Indeed, biochemical approaches to intelligence should be seen as complementary to, but not replacing, other approaches at this time. Although relatively little is known about the biological bases of intelligence, some progress has been made in conducting brain hemispheric, blood-flow, and brain-wave studies (Sampson, 1993; Sperry, 1993). Jensen (1981) reported correlations between a general factor of intelligence *g* and certain brain process including the speed and amplitude of evoked electrical potentials in the brain. Ittenbach, Esters, and Wainer (1997) foresee the potential of brain mapping, the recording of brain activity during the performance of particular tasks, for contributing to the assessment of cognitive ability. They also suggest not discounting the contributions and implications of behavioral genetics, and hypothesize the potential for the future of intellectual assessment to lie within the analysis of DNA.

Measurement of Intelligence: IQ Tests

Any discussion of intellectual assessment, nonverbal or verbal, would be lacking without a mention of intelligence testing ("IQ testing"). Binet and his collaborator Simon in France first developed systematic mental testing. Binet's early test was taken to the United States by Terman, whose version became known as the Stanford-Binet test, which has been repeatedly revised and continually used.

IQ tests became quite popular during World War I (1914-1918), at a time when they were employed to quickly assess and classify large numbers of men. Nonverbal (performance) IQ tests were especially important during this war because the majority of recruits were functionally illiterate. The tests also were used in peacetime on people who were non-English speaking or hearing impaired. As the primary producer of psychological tests during this early period, the Stoelting Company (founded circa 1886) is perhaps the oldest psychological testing company still in operation in America. The company's earliest products included research instruments as well as psychological tests, both mostly used by academic institutions. Stoelting also produces the Leiter International Performance Scale-Revised.

In response to criticism of the 1937 version of the Binet scale's emphasis on language/verbal skills, Wechsler included in his IQ test an entire scale that provided a measure of nonverbal intelligence into his assessment measure. This performance scale consists of tasks that require the subject to do something rather than merely to answer questions.

Nonverbal Assessment of Intelligence

There exists some disagreement regarding the use of the term "nonverbal intelligence" and "nonverbal intellectual assessment." Bracken and McCallum (1998) and McCallum, Bracken, and Wasserman (2001) use the term nonverbal intellectual assessment to describe a process of assessing general intelligence using nonverbal administration techniques. However, other authors use the term nonverbal intelligence to describe the construct "nonverbal reasoning" or "nonverbal abilities" (Brown, Sherbenou, & Johnsen, 1997; Hammill, Pearson, & Wierderholt, 1996). This is not merely an argument about semantics, but is important in considering the appropriateness of using these tests to generalize about overall intellectual functioning. If such tests do measure a theoretical construct that is significantly different than that assessed by traditional intelligence tests, then the utility of these nonverbal tests in making educational decisions and describing overall intellectual functioning is suspect (Bracken, McCallum, & Wasserman, 2001).

In the United States, nearly 32 million Americans do not speak English as their primary language, and nearly 2 million have no English-speaking capabilities at all (U.S. Bureau of the Census, 2000). The Bureau of the Census projected that the numbers of non-English speaking Americans is increasing exponentially. McCallum, Bracken and Wasserman (2001) reported that over 200 languages are spoken by students attending the Chicago city schools. They also report a similar rise in linguistic diversity in suburban areas (e.g., Cobb County, GA) and mid-sized cities (e.g., Knoxville, TN). Both groups, children with verbal language impairments (i.e., either expressive or receptive) and nonfluent English speakers have been shown to be at a distinct disadvantage when assessed with traditional verbally-loaded intelligence tests, a phenomenon often referred to as "bias" in testing (Jensen, 1980; Naglieri & Prewett, 1999).

While often framed in the context of social, economic, and political arguments, the issue of test bias may be better understood in terms of construct irrelevance. Construct irrelevance is "the extent to which test scores are influenced by factors that are irrelevant to the construct that the test is intended to measure" (AERA, APA, NCME, 1999, pp. 173-174, as cited in Braden, 2000). Therefore, for individuals who lack English language proficiency, knowledge of the dominant culture, and/or have an interfering emotional condition, then test scores may be distorted by the introduction of these construct irrelevant factors into the assessment process. Because scores on intelligence tests are often used to make judgments about an individual's academic or occupational potential and/or to determine eligibility for special education services, distorted scores may be viewed as unfair representations of the construct (e.g., intelligence) purported to be measured by the tests. Individuals from cultural and linguistically diverse backgrounds, and persons with disabilities, often do not have equal access to the same linguistic and cultural experiences as members of the dominant culture and their nondisabled peers. Consequently, inferences that low scores on tests that contain large amounts of construct irrelevant linguistic and cultural content are indicative of significantly below average cognitive abilities may promote stereotyping or other disparate treatment of individuals with disabilities or those who are member of minority groups. An additional consequence of using tests with large amounts of construct irrelevant content to make inferences about overall cognitive ability may be the disproportionate representation of children from cultural and linguistic minority backgrounds in special education.

This is particularly problematic for the disproportionate placement of minority children in special education under the IDEA category of mentally retarded (Donovan & Cross, 2002; MacMillan & Reschly, 1998). MacMillan and Reschly (1998) reported that while African American students comprise only 17% of the population of students in general education, nearly 35% of the students in special education programs for students with mental retardation are African Americans.

The idea of bias in the referral and assessment process remains controversial; however Donovan and Cross (2002) noted that "research shows that context, including familiarity with test taking and the norms and expectations of school, may depress the scores of students whose experiences prepare them less well for the demands of classrooms and standardized tests." Criticisms regarding the overrepresentation of minority students in special education programs, and particularly the disproportionate representation of African-American students in the category of mentally retarded are becoming more frequent and focused (Patton, 1998). However, data regarding whether or not placement in special education is a benefit or a risk do not currently exist (Donovan & Cross, 2002). Nonetheless, state and local education agencies are increasingly seeking ways to reform the referral and assessment process to reduce the disproportionate representation of minority students in special education. This action is a result of increasing pressure from minority advocacy groups, investigations by the federal Office of Civil Rights (OCR) and subsequent corrective action plans, and an increased emphasis on early identification of risk factors and interventions in general education programs.

Local education agencies have responded to the issue of disproportionate representation of minorities in special education by providing staff-development to teachers, psychologists, and other personnel regarding culturally responsive instruction, and education in the beliefs, values, cultural practices, discourse styles, and other aspects of students' lives that may negatively affect academic performance in the current educational system. These types of systemic changes that entail modifying long-held beliefs and attitudes can take years to accomplish. In the interim, many systems have chosen to reform the special education assessment process to consider the factors mentioned above, and use tests that minimize the effects of construct irrelevant factors on the obtained scores. Nonverbal measures of intelligence appear to be gaining widespread acceptance as a way to minimize construct irrelevance in the assessment of minority students' cognitive abilities. Consequently, there has been an increasing demand within the field of education for intellectual assessment methods are less verbally and culturally loaded, and as such do not rely on receptive or expressive English-language skills.

There are currently two primary methods used to assess intelligence with those persons for whom cultural or linguistic difference may introduce bias into the evaluation process. First, the same intelligence tests designed for populations are fluent in English are adapted for use with other populations who are not fluent in English. Two popular examples include the Stanford Binet Intelligence Scale, 4th Edition (SB-IV; Thorndike, Hagan, & Sattler, 1986) and the Wechsler Intelligence Scale for Children, 3rd Edition (WISC-III; Wechsler, 1991), both of which contain some nonverbal assessment subtests. However, critics of these tests claim that the SB-IV's nonverbal reasoning scale and the WISC-III's nonverbal abilities assessed through its performance scale include activities that demand language-based processes for completion. For instance, the SB-IV relies greatly on language capability throughout and thus "is totally inappropriate for use with [the] hearing-impaired," as one example (Sullivan & Burley, 1999, p. 783). The same might also be said of the WISC-III.

The other method is to use specialized intelligence tests designed to assess intelligence via test items and tasks that do not require language-oriented responses. Three popular examples of nonverbal IQ tests are the Matrix Analogies Test-Expanded Form (MAT-EF) (Naglieri, 1985), Raven's Progressive Matrices (Raven, Court, & Raven, 1986), and Test of Nonverbal Intelligence (TONI-3; Brown, Sherbenou, &

Johnsen, 1990, 1997). However, each of these tests has its critics with regard to various deficiencies or psychometric flaws, including the narrow range of abilities measured (Sattler, 1992, McCallum, Bracken, & Wasserman, 2001). Indeed, the TONI-3 measures nonverbal IQ solely through logical problem solving, and both the Raven's Progressive Matrices and MAT-EF measure nonverbal intelligence with figures and matrices.

According to critics, the problem with these types of nonverbal IQ tests is their inability to evaluate intelligence as comprehensively as general intelligence tests (Athanasious, 2000). To this end, Naglieri and Prewett (1999) recommended that, for a nonverbal IQ test to be effective, it must offer "a more complete evaluation of the cognitive processing of individuals with hearing impairment, physical limitations, limited knowledge of the English language, and language/communication disorders, as well as those of normal persons" (p. 368). With Naglieri and Prewett's (1999) comments in mind, two tests designed specifically for use without language-based skills have shown exceptional promise with respect to the nonverbal assessment of intelligence in a multi-dimensional manner: The Leiter and the UNIT (e.g., Farrell & Phelps, 2000).

Statement of the Problem

The current study will address the validity of the UNIT and the Leiter-R by determining their concurrent validity and their relative capability to predict academic achievement as measured by end of the year state-mandated group achievement tests. In their review of the UNIT, Young and Assing (2000) commented "it was difficult to thoroughly judge the concurrent validity for the UNIT when it is used with majority, nondisabled, English-speaking examinees because validity studies included only examinees from clinical/exceptional or minority samples (p. 284). They also noted a need for additional research to examine the criterion-related/predictive validity of the UNIT when used with native English-speaking students who are have not been identified with a disability and are not members of a minority group. These same criticisms could be leveled against the Leiter-R. The present study will address both of these issues by examining the concurrent and predictive validity of the UNIT and Leiter-R with a sample of native English speaking students who have not been identified with a disability.

<u>Research Questions</u>

1. What are the relationships among the various factor and composite scores on the UNIT and Leiter-R as determined by correlation coefficients?
2. Is there a significant mean score difference between the FSIQ scores on the UNIT and Leiter-R?
3. What is the relative capability of overall cognitive ability as identified by Full Scale IQ score obtained on the UNIT and Leiter-R to predict reading skills (as measured by the Total Reading score from the Terra Nova group achievement test)?
4. What is the relative capability of overall cognitive as identified by Full Scale IQ score obtained on the UNIT and Leiter-R to predict language skills as measured by the Total Math score from the Terra Nova?
5. What is the relative capability of overall cognitive ability as identified by Full Scale IQ score obtained on the UNIT and Leiter-R to predict mathematics skills as measured by the Total Math score from the Terra Nova?

6. What is the relative capability of various intellectual subconstructs as identified by composite scores obtained on the UNIT (Memory Quotient, Reasoning Quotient, Symbolic Quotient, Nonsymbolic Quotient), and Leiter-R (Fluid Reasoning) to predict reading skills as measured by the Total Reading score from the Terra Nova?
7. What is the relative capability of various intellectual subconstructs as identified by composite scores obtained on the UNIT (Memory Quotient, Reasoning Quotient, Symbolic Quotient, Nonsymbolic Quotient), and Leiter-R (Fluid Reasoning) to predict math skills as measured by the Total Math score from the Terra Nova?
8. What is the relative capability of various intellectual subconstructs as identified by composite scores obtained on the UNIT (Memory Quotient, Reasoning Quotient, Symbolic Quotient, Nonsymbolic Quotient), and Leiter-R (Fluid Reasoning) to predict Language skills as measured by the Total Language score from the Terra Nova?

CHAPTER 2

Methods

Participants

Data for the concurrent validity component of this study were obtained from an existing data set provided by Riverside Publishing Company and Measurement/Learning/Consultants, LLC. The data set contains one hundred students in the participating sample from grades one through eight. Participants were selected from elementary and middle schools in East and Southeast Tennessee. For the predictive analyses, a subsample of thirty-eight students from a public middle school in will be used. Cases in the subsample were matched for analysis by the school in cooperation with Riverside before being provided for the current investigation. Therefore, no identifying information was contained in the data set analyzed.

Instruments

Universal Nonverbal Intelligence Test (UNIT)

Distributed by Riverside Publishing, the Universal Nonverbal Intelligence Test (UNIT) is designed to provide a culturally fair, standardized, and comprehensive measure of general intelligence based on entirely nonverbal response and administration protocols (Bracken & McCallum, 1998; Farrell & Phelps, 2000). In other words, the UNIT measures a wide array of complex reasoning and memory abilities, including those involving internal processes of verbal ("symbolic") mediation, as well as those not involving such mediation ("nonsymbolic"). Developed for use with children ranging age from 5 through 17 years, the UNIT effectively assesses IQ in persons who are verbally noncommunicative; who have language, speech, or hearing impairments; and who come from non-English language or cultural backgrounds.

One primary goal in creating the UNIT was to ensure impartiality for all examinees regardless of gender, language, race, ethnicity, or hearing status (Bracken & McCallum, 1998). In contrast to most nonverbal IQ tests that are based exclusively on matrices, the UNIT's subtests uses multiple response modes (e.g., pencil and paper activities, pointing, touching). Special studies are reported in the manual describing use of the UNIT with African Americans, Asians, Hispanics, Native Americans, and individuals with limited English proficiency.

Each of the six subtests yield scaled scores, each with a mean of 10 and a standard deviation of 3. Furthermore, there are 5 quotient scores available, each with a mean of 100 and a standard deviation of 15: Full Scale Intelligence Quotient (FSIQ), Reasoning Quotient (RQ), Memory Quotient (MQ), Symbolic Quotient (SQ), and Nonsymbolic Quotient (NSQ). As such, the UNIT can provide diagnostic information relevant to common educational differences, including learning disabilities, mental retardation, and giftedness.

<u>Normative Data.</u> Standardized through a carefully designed stratified random sampling plan, the UNIT resulted in a sample that closely matches the U.S. population according to 1995 census data. Normative data were acquired from a large U.S. sample of 2,100 children and teens between the ages of 5 years 0 months through 17 years 11 months. Researchers collected data in 108 sites across 38 states. The following 9 variables were used to select participants for the standardization sample: age, gender,

race/ethnicity, Hispanic origin, community setting, region, classroom placement, special education services, and parental level of educational. A total of 3,865 children and teens were tested (when subjects from validity, reliability, and cultural relativity studies are included in the standardization sample) (Bracken & McCallum, 1998; Reed & McCallum, 1995).

Reliability. Reliabilities for the UNIT are reportedly high, and meet or exceed the usual technical standards often quoted by social scientists for both clinical and standardization samples. Reliability figures for subtests are reported in the UNIT manual as follows: Symbolic Memory .85, Cube Design .91, Spatial Memory .81. Analogic Reasoning .79, Object Memory .76, and Mazes .64. Average reliability coefficients for the scales are as follows: Extended Battery Memory .90. Extended Battery Reasoning, .86, Extended Battery Symbolic .89, Extended Battery Nonsymbolic .87; and Full Scale, Abbreviated Battery FSIQ .91, Standard Battery FSIQ .93, and Extended Battery FSIQ .93 - all of which are positive and high (Bracken & McCallum, 1998, McCallum, Bracken, and Wasserman, 2001).

Validity. Validity studies on the UNIT have included exploratory and confirmatory factor analyses, which provide consistent support for the structure of the UNIT. Validity studies show strong concurrent and predictive validity with many other measures of intelligence (e.g., the WISC-III), and the UNIT appears to be a good predictor of eventual educational success (Reed & McCallum, 1995). In addition, evidence of discriminant validity demonstrates that the UNIT distinguishes between students with mental retardation, learning disabilities, speech-language impairments, or those that are gifted (Bracken & McCallum, 1998).

Leiter International Performance Scale-Revised

This new revision of the Leiter, which was in use for over 50 years (Radcliff & Radcliff, 1979), is a nonverbal measure of intelligence that requires no writing or speaking from either the examiner or examinee (Roid & Miller, 1997). Claiming to be "culture free," this untimed (except for bonus points for speedy completion of items on several subtests), individually administered test examines conceptual ability but does not require speech responses. During testing, the child solves puzzle-like problems applying spatial, visual, and linguistically mediated logical reasoning. The Leiter-R is fairly quick to administer (the manual estimates 40 minutes for the six subtests on the Visualization and Reasoning Battery), but the child must exhibit some sustained attention. The Leiter-R also yields standard scores beyond the range of other IQ tests: 30 to 170. It spans ages 2 years 0 months to 20 years 11 months, exceeding that of the CTONI, TONI-2, and UNIT, and it fully covers the ranges of the WISC-III and WPPSI-R (Moore, O'Keefe, & Lawhon, 1998). The test also consists of 2 nationally standardized batteries: 1) new Attention and Memory domains, and 2) a revision of the original Visualization and Reasoning domains.

The Leiter-R offers a complete cognitive profile developed around the hierarchical models of Gustafson (1984), Woodcock (1990), Carroll (1993) and others. In fact, Roid and Miller (1997) indicated that the work of Gustafson and Carroll were carefully studied during the design phase of the Leiter-R development. The test is comprised of 20 subtests and numerous composites that measure both general intelligence and secondary cognitive ability areas. The subtests fall into these broad areas: Reasoning, Visualization (spatial), Memory, and Attention. The Visualization and Reasoning Battery yield composite scores that include a Full Scale IQ score, a Brief IQ screen, Fluid Reasoning, Fundamental Visualization (ages 2-5), and Spatial Visualization (ages 11-20). The Attention and Memory Battery yields a Memory Screen composite and five other composite memory scores. Subtests on both batteries yield standard scores (mean = 10, sd = 3), which when combined are used to arrive at a composite IQ score (mean = 100, sd = 15).

In contrast to most other verbally laden IQ tests, the Leiter-R emphasizes fluid intelligence (Athanasiou, 2000). There is some evidence to suggest that Leiter-R is a fairer instrument because the quality level of an examinee's academic, family, or social experience does not appreciably influence the IQ score. For instance, on the Leiter-R, scores for English as Second Language (ESL) persons and hearing impaired persons are reported at .33 sd's below the mean, compared to typical findings of a full SD difference for these groups on verbally-oriented tests (Roid & Miller, 1997). Moreover, because the Leiter-R measures fluid intelligence, which does not change considerably in the adult years, it can be used effectively with older subjects.

The Leiter-R consists of two nationally standardized batteries: 1) a revision of the original Visualization and Reasoning (VR) domains for measuring IQ, and 2) the new Attention and Memory (AM) domains. Included in these batteries are novel "growth" scores that discriminate small improvements in children with significant cognitive disabilities. The Attention and Memory Battery may have some utility in distinguishing typical children from those with attention deficit/hyperactivity disorder (ADHD) or other neuropsychological deficits. However, the validity of these claims has not been adequately evaluated.

Normative Data. With a stratification plan based on 1993 census U.S. statistics, the Leiter-R was normed on 1,719 normal children and teens, as well as 692 children representing 9 clinical groups, all ranging in age from 2.0 to 20.11. The standardization was carefully constructed to accurately represent the child's age, gender, and SES. In addition, representative proportions of U.S. children who are Caucasian, Hispanic-American, African-American, Asian-American, and Native American were included. Psychometric studies on Native American, Hispanic, and African-American groups have shown the Leiter-R to evince cultural fairness for all represented backgrounds (Roid & Miller, 1997). For example, Navajo children averaged 98.0 on the Leiter-R brief IQ, while WISC-III averaged 84.5.

Reliability. Extensive studies of internal consistency, test-retest and decision-consistency reliability are reported in the Leiter-R's test manual (Roid & Miller, 1997). In addition to test information curves based on item-response theory, subtest and composite IQ scores demonstrate high levels of reliability. For example, internal consistency reliability of screening and IQ scores are shown to range from .88 to .93.

Validity. Research studies on construct, predictive, content, and criterion validity are also referenced in the Leiter-R's test manual in an extensive validity chapter. Factor analysis indicates the test fits a "g" model of intellectual abilities comprised of nonverbal memory, reasoning, and assorted attention aspects. Research into such criterion groups as children with cognitive impairment or "giftedness" has found significant differences in the anticipated direction from subjects in the normative sample (Roid & Miller, 1997).

Terra Nova

Group achievement data that will be used in the predictive validity analyses will be obtained from scores on the *Terra Nova* tests utilized by the Tennessee Department of Education as part of its Tennessee Comprehensive assessment program (TCAP). The *Terra Nova* is the fifth revision of the *Comprehensive Test of Basic Skills* (CTBS-5) which provides a norm-referenced measure of academic achievement in the areas of reading, math, language, spelling, social studies, and science (Terra Nova Technical Bulletin, 1997). It should be noted that Terra Nova is actually the name of several distinct overlapping products. Depending on the needs of the user several different batteries may be selected that include the subtests of Reading-Language Arts, Mathematics, Science, and Social Studies, Language Mechanics-Word Analysis, Vocabulary, Spelling, and Mathematics Computation. When the Survey Battery Plus product is used, composite scores are provided in reading, language, and mathematics (Nitko, 1998). Reliability coefficients for the subtests and composite scores on the *TerraNova* were consistently high in the .80's and .90's. Only the Spelling subtest was found to have consistently lower reliability coefficients. For the present study, the Total Reading, Total Math, Total Language, and Total Test Composite will be used as the criterion measures of academic achievement in the predictive validity analyses.

Procedures

Data for this study was collected by Riverside Publishing Company at several elementary and middle schools in Southeast Tennessee. The UNIT, Leiter-R, Comprehensive Test of Nonverbal Intelligence, and select subtests from the standardization edition of the Woodcock Johnson Tests of Cognitive Ability – Third Edition were administered to 100 examinees by certified school psychologists and advanced graduate students in school psychology. All examiners received training on the administration of each instrument before data collection began. The tests were administered in counterbalanced order to minimize the effects of test administration order. End of the year group achievement test scores were obtained from school records and matched with subjects' intelligence testing data by school personnel before the data was released to the investigator for the present study. The data set also contains demographic information on each student, including gender, age, race, grade in school, parental education level, and parental occupation. Only data from the UNIT, Leiter-R, and end of the year group achievement tests will be used for this study.

Data Analyses

Relevant descriptive statistics were obtained including means and standard deviations of the included variables. Correlational and mean difference analyses were performed to determine the comparability of scores obtained on the UNIT and Leiter-R. Multiple regression equations were calculated to determine the relative contributions of each of the intellectual measure's factor scores to the prediction of scores in specific academic areas.

CHAPTER 3

Results

This study addressed the validity of the UNIT and the Leiter-R by determining their concurrent validity and their relative capability to predict academic achievement as measured by end-of-the-year, State-mandated, group achievement tests. Table 1 displays the descriptive statistics for these scales. The means and standard deviations are displayed for the Leiter-R Fluid Reasoning and Full Scale IQ scores, as well as the UNIT Full Scale IQ with the Memory Quotient, Reasoning Quotient, Symbolic Quotient, and Nonsymbolic Quotient subscales. In addition, three measures of academic progress were included: the Terra Nova Total Reading, Total Math, and Total Language scores (Table 1). All these scores are similar to population parameters, i.e., the values are within a few points of population means and standard deviations. Table 2 displays the intercorrelations among selected variables, which include the seven cognitive variables, as well as the three academic performance variables.

Research Question One: Relationships Global UNIT, Leiter-R and Terra Nova Scores

The relationships between the various factor and composite scores on the UNIT and Leiter-R were determined, in part, by correlation coefficients. The UNIT Full Scale IQ score and the Leiter-R Full Scale IQ score correlated at $r = .72$. The Leiter-R Fluid Reasoning Score correlated $r = .74$ with the Leiter-R Full Scale Score. The four UNIT Quotients correlated with the UNIT Full Scale IQ, and these varied between $r = .88$ to $r = .90$ (Table 2). All of these correlations were statistically significant.

Of additional interest in Table 2 are the correlations between the seven IQ scores and the three achievement test scores. The coefficients between Terra Nova Total Reading scores and the seven IQ scores ranged from a low of $r = .54$ to a high of $r = .74$, with a median correlation of $r = .61$. For the Terra Nova Total Math score, the correlations ranged from $r = .48$ to $r = .64$, with a median correlation of $r = .58$. For the Terra Nova Total Language score, the correlations with the seven IQ scores ranged between $r = .43$ to $r = .64$, with a median correlation of $r = .56$ (Table 2). All these are statistically significant and are similar in magnitude.

Research Question Two: Examination of Mean Score Differences Between UNIT and Leiter-R Full Scale IQ Scores

To determine whether a significant mean score difference exists between the Full Scale IQ scores for the UNIT and the Leiter-R, a correlated *t* test was computed. The Leiter-R Full Scale IQ mean of 97.65, (standard deviation of 14.57) was compared to a UNIT Full Scale IQ mean of 102.90, (standard deviation of 14.80). The difference of 5.25 was statistically significant, $t\ (99) = 4.73$, $p < .001$. The effect size of .35 is considered between small and moderate using Cohen's (1988) criteria for determining the magnitude of an effect size.

A note of caution for practice is urged based on the finding of statistically significant difference of approximately five points between the mean UNIT and Leiter-R Full Scale IQ scores in this study. While this difference appears clinically small, it could have practical implications when considered in the context of special education eligibility criteria, which often rely heavily upon numerical "cutoffs" and discrepancies.

Research Question Three: Relative Predictive Capability of UNIT and Leiter-R Full Scale IQ Scores with Terra Nova Group Reading Achievement Scores

The relative capability of overall cognitive ability as identified by the Full Scale IQ scores obtained from the UNIT and the Leiter-R to predict reading skills (as measured by the Total Reading score from the Terra Nova Group Achievement Test) was determined via multiple stepwise regression analyses. Table 3 displays the stepwise regression summary of the Terra Nova Total Reading score, based on the Full Scale IQ scores. In the first step of the model predicting Total Reading, the UNIT Full Scale IQ was entered first, based on it's stronger correlation with Total Reading than the Leiter-R Full Scale IQ score. In the first model, the UNIT Full Scale IQ score was found to be significant, F (1, 35) = 42.22, p < .001. This model accounted for 54.7% of the variance in Total Reading. In the second step of the model, the Leiter-R Full Scale IQ was entered. This added only 0.60% to the variance explained, and was not significant (p = .49) (Table 3). Inspection of the partial correlations revealed the UNIT Full Scale IQ (pr = .54) to be over four times larger than the partial correlation for the Leiter-R Full Scale IQ (pr = .12). Therefore, the unique relationship between the UNIT Full Scale IQ score and the Terra Nova Total Reading was over four times stronger than the Leiter-R Full Scale IQ and the Terra Nova Total Reading score when removing the interaction effects among the variables.

Research Question Four: Relative Predictive Capability of UNIT and Leiter-R Full Scale IQ Scores with Terra Nova Group Language Achievement Scores

The relative capability of overall cognitive ability, as identified by the Full Scale IQ score obtained on the UNIT and Leiter-R to predict language skills as measured by the Total Language score from the Terra Nova was determined via stepwise multiple regression analyses. Table 4 displays the stepwise regression summary for this model. In the stepwise prediction of Total Language, the UNIT Full Scale IQ was significantly related to Total Language, $F\ (1, 35) = 22.05$, $p < .001$. In the second step of the model, the Leiter-R Full Scale IQ was entered. This added only 1.9% of the variance explained in Total Language and was not significant ($p = .30$) (Table 4). Inspection of the partial correlations found the unique relationship of the UNIT Full Scale IQ ($pr = .37$) with the Terra Nova Total Language score to be twice as strong as the Leiter-R Full Scale IQ ($pr = .18$) (Table 4).

Research Question Five: Relative Predictive Capability of UNIT and Leiter-R Full Scale IQ Scores with Terra Nova Group Math Achievement Scores

The relative capability of overall cognitive ability as identified by Full Scale IQ scores obtained on the UNIT and the Leiter-R to predict math skills as measured by the Total Math score on the Terra Nova was also examined through stepwise multiple regression analyses. Table 5 displays the stepwise regression summary for those results. In the prediction of the students' Total Math score, the UNIT Full Scale IQ was entered first, and found to be significant, $F\ (1, 35) = 24.63$, $p < .001$. In the second step of the model, the Leiter-R Full Scale IQ was entered. However, it added only 1.2% to the variance explained in the Total Math score, and was not significant (*p* = .41) (Table 5). Inspection of the partial correlations revealed that the UNIT Full Scale IQ (*pr* = .41) was almost three times larger than the partial correlation for the Leiter-R Full Scale IQ (*pr* = .14) (Table 5). This result indicates a significantly stronger unique relationship between the UNIT Full Scale IQ and the Total Math score when the interaction effects between the variables are removed.

<u>Research Question Six: Relative Predictive Capability of Intellectual Subconstructs Measured by the UNIT and Leiter-R with Terra Nova Group Reading Achievement Scores</u>

Two stepwise multiple regression analyses permitted the examination of the relative capability of various intellectual subconstructs as identified by the composite scores obtained on the UNIT (Memory Quotient, Reasoning Quotient, Symbolic Quotient, and Nonsymbolic Quotient), and the Leiter-R (Fluid Reasoning) to predict reading skills as measured by the Total Reading score of the Terra Nova. Multicollinearity was a problem with this set of independent variables, because of overlapping subtests across the various UNIT global scores. The Variance Inflation Factor (VIF) collinearity statistics ranged from 10.50 to 22.91 for the four UNIT quotient scores. Consequently, separate stepwise multiple regression analyses were conducted for the various subcomponents. First, the relative predictive power of the Leiter-R Fluid Reasoning, and the UNIT Memory and Reasoning scores was determined, followed by a second set of analyses, which evaluated the relative predictive capability of the Leiter-R Fluid Reasoning and UNIT Symbolic and Nonsymbolic scores. Additional variables were entered or removed from each model based on *F* to enter criteria of .10 and *F* to exclude criteria of .20. Tables 6 and 7 display the stepwise regression summary for these analyses.

In the first step of the model for Total Reading, the UNIT Reasoning Quotient was significantly related to total reading, $F\ (1, 35) = 38.02$, $p = .001$. In the second step of the model, the UNIT Memory Quotient added 4.6% to the variance explained. This additional variable was found to be significant at the $p = .07$ level. The third potential variable for the model, Leiter-R Fluid Reasoning, was excluded because it did not reach the $p = .10$ entry criteria (Table 6). Inspection of the partial correlation coefficients revealed that the unique relationship between the UNIT Reasoning Quotient ($pr = .59$) and the Terra Nova Total Reading score was about twice as strong as the UNIT Memory Quotient ($pr = .31$) with the Terra Nova Total Reading score.

In the second set of analyses, the relative predictive capability of the UNIT Nonsymbolic Quotient, UNIT Symbolic Quotient, and the Leiter-R Fluid Reasoning scores were examined in a series of stepwise regression equations in the prediction of Total Reading (Table 7). In the first step of the model predicting Total Reading, the UNIT Nonsymbolic Quotient was entered, and this variable accounted for 52.7% of the variance in Total Reading, $F\ (1, 35) = 39.04$, $p < .001$. Neither the Leiter-R Fluid Reasoning nor the UNIT Symbolic Quotient provided further significant explanation of the variance in Total Reading based on the entry criteria for $p = .10$ (Table 7).

Research Question Seven: Relative Predictive Capability of Intellectual Subconstructs Measured by the UNIT and Leiter-R with Terra Nova Group Math Achievement Scores

The relative capability of the various intellectual subconstructs as identified by the composite scores obtained by the UNIT (Memory Quotient, Reasoning Quotient, Symbolic Quotient, and Nonsymbolic Quotient), and the Leiter-R (Fluid Reasoning) to predict math skills, as measured by the Total Math score from the Terra Nova was evaluated by two stepwise multiple regression analyses. As in Research Question Six, VIF collinearity statistics ranging from 10.50 to 22.91 for the four UNIT scale scores suggesting strong multicollinearity. Consequently, separate stepwise multiple regression analyses were conducted for the various subcomponents. Additional variables were entered or removed from each model based on F to enter criteria of .10 and F to exclude criteria of .20. First, the relative predictive power of the Leiter-R Fluid Reasoning, and the UNIT Memory and Reasoning scores was determined, followed by a second set of analyses, which evaluated the relative predictive capability of the Leiter-R Fluid Reasoning and UNIT Symbolic and Nonsymbolic scores. Tables 8 and 9 display the stepwise multiple regression summary for these analyses.

In the first stepwise prediction of Total Math, the UNIT Memory Quotient was the first variable added into the model, and explained 33.6% of the variance in the total math score, $F(1, 35) = 17.68$, $p < .001$. In the second step of the model, the Leiter-R Fluid Reasoning was added, which accounted for an additional 15.2% of the variance in Total Math. This additional variable was significant ($p = .003$). The last potential variable of UNIT Reasoning Quotient was not added to the model due to not reaching the $p = .10$

entry criteria (Table 8). Inspection of the partial correlations revealed that both the Leiter-R Fluid Reasoning score (*pr* = .48) and the UNIT Memory Quotient (*pr* = .50) were positively related to the Terra Nova Total Math score, and the unique relationships were approximately equal when removing the interaction effects of the other variables (Table 8).

In the second set of analyses predicting Total Math (Table 9), the first step of the model included the UNIT Nonsymbolic Quotient, which accounted for 41.6% of the variance, $F(1, 35) = 24.93$, $p < .001$. In the second step of the model, the Leiter-R Fluid Reasoning Score was added, explaining an additional 5.2% of the variance, and was significant at the $p = .08$ level. The UNIT Symbolic Quotient was not included in this model (Table 9). Inspection of the partial correlations revealed that both the Leiter-R Fluid Reasoning score (*pr* = .30) and the UNIT Nonsymbolic Quotient (*pr* = .47) were positively related to the Terra Nova Total Math score with the UNIT Nonsymbolic Quotient demonstrating a significantly stronger unique relationship with the Total Math score when controlling for the interaction effects of the other variables (Table 9).

<u>Research Question Eight: Relative Predictive Capability of Intellectual Subconstructs Measured by the UNIT and Leiter-R with Terra Nova Group Language Achievement Scores</u>

The relative capability of the various intellectual subconstructs as identified by composite scores obtained by the UNIT (Memory Quotient, Reasoning Quotient, Symbolic Quotient, and Nonsymbolic Quotient), and the Leiter-R (Fluid Reasoning) to predict language skills as measured by the Total Language score of the Terra Nova was examined through two stepwise multiple regression analyses. As in previous analyses, VIF collinearity statistics ranged from 10.50 to 22.91 for the four UNIT scale scores suggesting strong multicollinearity. Consequently, separate stepwise multiple regression analyses were conducted for the various subcomponents. Additional variables were entered or removed from each model based on F to enter criteria of .10 and F to exclude criteria of .20. First, the relative predictive power of the Leiter-R Fluid Reasoning, and the UNIT Memory and Reasoning scores was determined, followed by a second set of analyses, which evaluated the relative predictive capability of the Leiter-R Fluid Reasoning and UNIT Symbolic and Nonsymbolic scores. Tables 10 and 11 display the stepwise multiple regression summary for these analyses.

In the first stepwise prediction of Total Language, the UNIT Reasoning Quotient was the first variable entered into the model, and accounted for 34.2% of the variance in Total Language, $F(1, 35) = 18.19, p < .001$. In the second step of the model, the Leiter-R Fluid Reasoning Score was added, and explained an additional 5.4% of the variance. This additional variance was significant at the $p = .09$ level. In the third step of the model, the UNIT Memory Quotient was added, and it accounted for another 5.2% of the variance in the total language score. This addition was significant at the $p = .09$ level. In the last step of the model, the UNIT Reasoning Quotient was removed, because its significance level ($p = .21$) was greater than the removal criteria of $p > .20$ (Table 10). Inspection of the individual partial correlations revealed that the Leiter-R Fluid Reasoning ($pr = .31$), UNIT Memory Quotient ($pr = .29$), and the UNIT Reasoning

Quotient (*pr* = .22) were positively related to the Terra Nova Total Language score (Table 10).

In the second set of analyses predicting Terra Nova Total Language, the UNIT Nonsymbolic Quotient was entered first (Table 11). This predictor accounted for 42.6% of the variance in the Total Language score. The remaining potential predictors, Leiter-R Fluid Reasoning and the UNIT Symbolic Quotient, were not entered into the model, as they did not significantly explain additional portions of the variance (Table 11).

CHAPTER 4

Discussion

The results of this study are similar to the results of other studies, which have found the UNIT and the Leiter-R to provide comparable nonverbal measures of intelligence (Farrell & Phelps, 2000). There are salient similarities and differences between the two tests. For example, both demonstrate ample technical qualities in terms of reliability and validity, both are multidimensional in their coverage of various intellectual subconstructs, and both use nonverbal instructional methods to communicate the directions for each subtest to the examinee. However, task demands of the subtests differ significantly, and the UNIT utilizes a much more standardized set of gestures than the Leiter-R. The standardization of gestures used in administration promotes greater consistency in evaluation procedures among different examiners with the UNIT as compared to the Leiter-R.

One of the purposes of this study was to conduct research to evaluate the concurrent and predictive validity of the UNIT as an intellectual measure with majority culture, nondisabled, English-speaking examinees, as suggested by Young and Assing (2000) in their review of the UNIT. Prior validity studies with the UNIT have included primarily subjects from clinical/exceptional or minority populations. Prepublication validity studies with the UNIT found that it correlated highly with established traditional standardized measures of general intelligence such as the WISC-III, *Woodcock-Johnson Tests of Cognitive Ability – Revised* (WJ-R-COG, Woodcock & Johnson, 1989) and the *Kaufman Brief Intelligence Test* (KBIT, Kaufman & Kaufman, 1990), and commonly used individual achievement tests including the *Woodcock-Johnson Tests of Achievement – Revised* (WJ-R-ACH, Woodcock & Johnson, 1989) and the *Peabody Individual Achievement Test – Revised* (PIAT-R, Markwardt, F.C., Jr., 1989) in samples of clinical/exceptional and minority examinees. One predictive validity study was reported by the test authors using the UNIT and the WIAT with a nondisabled, English-speaking sample. This study found significant correlations between the UNIT Full Scale IQ score and the *Wechsler Individual Achievement Test* (WIAT; Wechsler, 1992) Total Composite Score.

The Leiter-R could also be criticized because of insufficient evidence of concurrent and predictive or criterion related validity with nondisabled, English speaking examinees. One validity study with Leiter-R and the WISC-III was conducted with a sample in which only 47 percent of the subjects were identified as nondisabled or not gifted. A second study in which Leiter-R scores were correlated with archived WISC-III, and individual and group achievement test scores was conducted with a sample of 84 subjects, of which only 14 percent were identified as "typical children" (Roid & Miller, 1997, p. 181). These studies found the Leiter-R generally correlated favorably with the WISC-III and achievement measures. The Leiter-R Full Scale IQ correlated with the WISC-III Full Scale IQ at .86. Correlation coefficients for the Leiter-R Full Scale IQ and various academic achievement composites from the WIAT and WJ-R-ACH ranged from .69 to .82. However, both obtained and corrected correlations were not reported.

In the present study, correlational analyses illustrated considerable concurrent validity between the UNIT and Leiter-R. Intercorrelations between UNIT and Leiter-R

Full Scale IQ scores were significant *r* = .72. Intercorrelations between UNIT and Leiter-R composite scale scores, measuring various intellectual subconstructs, were also significant and range from .33 to .90. This was similar to the results of Farrell and Phelps (2000), who found substantial concurrent validity between the two tests in a sample of subjects identified with a language impairment. While correlations between the Leiter-R Fluid Reasoning scale and the UNIT quotient scores were significant, higher correlations were obtained between the UNIT quotient scores and the Leiter-R Full Scale IQ score. This result was also observed by Farrell and Phelps. Although, the mean Full Scale IQ scores on these tests were found to be significantly different using a paired *t*-test (Research Question 1), the effect size was only small to moderate, and the strength and direction of the correlation (Research Question 1) was large enough to suggest that the UNIT and Leiter-R provide a similar measure of general intelligence. The finding that the mean UNIT Full Scale IQ score was higher than the Leiter-R Full Scale IQ score is similar to the findings of Farrell and Phelps (2000) in direction (higher UNIT mean Full Scale IQ score), but different in that that magnitude of the difference in that study was too small to provide a statistically significant outcome. One possible explanation for the higher UNIT mean score is that the use of standardized pantomime gestures, demonstration items, and checkpoint items on the UNIT results in a better understanding of task requirements than on the Leiter-R, and thus produces a higher mean Full Scale IQ score. Both are multidimensional, and the choice between the two may depend on a host of quantitative and qualitative considerations as discussed above.

Intelligence tests have historically been used for the purpose of identifying students who may experience difficulty (or need acceleration) in their grade-level curriculum. Therefore, the predictive or criterion related validity of intelligence tests has been the subject of a multitude or research studies. Typically, new intelligence tests are not published without prior studies demonstrating the predictive validity of the instrument with widely accepted measures of academic achievement. Due to a variety of practical factors, it is not possible to demonstrate the predictive validity of a new intelligence test with all possible populations with which it may be used. Additionally, accepted scientific precepts demand replication of validity findings by test authors and publishing companies through additional research in the interest of productive academic discourse. In an era of increasing demand for valid, reliable, and culturally fair measures of general intelligence, it is imperative to examine recently published nonverbal intellectual assessment measures to determine the extent to which they exhibit concurrent and predictive validity if they are to be used to make important high stakes decisions regarding a student's educational program or disability status. Therefore, academicians and practitioners are interested in accessing research findings that may support the differential use of one nonverbal intellectual assessment measure over another.

Correlational analyses show the UNIT and the Leiter-R to correlate significantly with the Terra Nova group achievement tests in the areas of reading, math, and language. All three correlation coefficients with the UNIT were above the .60 cited by Sattler (1992) as being the typical correlation between measures of ability and achievement. For the Leiter-R, only the Terra Nova Total Reading Score was correlated at above the .60

level (.61). These are similar to the results obtained during the UNIT and Leiter-R prepublication studies (Bracken & McCallum, 1998, McCallum, Bracken & Wasserman, 2001; Roid & Miller, 1997).

Stepwise multiple regression analysis revealed the UNIT to be a significantly better predictor of reading achievement, as measured by the Terra Nova Total Reading Score, than the Leiter-R. In fact, the analyses revealed that the UNIT accounted for 55 percent of the total variance in the Terra Nova Total Reading score. The Leiter-R Full Scale IQ score increased the variance accounted for by the UNIT Full Scale IQ score by less than 1 percent (See Table 3). Similar results were obtained through stepwise multiple regression analyses with the Terra Nova Total Language and Terra Nova Total Math scores. In all three analyses, the UNIT was a significantly better predictor of academic achievement (Research Questions 3, 4 & 5) in these areas than the Leiter-R. The UNIT accounted for substantially greater variance in the academic achievement scores than did the Leiter-R. This is the first study to directly compare the predictive capability of the UNIT and Leiter-R Full Scale IQ scores and the first to show that the UNIT may be a better predictor of academic achievement than the Leiter-R.

There has been substantial interest recently in examining the relationships between intellectual subscontructs and academic achievement as described in the CHC Theory of intelligence. Both the UNIT and Leiter-R have demonstrated that they produce valid measures of overall intellectual functioning or *g* (Bracken & McCallum, 1998; McCallum, Bracken & Wasserman, 2001, Roid & Miller, 1997). The UNIT appears to provide valid measures of visual memory (MV) and reasoning (fluid reasoning-Gf, visual reasoning-Gv) with its Memory and Reasoning Quotients, while the Leiter-R provides a measure of reasoning (Gf) with its Fluid Reasoning score and primarily samples visual reasoning (Gv) with the remainder of its subtests on the IQ portion of the battery. This is consistent with the classification of the UNIT under CHC Theory by McGrew and Flanagan (1998) based on a reanalysis of the Reed and McCallum UNIT (1995) data.

Using CHC Theory model of intelligence, Evans, Floyd, McGrew, and Leforgee (2002) found that Comprehension-Knowledge (Gc) demonstrated the strongest relations with the components of reading achievement, while Short-term Memory (Gsm) demonstrated moderate relations. Moderate relations were also found between Auditory Processing (Ga), Long-Term Retrieval (Glr), and Processing Speed (Gs). This was in contrast to findings that Fluid Reasoning (Gf) and Visual-Spatial Thinking (Gv) were not significantly related to reading achievement. Previous studies have found evidence of the predictive validity of Fluid Reasoning (Gf), Comprehension-Knowledge/Crystallized Abilities (Gc), and Processing Speed (Gs) (McGrew, 1993; Williams, McCallum, & Reed, 1996). Obviously, nonverbal tests of intelligence will not contain measures of all the CHC components; however, the UNIT and Leiter-R are the most inclusive of the nonverbal tests currently available. Further research to investigate the relative predictive utility of constructs measured by these tests would be helpful. Even so, much disagreement remains in the academic community regarding the validity of using intellectual subscontructs to predict achievement and proscribe specific courses of intervention, and some believe these procedures to be useless (Watkins, Youngstrom, & Glutting, 2002; Gresham & Witt, 1997).

Multiple regression analyses of the four UNIT quotient scores (Memory, Reasoning, Symbolic, and Nonsymbolic) and the Leiter-R Fluid Reasoning score were divided into two separate analyses for each area of academic achievement (Reading, Math, and Language) because multicollinearity was a problem in conducting multiple regression analyses with these independent variables. The UNIT Memory and Reasoning Quotients were significant predictors of reading achievement, while the Leiter-R Fluid Reasoning score did not contribute a significant amount of additional predictive power in this model. Partial correlation coefficients did reveal positive relationships between the UNIT Memory and Reasoning Quotient and reading achievement. The UNIT Nonsymbolic Quotient was also a significant predictor of reading achievement, while the UNIT Symbolic Quotient and Leiter-R Fluid Reasoning score did not add significantly to the predictive power of the UNIT Nonsymbolic Quotient alone.

The UNIT Memory Quotient was found to be a significant predictor of math achievement, accounting for 33.6 percent of the variance in the Terra Nova Total Math score. The addition of the Leiter-R Fluid Reasoning score to the model was significant and increased the explanatory power of the model by 15.2 percent. The UNIT Memory Quotient and Leiter-R Fluid Reasoning score appear to have a relatively equal unique relationship with math achievement, based on inspection of the partial correlation scores.

The UNIT Nonsymbolic Quotient was a significant predictor of math achievement, accounting for 41.6 percent of the variance in the Terra Nova Total Math score. The addition of the Leiter-R Fluid Reasoning score to the model was significant, but only increased the explanatory power of the model by 5.2 percent.

The UNIT Reasoning Quotient was also found to be a significant predictor of language achievement, accounting for 34.2 percent of the variance in Terra Nova Total Language Scores. Memory and reasoning abilities as measured by the UNIT and Leiter-R also appear to contribute additional explanatory power to the prediction model. The UNIT Nonsymbolic Quotient was a significant predictor of language achievement in this model, accounting for 42.6 percent of the variance in Terra Nova Total Language scores.

The finding that the UNIT Symbolic Quotient was not a significant predictor of language achievement is surprising and somewhat counterintuitive. The subtests that comprise the UNIT Symbolic Quotient (Symbolic Memory, Analogic Reasoning, and Object Memory) contain stimulus materials that are meaningful and it is thought that these tasks may be mediated by "private speech". Therefore, it would seem reasonable that there would be a stronger positive relationship between the UNIT Symbolic Quotient and language achievement (on the Terra Nova Total Language score) than between the UNIT Nonsymbolic and Terra Nova Total Language score. Perhaps some unique characteristics of the small sample size introduced unforeseen variance into the predictive models (e.g., overall above average Terra Nova scores). However, the UNIT's rather novel assessment of symbolic/nonsymbolic processing did not yield significantly different means (Symbolic Quotient vs. Nonsymbolic Quotient) in a sample of subjects with a language impairment (Farrell & Phelps, 2000). Perhaps, the Symbolic/Nonsymbolic dichotomy used to interpret the UNIT results is not meaningful. In fact, examinees may mediate Nonsymbolic tasks using "private speech" to the same degree as for Symbolic tasks. Another explanation may depend on the nature of the criterion variable. The Terra

Nova Total Language score primarily incorporates language mechanics, sentence structure, and editing skills, which may not be sensitive to symbolic language mediation. Additional investigation of the diagnostic and predictive utility of Symbolic/Nonsymbolic dichotomy would be helpful.

Implications

As the population of school-children increasingly becomes culturally and linguistically diverse, efforts to adopt assessment practices which minimize the effects of construct irrelevant bias in the assessment of cognitive ability will likely also increase. Therefore, the need for intellectual assessment instruments that have been shown to minimize construct irrelevant bias will be in increasing demand. Based on the results of this study, both the UNIT and Leiter-R can be recommended as valid and reliable measures of general intelligence assessed nonverbally. However, it does appear that the UNIT may be a significantly stronger instrument in predicting academic achievement than the Leiter-R with nondisabled, English speaking examinees. Therefore, as school districts and school psychologists seek to expand their repertoire of assessment instruments to include nonverbal intellectual assessment tests, they may prefer the UNIT as part of a comprehensive psychoeducational battery for evaluating culturally and linguistically diverse students.

However, as has been noted in previous studies a fairly large percentage of unique variance in academic achievement scores is not explained by IQ scores. Unique variance in academic achievement that was not accounted for by the UNIT Full Scale IQ scores in multiple regression analyses ranged from 46 to 63 percent. This is consistent with studies that examined the relative predictive capability of traditional verbally laden tests, with measure of academic achievement which have found 35 to 75 percent of variance unaccounted for by IQ scores (Flanagan, Andrews, & Genshaft, 1997; Sattler, 1992; Jensen, 1980). However, this was significantly greater than the unique variance accounted for by the Leiter-R Full Scale IQ score in these analyses. The addition of the Leiter-R Full Scale IQ score to the regression analyses yielded little additional explanatory power in the various prediction models. These data further support the use of the UNIT over the Leiter-R for examining expected achievement outcomes based on Full Scale IQ scores.

Limitations and Additional Research

The findings of this study are consistent with other studies examining the concurrent validity of the UNIT and Leiter-R in certain clinical/exceptional populations. However, this is apparently the first study to examine both the concurrent and predictive validity of these two instruments. The generalizability of these results may be limited due to the small sample size, particularly the small size of the subsample used in the predictive analyses. Restricted geographic representation may also be affect the generalizability of these results as all subjects were recruited based on availability and parental consent from elementary and middle schools in East and Southeast Tennessee. Therefore, replication of these findings using a larger national sample is recommended.

Future research may examine the diagnostic and predictive utility of the specific intellectual subconstructs measured by the UNIT and Leiter-R in terms of CHC Theory with a larger sample. One analysis of interest, evaluation of the factor structure, could be used to reexamine the structure of both instruments in terms of the increasingly popular

CHC Theory. Findings from such an analysis may assist in validating the current CHC factor structure that has been proposed for these instruments by McGrew and Flanagan (1998) based primarily on expert opinion and limited statistical evidence. The resulting factors that emerge from this analysis may be of value in examining the predictive validity of the intellectual subconstructs measured by these instruments.

REFERENCES

REFERENCES

Adelson, J. (1996). Up with feelings: Emotional intelligence by Daniel Goleman. Commentary, 101(1), 59-60.

Anderson, M. (1988). Inspection time, information processing, and the development of intelligence. British Journal of Developmental Psychology, 6, 43-57.

Anderson, M. (1992). Intelligence and development: A cognitive theory. New York: Blackwell.

Athanasiou, M. S. (2000). Current nonverbal assessment instruments: A comparison of psychometric integrity and test fairness. Journal of Psychoeducational Assessment, 18(3), 211-229.

Braden, J. P. (2000). Editor's introduction: Perspectives on the nonverbal assessment of intelligence. Journal of Psychoeducational Assessment, 18(3), 204-210.

Bracken, B. A., & McCallum, R. S. (1998). Universal Nonverbal Intelligence Test examiner's manual. Itsaka, IL: Riverside.

Brody, N. (1999). What is intelligence?. International Review of Psychiatry. 11, 19-25.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1990). Test of Nonverbal Intelligence: Second edition. Austin, TX: PRO-ED.

Carroll, J. B. (1993) Human cognitive abilities. New York: Cambridge University Press.

Caruso, J. C., & Cliff, N. (1999). The relative properties of equally and differentially weighted WAIS-III factor scores. Psychological Assessment, 11, 198-206.

Caruso, J. C., & Witkiewitz, K. (2001). Memory and reasoning abilities assessed by the universal nonverbal intelligence test: A reliable component analysis (RAC) study. Educational and Psychological Measurement, 61(1), 5-22.

Cattell, R. (1971). Abilities: Their structure, growth, and action. Boston: Houghton Mifflin.

Cattell, R. B. (1982). The clinical use of difference scores: Some psychometric problems. Multivariate Experimental Clinical Research, 6, 87-98.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd Ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. Journal of Experimental Psychology, 120(4), 339.

Farrell, M., & Phelps, L. (2000). A comparison of the Leiter-R and the Universal Nonverbal Intelligence Test (UNIT) with children classified as language impaired. Journal of Psychoeducational Assessment, 18,(3), 268-274.

Flanagan, D., Andrews, T.J., & Genshaft, J. (1997). The functional utility of intelligence tests with special education populations. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), Contemporary intellectual assessment: theories, tests, and issues (pp. 457-483). New York: Guilford.

Flanagan, D., Genshaft, J., & Harrison, P. (Eds.). (1997). Contemporary intellectual assessment: Theories, tests, and issues. New York, NY: Guilford Press.

Flanagan, D., & Ortiz, S. (2001). Essentials of cross-battery assessment. New York: John Wiley & Sons.

Flemmer, D. & Roid, G. (1997). Nonverbal intellectual assessment of Hispanic and speech-impaired adolescents. Psychological Reports, 80(3, Pt. 2), 1115-1122.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin, 101(2), 171.

Feuerstein, R., Rand, Y., & Hoffman, M. (1979). Dynamic assessment of retarded performers. Baltimore: University Park Press.

Gardner, H. (1983, 1993). Frames of mind: The theory of multiple intelligences. New York: Basic Books.

Gardner, H. (1993). Multiple intelligences: The theory in practice. New York: Basic Books.

Goleman, D. (1999). Daniel Goleman talks about emotional intelligence. Scholastic Early Childhood Today, 13(4), 29-30.

Goleman, D. (1995). Emotional intelligence: Why it can matter more than IQ. New York: Bantam Books.

Goleman, D. (1998). Working with emotional intelligence. New York: Bantam Books.

Gresham, F.M., & Witt, J.C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. School Psychology Review, 12, 249-267.

Guilford, J. P. (1967). The nature of human intelligence. New York: McGraw-Hill.

Gustafsson, J.E. (1984). A unifying model for the structure of intellectual abilities. Intelligence, 24(1), 179-203.

Horn, J. L. (1982). The theory of fluid and crystallized intelligence in relation to concepts in cognitive psychology and aging in adulthood. In F. I. M. Craik & S. Trehub (Eds.), Aging and cognitive processes. New York: Plenum.

Horn, J.L. (1991). Measurement of intellectual capabilities: Gf-Gc theory. In K.S. McGrew, J.K. Werder, & R.W. Woodcock, WJ-R technical manual (pp. 197-232). Itasca, IL: Riverside Publishing.

Horn, J.L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), Contemporary intellectual assessment: theories, tests, and issues (pp. 53-93). New York: Guilford.

Ittenbach, R. F., Esters, I. G., & Wainer, H. (1997). The history of test development. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), Contemporary intellectual assessment: theories, tests, and issues (pp. 17-31). New York: Guilford.

Ivey, A. E. (1994). Intentional interviewing and counseling. Belmont, CA: Brooks/Cole.

Jensen, A. R. (1980). Bias in mental testing. New York: Free Press.

Jensen, A.R. (1981). Straight talk about mental tests. New York: Free Press

Jensen, A. R. (1998). The g factor. Wesport, CT: Praeger.

Kaufman, A.S., & Kaufman, N.L. (1990). Kaufman Brief Intelligence Test. Circle Pines, MN: American Guidance Service.

Klein, R.M. (1999). The Hebb legacy. Canadian Journal of Experimental Psychology, 53(1), 1-3.

Lidz, C. S. (1991). Practitioner's guide to dynamic assessment. New York: Guilford Press.

MacMillan, D.L., & Reschly, D. J. (1998). Overrepresentation of minority students: The case for greater specificity or reconsideration of the variables examined. Journal of Special Education, 32(1), 15-24.

Markwardt, F.C., Jr. (1989). Peabody Individual Achievement Test – Revised. Circle Pines, MN: American Guidance Service.

McCallum, R., Bracken, B., & Wasserman, J. (2001). Essentials of nonverbal assessment. New York: Wiley & Sons.

McCallum, R.S., & Bracken, B.A. (1997). The Universal Nonverbal Intelligence Test. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 268-280), New York: Guilford.

McGrew, K.S. (1993). The relationship between the Woodcock-Johnson Psycho-Educational Battery – Revised Gf-Gc cognitive clusters and reading achievement across the life-span. Journal of Psychoeducational Assessment Monograph Series: Woodcock-Johnson Psycho-Educational Battery – Revised, 39-53. Cordova, TN: Psychoeducational Coporation.

McGrew, K.S., & Flanagan, D.P. (1998). The Intelligence Test Desk Reference: Gf-Gc Cross Battery Assessment. Needham Heights, MA: Allyn & Bacon.

Moore, C., O'Keefe, S., Lawhon, D. (1998). Concurrent validity of the Snijders-Oomen Nonverbal Intelligence Test-Revised with the Wechsler Preschool and Primary Scale of Intelligence-Revised. Psychological Reports, 82(2), 619-625.

Naglieri, J. A. (1985). Matrix Analogies Test, Expanded Form: Examiner's manual. San Antonio, TX: Psychological Corporation.

Naglieri, J. A., & Prewett, P. N. (1999). Nonverbal intelligence measures: A selected review of instruments and their use. In C. R. Reynolds & R. W Kamphaus (Eds.), Handbook of psychological and educational assessment of children: Intelligence and achievement (pp. 348-370). New York: Guilford.

Newell, A., & Bickerton, D. (1992). Precision of unified theories of cognition. Behavioral and Brain Sciences, 15(3), 425.

Nitko, A. J. (1998). Review of the TerraNova: Comprehensive Tests of Basic Skills, Fifth Edition. In J. J. Kramer and J. C. Conoley (Eds.), The twelfth mental measurements yearbook. Lincoln, NE: The Buros Institute of Mental Measurements.

Patton, J. M. (1998). The disproportionate representation of African-Americans in special education: Looking behind the curtain for understanding and solutions. The Journal of Special Education, 32 (1), 25-31.

Radcliff, J. K. & Radcliff, M. W. (1979). The Leiter scales: A review of validity findings. American Annals of the Deaf, 124, 38-45.

Raven, J. C., Court, J. H., & Raven, J. (1986). Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 3): Standard progressive matrices (1986 edition, with U.S. norms). London: Lewis.

Reed, M., & McCallum R. S. (1995). Construct validity of the universal nonverbal intelligence test (UNIT). Psychology in the Schools, 32(4), 277-290.

Roid, G. H., & Miller, L. J. (1997). Leiter international performance scale-revised. Wood Dale, IL: Stoelting.

Roid, G. H., & Woodcock, R. W. (2000). Uses of Rasch scaling in the measurement of cognitive development and growth. Journal of Outcome Measurement, 4(2), 579-94.

Rumelhart, D. E., Widrow, B., & Lehr, M. A. (1994). The basic ideas in neural networks. Association for Computing Machinery: Communications of the ACM, 37(3), 86-93.

Sampson, E. E. (1994). Sperry's cognitive revolution. The American Psychologist, 49(9), 818-819.

Sattler, J. M. (1992). Assessment of children (3rd ed., revised). San Diego, CA: Author.

Spearman, C. (1927). The abilities of man: Their nature and measurements. New York: Macmillan.

Sperry, R. G. (1993). The impact and promise of the cognitive revolution. The American Psychologist, 48(8), 878-885.

Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. Cambridge, UK: Cambridge University Press.

Sternberg, R. J., Wagner, R. K., Williams, W. W., & Horvath, J. A. (1995). Testing common sense. The American Psychologist, 50(11), 912-928.

Sternberg, R. J., (1997). The triarchic theory of intelligence. In Flanagan, D.P., Genshaft, J.L., & Harrison, P.L., Contemporary Intellectual Assessment (pp. 92-104). New York: Guilford.

Sullivan, P. M., & Burley, S. (1999). Mental testing of the hearing impaired child. In C. R. Reynolds & R. W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children: Intelligence and achievement (pp. 761-788). New York: Guilford.

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

Thurstone, L. L. (1960). The nature of intelligence. Littlefield: Adams.

U.S. Bureau of the Census. (2000). Language use [Online]. Available: http://www.census.gov/population/www/socdemo/lang_use.html

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

Wechsler, D. (1992). Wechsler Individual Achievement Test. San Antonio, TX: Psychological Corporation.

Williams, P.C., McCallum, R.S., & Reed, M.T. (1996). Predictive validity of the Cattell-Horn Gf-Gc constructs to achievement. Assessment, 3, 43-51.

Woodcock, R.W., Johnson, M.B. (1989/1990). Woodcock-Johnson Psychoeducational Battery – Revised. Itasca, IL: Riverside Publishing.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). Woodcock-Johnson Psychoeducational Battery – Third Edition. Itasca, IL: Riverside Publishing.

Young, E. L., & Assing, R. (2000). Review of the Universal Nonverbal Intelligence Test, Journal of Psychoeducational Assessment, 18(3), 280-2

APPENDICES

Table 1

Descriptive Statistics of UNIT, Leiter-R, and Terra Nova Cognitive and Academic Achievement Variables (N = 100[a])

	Low	High	*M*	*SD*
Leiter-R Fluid Reasoning	62.00	131.00	99.07	14.30
Leiter-R Full Scale IQ	50.00	132.00	97.65	14.57
UNIT Memory Quotient	66.00	140.00	101.54	13.69
UNIT Reasoning Quotient	50.00	141.00	103.91	16.31
UNIT Symbolic Quotient	61.00	140.00	101.59	15.19
UNIT Non-symbolic Quotient	62.00	137.00	103.67	14.79
UNIT Full Scale IQ	62.00	139.00	102.90	14.80
Terra Nova Total Reading[a]	605.00	782.00	685.70	40.87
Terra Nova Total Math[a]	614.00	764.00	693.95	43.00
Terra Nova Total Language[a]	615.00	748.00	683.41	35.24

[a] Terra Nova Scales were computed using $n = 37$.

Table 2

Intercorrelations Among UNIT, Leiter-R, and Terra Nova Cognitive and Academic Achievement Variables $(N = 100^a)$

	1	2	3	4	5	6	7	8	9
1. Leiter-R Fluid Reasoning									
2. Leiter-R Full Scale IQ	.74**								
3. UNIT Memory Quotient	.33**	.52**							
4. UNIT Reasoning Quotient	.52**	.74**	.59**						
5. UNIT Symbolic Quotient	.36**	.55**	.90**	.70**					
6. UNIT Non-symbolic Quotient	.49**	.72**	.64**	.90**	.57**				
7. UNIT Full Scale IQ	.48**	.72**	.88**	.90**	.89**	.88**			
8. Terra Nova Total Reading[a]	.54**	.61**	.58**	.72**	.64**	.73**	.74**		
9. Terra Nova Total Math[a]	.56**	.55**	.58**	.56**	.54**	.645**	.64**	.71**	
10. Terra Nova Total Language[a]	.55**	.56**	.51**	.59**	.49**	.65**	.62**	.81**	.82**

$^{**}p < .001$. [a] Note: For all correlations $N = 100$, except correlations with Terra Nova scales. For these $n = 37$.

Table 3

Multiple Stepwise Regression Analysis Predicting Total Reading From UNIT and Leiter-R Full Scale IQ Scores (N= 37)

Step and predictor variable	R^2	ΔR^2	ΔF	*pr*
Step 1	.55*	.55*	42.22*	
UNIT Full Scale IQ				.74*
Step 2	.55*	.01	0.50	
UNIT Full Scale IQ				.54*
Leiter-R Full Scale IQ				.12

* $p < .001$.

Table 4

Multiple Stepwise Regression Analysis Predicting Total Language From UNIT and Leiter-R Full Scale IQ Scores (N= 37)

Step and predictor variable	R^2	ΔR^2	ΔF	*pr*
Step 1	.39**	.39**	22.05**	
UNIT Full Scale IQ				.62**
Step 2	.41**	.02	1.11	
UNIT Full Scale IQ				.37*
Leiter-R Full Scale IQ				.18

* $p < .05$. ** $p < .001$.

Table 5

Multiple Stepwise Regression Analysis Predicting Total Math From UNIT and Leiter-R Full Scale IQ Scores (N= 37)

Step and predictor variable	R^2	ΔR^2	ΔF	*pr*
Step 1	.413**	.413**	24.63**	
UNIT Full Scale IQ				.64**
Step 2	.425**	.012	0.71	
UNIT Full Scale IQ				.41*
Leiter-R Full Scale IQ				.14

* $p < .01$. ** $p < .001$.

Table 6

Multiple Stepwise Regression Analysis Predicting Total Reading From UNIT and Leiter-R Cognitive Subcontstructs (N= 37)

Step and predictor variable	R^2	ΔR^2	ΔF	*pr*
Step 1	.521**	.521**	38.02**	
UNIT Reasoning Quotient				.72**
Step 2	.567**	.046*	3.65*	
UNIT Reasoning Quotient				.59**
UNIT Memory Quotient				.31

* $p < .10$. ** $p < .001$.

Excluded ($p = .10$ entry criteria): Leiter-R Fluid Reasoning.

Table 7

Multiple Stepwise Regression Analysis Predicting Total Reading From UNIT and Leiter-R Cognitive Subcontstructs (N= 37)

Step and predictor variable	R^2	ΔR^2	ΔF	*pr*
Step 1	.527*	.527*	39.04*	
UNIT Nonsymbolic Quotient				.73*

* $p < .001$.

Excluded ($p = .10$ entry criteria): Leiter-R Fluid Reasoning, UNIT Symbolic Quotient.

Table 8

Multiple Stepwise Regression Analysis Predicting Total Math From UNIT and Leiter-R Cognitive Subcontstructs (N= 37)

Step and predictor variable	R^2	ΔR^2	ΔF	*pr*
Step 1	.336**	.336**	17.68**	
UNIT Memory Quotient				.58**
Step 2	.488**	.152*	10.11*	
UNIT Memory Quotient				.50*
Leiter-R Fluid Reasoning				.48*

* $p < .005$ ** $p < .001$.

Excluded ($p = .10$ entry criteria): UNIT Reasoning Quotient.

Table 9

Multiple Stepwise Regression Analysis Predicting Total Math From UNIT and Leiter-R Cognitive Subcontstructs (N= 37)

Step and predictor variable	R^2	ΔR^2	ΔF	*pr*
Step 1	.416***	.416***	24.93***	
UNIT Nonsymbolic Quotient				.65***
Step 2	.468***	.052*	3.30*	
UNIT Nonsymbolic Quotient				.47**
Leiter-R Fluid Reasoning				.30*

* $p < .10$. ** $p < .005$. *** $p < .001$.

Excluded ($p = .10$ entry criteria): UNIT Symbolic Quotient.

Table 10

Multiple Stepwise Regression Analysis Predicting Total Language From UNIT and Leiter-R Cognitive Subcontstructs (N= 37)

Step and predictor variable	R^2	ΔR^2	ΔF	*pr*
Step 1	.342*****	.342*****	18.19*****	
UNIT Reasoning Quotient				.59*****
Step 2	.396*****	.054*	3.06*	
UNIT Reasoning Quotient				.37**
Leiter-R Fluid Reasoning				.29*
Step 3	.448*****	.052*	3.11*	
UNIT Reasoning Quotient				.22
Leiter-R Fluid Reasoning				.31*
UNIT Memory Quotient				.29*
Step 4	.421*****	-.027	1.62[a]	
Leiter-R Fluid Reasoning				.47****
UNIT Memory Quotient				.42***

* $p < .10$. ** $p < .05$. *** $p < .01$. **** $p < .005$. ***** $p < .001$.
Excluded ($p = .10$ entry criteria): UNIT Reasoning Quotient.

[a] $p > .20$ Removal Criteria

Table 11

Multiple Stepwise Regression Analysis Predicting Total Language From UNIT and Leiter-R Cognitive Subcontstructs (N= 37)

Step and predictor variable	R^2	ΔR^2	ΔF	*pr*
Step 1	.426*	.426*	26.03*	
UNIT Nonsymbolic Quotient				.65*

* $p < .001$.

Excluded ($p = .10$ entry criteria): Leiter-R Fluid Reasoning, UNIT Symbolic Quotient.

VITA

Vincent Scott Hooper was born in Whitwell, Tennessee. He attended public schools in the Marion County, Tennessee, school system. He was graduated from Whitwell High School, Whitwell, Tennessee, in May of 1987. He was graduated from the University of Tennessee at Chattanooga in August of 1991 with a B.S. in Economics. After working in the juvenile court system for a year, the author began graduate school at the University of Tennessee at Chattanooga in the School Psychology program under the direction of Dr. George Helton. He received a Master of Science degree in Psychology with a concentration in School Psychology in August 1994. The author was accepted into the doctoral program in School Psychology at the University of Tennessee at Knoxville and enrolled in August 1994. Graduate assistantships were held in the Psychoeducational Studies Unit under Dr. Steve McCallum and in the Athletic Department as a tutor and mentor to students identified with learning disabilities. The predoctoral internship was completed through the Louisiana School Psychology Internship Consortium in New Orleans, Louisiana. The LAS*PIC internship program is accredited by the American Psychological Association. Since August of 2000, the author has been employed by the Hamilton County Department of Education in Chattanooga, Tennessee. He has worked as a school psychologist with the HCDE Preschool Developmental Assessment Team, and in August of 2002, he was promoted to Lead School Psychologist and now provides technical support, staff development, and mentoring to 45 school psychologists in a diverse system of urban, suburban, and rural schools. The Doctor of Philosophy degree was conferred in December of 2002.