8-2008

Concurrent and Predictive Validity of the Universal Nonverbal Intelligence Test-Group Ability Test

Stacy L. Bliss

University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Stacy L. Bliss entitled "Concurrent and Predictive Validity of the Universal Nonverbal Intelligence Test-Group Ability Test." 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:

Robert Williams, Sherry Bain, Robert Saudargas

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 Stacy L. Bliss entitled "Concurrent and Predictive Validity of the Universal Nonverbal Intelligence Test-Group Ability Test." 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:

Robert Williams_____

Sherry Bain_____

Robert Saudargas_____

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the
Graduate School

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

Concurrent and Predictive Validity of the Universal Nonverbal Intelligence Test-Group Ability Test

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

Stacy L. Bliss
August 2008

Dedication

For everyone who supported me

"Are you closer right now to where you want to be than you were a half-hour ago?"
---Robert Cooper

Acknowledgement

I would like to extend my appreciation to four people who were responsible for guiding me during the past four years. Drs. Bain, McCallum, Skinner, and Williams, you taught me how to be a good researcher, instructor, professional, and person, thank you. I would also like to thank several people who were integral in the collection of this data, Drs. Draper, Gettlefinger, and Morse and the principals and teachers at Lenoir City Schools. Thank you for allowing me to disrupt your classrooms. Additionally, to the staff at Riverside Publishing, Tara, Tommie, Jon, and Melanie, thank you for all your assistance. Finally, Emily, you are and always will be, my favorite inter-rater.

Abstract

In order to determine the concurrent and predictive validity of the Universal Nonverbal Intelligence Test- Group Ability Test (UNIT-GAT; McCallum & Bracken, in press), the UNIT-GAT and the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997a) were administered in counter-balanced order to 93 students. In addition, 40 students were rated on the Universal Nonverbal Intelligence – Gifted Screening Scales (UNIT-GSS; McCallum & Bracken, in press). The correlation coefficient of r = .36 between the UNIT-GAT total raw score and the NNAT was statistically significant at the $p < .01$ level. The UNIT-GAT scale score correlations with the NNAT total ranged from r = .18 for the Symbolic Scale to r= .53 ($p< .01$) for the Nonsymbolic Scale. The UNIT-GAT total raw score correlations with the UNIT-GSS composite and scales ranged from r = -.06 between both the Emotional and Science scales to r = .19 on the Creative Scale. None of the correlations were statistically significant. The correlations between the scales of the UNIT-GAT and composites of the UNIT-GSS ranged between r= -.05 (UNIT-GAT Memory Scale and UNIT-GSS General Aptitudes Composite) to r = .20 (UNIT-GAT Reasoning Scale and UNIT-GSS General Aptitudes Composite). Correlations between the scales of the UNIT-GAT and the scales of the UNIT-GSS ranged from r = -.30 between the UNIT-GAT Memory Scale and UNIT-GSS Emotional Scale to r = .25 between the UNIT-GAT Nonsymbolic Scale and UNIT-GSS Creative Scale.

Stepwise multiple regression analysis did not reveal any significant utility by the UNIT-GAT total raw score or the NNAT total raw score to predict teacher-ratings on the UNIT-GSS General Aptitude and Specific Academic Aptitude Composites. Implications and future directions for research are discussed.

TABLE OF CONTENTS

List of Tables

Table Page

Chapter 1

Introduction

Although there are a number of group-administered nonverbal intelligence tests currently in use, none are completely without some form of verbal mediation. Additionally, none of these tests provide measures of both reasoning and memory (i.e., all assess reasoning only). Consequently, the Universal Nonverbal Intelligence Test – Group Ability Test (UNIT-GAT; McCallum & Bracken, in press) was developed to address these limitations. The UNIT-GAT is a nonverbal group-administered intelligence test that measures both memory and reasoning. The test is currently in the experimental phase and validity studies are necessary to determine the technical properties of the test. The purpose of this study is to determine the concurrent validity of the UNIT-GAT.

History of Intelligence Testing

Psychological testing can trace its roots to the use of civil service examinations in 2200 B.C. China (DuBois, 1970), where Chinese officials were examined every third year to determine their fitness for remaining in office. In the early 19th century, British diplomats to China brought the model of competitive examinations back to England, and the practice quickly spread to the United States. These early efforts to assess "intelligence" were highly primitive and relied on the measurement of characteristics of human faculty, blocks of varying weights and visual images (DuBois) into the early 20th century. In the first large-scale collection of data, Galton opened his Anthropometric Laboratory in 1884, where he took the physical measurements of close to 10,000 people. His techniques were introduced to the U.S. by James Cattell, a student of Galton who founded the Psychological Laboratory at Columbia University (DuBois). He, like Galton,

developed a battery of tests that was intended to focus on the measurement of the body and senses (Cattell & Farrand, 1896).

Modern intelligence testing is generally considered to have begun with Alfred Binet (Bartholomew, 2004; Wasserman & Tulsky, 2005). Binet, citing "limitations" of the assessments of Galton and Cattell (Wasserman & Tulsky), developed his own scale of intelligence and presented it at the 1905 International Congress of Psychology. This scale, intended to provide examinations to students who were not benefiting from general education classes, was revolutionary in that individual items were ranked in order of difficulty and administration instructions were included within the test materials. The scale was translated into English and brought to the United States in 1908 by Henry Goddard. Termed the Binet - Simon Scale, it quickly became the standard for intelligence testing.

Two revisions were made to the Binet - Simon Scale in the year before the United States entered World War 1. The first structured the Binet - Simon Scale into a point-scale rather than a year-scale (Yerkes, Bridges, & Hardwick, 1915). The second, by Terman, extended the age range into adulthood and replaced the standard mental age with an overall intelligence quotient score. He also gave the Binet - Simon Scale several new subtests, including the form board originally developed by Seguin (Boake, 2002). This revision was termed the Stanford-Binet Intelligence Scale and became the principal test in the United States for measuring intelligence.

When the United States' entered World War I, the military was faced the daunting task of determining appropriate placement for thousands of foreign recruits. The result was the development of the Army Alpha and Army Beta exams. Developed by

psychologists and administered by army examiners, the Army Alpha, a verbal test, and Army Beta, a nonverbal test, represent the first successful attempt at group-administered intelligence testing.

One of the Army examiners, David Weschler, began his own investigations into intelligence. In 1939, he published the Weschler-Bellevue Intelligence Scale. The test yielded a verbal, performance, and total IQ score, eliminating the need for separate performance-based tests. The original test has been replaced by scales specific to adults, children, and preschool-aged examinees and has become the most widely used measures of intelligence (Naglieri, 2000). Revisions to these scales and other intelligence instruments continue today.

Currently, millions of intelligence tests are given each year for a variety of purposes, including psychological treatment, classification of students for special education services (Anastasi & Urbina, 1997), licensure, and placement determinations. However, many of the same difficulties (e.g., bias and validity issues for special populations) that early developers faced have not been addressed sufficiently. Compulsory school attendance and the increasing diversity of public schools have led to the need to be able to assess accurately the abilities of students who have hearing or linguistic challenges, cultural differences, and lower socioeconomic status.

Intelligence Testing for Students with Communication or Language Disorders

Soon after development of the Binet - Simon scale, researchers began to apply testing procedures to children who have difficulties communicating verbally. Pintner and Patterson (1915), the first to administer intelligence tests to hearing-impaired children, found that this group was consistently scoring in the range of mentally retarded.

Recognizing the heavy verbal bias in the intelligence tests of the time (Pintner & Patterson, 1921), they developed the Pintner Non-language Test (Pintner, 1924), which reduced the amount of verbal language required to complete tasks. While the results of this test indicated that hearing-impaired children were much closer to the normal population than the verbal tests concluded, mean scores were still significantly lower than those obtained by hearing children (in Vernon, 1968). Still other studies by Reamer (1921) and Day, Fusfield, and Pintner (1928) continued to report that hearing-impaired children scored well below average, perpetuating the common belief that children who were deaf or hearing-impaired were less intelligent than those without these limitations (Vernon). Drever and Collins (1928), who further reduced the language component and found that hearing children and hearing-impaired children scored similarly on measures of mental ability, challenged this belief. Since 1930, numerous studies show that hearing-impaired children score at similar levels of hearing children (see Braden, 1992). Nonverbal assessments have now become the standard for measuring the abilities of people with hearing or linguistic limitations.

Nonverbal Intelligence Tests

Today there are two primary methods of assessing the intelligence for those persons whose linguistic or cultural difference may introduce bias into evaluations. First, the traditional intelligence tests designed for populations fluent in English are adapted for use with populations who are not fluent in English. Common examples include the Weschler Intelligence Scales for Children, fourth edition (Weschler, 1991), and the Stanford Binet, fifth edition (Roid, 2003), both of which contain some form of nonverbal assessment subtests. Critics of the use of these tests claim that administration and

completion require language-based skills; consequently, they do not provide a true measure of nonverbal intelligence, but are only somewhat language-reduced (Hooper, 2004).

The second method is to use specialized intelligence tests designed to assess intelligence with items and tasks that do not require verbalizations to either (a) administer the test, (b) complete the items, or (c) administer or complete the test. Nonverbal intellectual measures typically require reasoning, spatial, and 2-dimensional visualization, memory, attention, concentration for complex tasks, and speed of processing complex information. These abilities do not require proficiency in perceiving and reasoning with words or numbers, or any other material traditionally defined as verbally laden (Roid & Miller, 1997). There are two types of nonverbal tests. The first uses only one method of assessment, such as progressive matrices or matrix analogies (e.g. Comprehensive Test of Nonverbal Intelligence, Test of Nonverbal Intelligence, third edition, and Naglieri Nonverbal Abilities Test), while the second uses a variety of methods to assess multiple facets of intelligence, including memory and reasoning tasks. Currently only two individually-administered, well-accepted, multi-faceted nonverbal tests are available, the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998) and the Leiter International Performance Scale, Revised (Roid & Miller, 1997). Group-administered screening tests have traditionally used an exclusively unidimensional format while individually-administered tests have typically required a multi-dimensional one. Recently the UNIT-GAT was introduced and is the only multi-faceted nonverbal group-administered test available.

Group-administered Nonverbal Intelligence Tests

The first group-administered intelligence tests were used to examine men to determine their suitability for the military during World War 1. Army Alpha was a verbal test intended for examinees fluent in the English language. Army Beta was a nonverbal group test given to those men who were not fluent in English, illiterate, and those who performed poorly on the Army Alpha (Yoakum & Yerkes, 1920). After the war, several new group intelligence tests appeared; many were verbal and similar to the Army tests but with written directions and content, still a problem for recent immigrants. While these tests allowed for the efficient testing of large numbers of subjects, they did not address the necessity of testing examinees who were illiterate, not fluent in English, or who had hearing-impairments. The need for a completely nonverbal group-administered test was apparent.

Naglieri Nonverbal Ability Test. The Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997a) is a group-administered progressive matrix test. The test is a revision of the Matrix Analogies Test Short Form (Naglieri, 1985) and is designed to provide a measure of general ability that uses nonverbal items in a group administration format. The NNAT is comprised of similar questions as the NNAT - Individual. Administration is completed by levels, each containing 38 items, and can be conducted by the classroom teacher. Starting level is determined by grade-level. Raw scores are converted to scaled scores (mean = 10, SD = 3), which are then converted to standard scores (mean = 100, SD = 15). The advantages of the NNAT include those expected for group tests, a brief administration time and machine scoring.

Limitations of the NNAT include a lack of information on the examinee's ability to verbally mediate tasks and validity evidence that is of poor quality and lacking in integration (Trevisan, 1999). Additionally, the instructions to the NNAT are administered verbally, making the test inappropriate for use with examinees that do not have an understanding of the English language. Stinnett (1999) also cautions that, as the NNAT only requires a "B" user classification, users of the NNAT may not be properly trained in the collection and use of intelligence data.

Otis-Lennon School Ability Test. The Otis-Lennon School Ability Test (Otis-Lennon; Otis & Lennon, 1996) was originally developed in 1918 as the Otis Group Intelligence Scale (DeStefano, 1999). The Otis-Lennon is a group-administered test comprised of both Verbal and Nonverbal components. The Otis-Lennon is arranged into seven levels, used for students in kindergarten through the 12th grade. Twenty-one different item types are organized into five clusters; Pictorial Reasoning, Figural Reasoning, Quantitative Reasoning, Verbal Comprehension, and Verbal Reasoning. The first three clusters combine to form the Nonverbal component and the latter two comprise the Verbal component. Within each cluster are several subtests, which are administered according to grade-level. Not all subtests are administered to each level. The Otis-Lennon was standardized using a stratified random sampling technique to obtain a sample proportionate to the U.S. public and private school enrollment. Component scores (both Verbal and Nonverbal) and the total score are represented as School Ability Indexes (mean = 100, standard deviation = 16).

Reliability information is reported in the technical manual (Otis & Lennon, 1996). Internal consistency reliability coefficients are generally in the .80s and .90s, although

several fall in the .70s and the Level A Verbal component for age 5 years 0 months to 5 years 2 months falls as low as .68. The primary evidence of validity is presented as correlational data between the sixth and seventh editions. Total score correlations range from .77 to .87 with Verbal and Nonverbal component score correlations falling slightly lower. Correlations between the Otis-Lennon and the Stanford Achievement Test Ninth Edition (Harcourt Brace Educational Measurement, 1996) are presented as a demonstration of the relationship between the test and academic achievement.

The Otis-Lennon's primary strengths lie in the rigorous development methods and the ease in which it can be administered to large groups of students (DeStefano, 1999). There are several weaknesses, the most serious of which is the insufficient validity evidence. Additionally, recommended uses are not addressed in the Examiner's Manual. Despite these limitations, the Otis-Lennon could be one of a variety of instruments used for screening purposes (DeStefano, 1999).

InView. The InView (CTB-McGraw Hill, 2000) is an updated version of the Test of Cognitive Skills (CTB, 1992). It is a group-administered test of cognitive ability. The test can be administered either directly in person or by computer. The InView was co-normed with the TerraNova, Second Edition, a group-administered measure of achievement. The InView consists of five subtests, three measuring nonverbal ability and two measuring verbal ability. The InView yields five subtest scores and three aggregate scores (Verbal, Nonverbal, and a Total score). A Cognitive Skills Index (mean = 100, standard deviation = 16) serves as a measure of general ability. Additional available scores include scale scores, grade equivalents, and percentile ranks.

Internal consistency reliability was measured using the Kuder-Richardson Formula 20 (KR-20). KR-20 values were generally in the .80s for the subtests, verbal, and nonverbal composites. Total score KR-20 values ranged from .95 to .96. Concurrent validity for subtests of the InView with the TerraNova are in the range of .40 to .70. According to Carney (2001) and Thompson (2001), confirmatory factor analysis indicates a good fit with the model of a single, general trait and verbal and nonverbal traits. In general, the InView provides an adequate group-administered measure of cognitive ability.

Cognitive Abilities Test, Form 6. The Cognitive Abilities Test Form 6 (CAT; Lohman & Haggen, 2001) is a group-administered test of general reasoning skills. The purpose of the test is to evaluate the level and pattern of cognitive development of students from kindergarten through grade 12. There are three author-identified uses for the CAT. The first is to guide instruction to match the cognitive abilities of each student in the classroom, the second is to provide an alternative measure of cognitive development relative to standardized achievement tests, and the third is to identify achievement-ability discrepancies. The test is administered by classroom teachers reading the instructions to students and then students are expected to complete each item on their own. The CAT is comprised of two editions, the Primary Edition which contains three levels, for students in kindergarten through second grade, and the Multilevel Edition which contains eight levels, for students in third grade through 12th grade. Both Editions include three test batteries (Verbal, Quantitative, and Nonverbal), with the Multilevel Edition containing three subtests in each battery and the Primary Edition has only two levels. Multiple scores are available for each of the batteries, including Standard Age

Scores (mean = 100, standard deviation = 16), and percentile ranks. These scores can be calculated using an age- or grade-based comparison group. The CAT also yields a Universal Scale Score which allows for comparison of performance across levels of the test.

Beta III. The Beta III (Kellogg & Morton, 1999) is a group-administered test for people between the ages of 16 and 89 years. It consists of five subtests with an administration time of approximately 30 minutes. The Beta III is intended to measure visual information processing, processing speed, spatial and nonverbal reasoning, and certain aspects of fluid intelligence (McCallum, Bracken, & Wasserman, 2001). The test is intended to be used with individuals for whom verbal assessment would be inappropriate, such as those who are non-English speakers, illiterate, or language-disordered.

The Beta III revision extended the age range, updated norms, improved the test content, and raised the ceiling of possible IQ scores up to 155 points. Available scores for the Beta III include scaled score (mean = 10, standard deviation = 3), an overall Beta III IQ (mean = 100, standard deviation = 15), and percentile ranks. While factor analysis indicates two tests (Coding and Clerical Checking) measure processing speed and the remaining three tests (Matrix Reasoning, Picture Completion, and Picture Absurdities) measure nonverbal reasoning, the test is best interpreted at the composite level (McCallum, Bracken, & Wasserman, 2001). The primary strengths of the Beta III include its outstanding standardization sample, high correlations with several indices of intelligence, and its ease of administration. Additionally, the authors include multiple practice problems. Limitations of the test include a lack of acceptable reliability

(McCallum, Bracken, & Wasserman, 2001), and construct validity (Bellah, 2001), speed requirements of subtests which make it inappropriate for examinees with motor impairments, and directions that are verbally administered, making it inappropriate for non-English speaking examinees.

Raven's Progressive Matrices. Raven's Progressive Matrices (Raven's Matrices; Raven, Raven, & Court, 1998) is an un-timed, individually- or group-administered collection of matrix reasoning tests. Originating in 1938, Raven's Matrices includes six major versions; Coloured Progressive Matrices, Coloured Progressive Matrices Parallel, Standard Progressive Matrices, Standard Progressive Matrices Parallel, Standard Progressive Matrices Plus, Matrices Plus, and Advanced Progressive Matrices. Raven's Matrices purports to measure the eductive component of g, or the ability to obtain meaning in confusion, forge new insights, and identify relationships (Raven, 2000), as defined by Spearman's theory of ability (McCallum, Bracken, & Wasserman, 2001). Recently, factor analytic and experimental evidence provide for the argument that Ravens Matrices' items measure two processes, perceptual and analogical (Van der Ven & Ellis, 2000). Directions are verbal, and take approximately five to 10 minutes to recite. Group testing is not recommended for children under the age of six. For examinees that are not proficient in English or have hearing impairments, Raven's Matrices can be administered without spoken directions through the use of pantomimed gestures. Raven's Matrices yield overall descriptive categories and percentile ranks.

The various versions of Raven's Matrices have been normed in Argentina (Angelini, Alves, Cutodino, & Duarte, 1989), Australia (Cotton, Kiely, Crewther, Thomson, Laycock, & Crewther, 2005), Canada (Yeudall, Fromm, Reddon, & Stefanyk,

1986), Egypt (Abdel & Ahmed, 1998), France (Bourdier, 1964), Hong Kong (Chan, 1989), India (Bhogle & Prakash, 1992), Kenya (Costenbader & Ngari, 2001), South Africa (Owen, 1992), the United Kingdom (Raven, Raven, & Court, 1990; Raven, Raven, & Court, 1998) and several other countries. Reliability studies indicate that generally the various versions of Raven's Matrices show good reliability of scores across cultures (Valencia, 1984; Mills & Tissot, 1995) and genders (Benbow & Minor, 1990). Test-retest reliability, with a delay of two weeks, on the Coloured Progressive Matrices resulted in alphas ranging from .69 to .85 (Abdel & Ahmed, 2005). Although, test-retest reliability studies with longer delays between test administration have found alphas ranging from .49 (Kazlauskaite & Lynn, 2002) to .74 (Vodegel-Matzen, van der Molen, & Dudink, 1994), reliability estimates across ages indicate that Raven's Matrices has lower reliability at lower ages (Barnabas, Kapur, & Rao, 1995).

The most prominent weakness of Raven's Matrices is the norming sample. Raven's Matrices are the only major nonverbal instrument that does not have adequate U.S. standardization norms (McCallum, Bracken, & Wasserman, 2001). The original test, the Standard Progressive Matrices, yielded inadequate discrimination among the upper and lower levels (Raven, 2000), a problem that may have not been adequately resolved (Gudjonsson, 1995). Additionally, the option of six versions can make choosing the appropriate test difficult, especially when the administrators need to select the appropriate norm reference for each version. In spite of these weaknesses, Raven's Matrices is the most extensively-researched nonverbal measure available (McCallum, Bracken, & Wasserman). Additionally, the test is easy to administer and shows good convergent validity with other intelligence tests.

Universal Nonverbal Intelligence Test-Group Ability Test. The Universal Nonverbal Intelligence Test-Group Ability Test (UNIT-GAT; McCallum & Bracken, in press) is a group-administered, multi-faceted, nonverbal intelligence test. The test contains two primary scales, Memory and Reasoning, and two secondary scales, Symbolic and Nonsymbolic. The purpose of the UNIT-GAT is to screen groups of students for the identification of giftedness or developmental delay. The UNIT-GAT is administered in an almost completely language-free manner, with the administrator using universal signs (e.g., shrugs and thumbs up) and three words (i.e., look, think, and stop) that are not related to test content. The test includes demonstration and sample items in each subtest to ensure that the examinee fully understands the task before moving to items scored for credit.

The Reasoning scale requires the examinee to recognize relationships between items depicted in boxes as they move across the rows. One box contains a question mark and the examinee must choose the correct picture from a list at the bottom to complete the analogy. Half of the pictures contain objects and symbols seen in everyday life, representing the Symbolic scale, while the remaining pictures are formed of geometric patterns, representing the Nonsymbolic scale. The Memory scale asks the examinee to study a series of pictures. The examinee must then pick the missing picture from a set of four possible responses. The items become progressively harder throughout the test. As with the Reasoning scale, half of the items are included in the Symbolic scale and formed of pictures of objects and symbols, while the other half are included in the Nonsymbolic scale and comprised of geometric patterns. Total Symbolic and Nonsymbolic scores are comprised of questions from both the Memory and Reasoning scales.

Identification of Students for Gifted Education.

Giftedness, as defined by the U.S. Department of Education (1993), consists of extraordinary intelligent and academically or artistically gifted students or high performance abilities in creative or leadership endeavors. Included in the definition are recommendations as to the best method of identifying students for gifted services. These recommendations specify the use of a multi-modal assessment, long promoted as the gold standard for all assessments (Kaufman & Harrison, 1986; Pfeiffer, 2001), and use a variety of sources such as traditional assessment, interviews, observations, work samples, and teacher reports or rating scales to develop a comprehensive picture of the student's abilities. This type of assessment is commonly used for the identification of students who have Attention Deficit Disorder, learning disabilities, and other disorders affecting school functioning. However, those abilities that are nonintellectual in nature (e.g., creativity and leadership) are not typically assessed (Alvino, McDonnel, & Richert, 1981). Faced with a limited knowledge about and number of nonintellectual assessment measures, schools have a considerable challenge in identifying students who meet these criteria. To alleviate this problem, practitioners have turned to rating scales to obtain information about potentially gifted students (Ashman & Vukelich, 1983; Haroutounian, 1995)). Teacher-completed rating scales have the benefit of assessing those areas ignored by intellectual scales (i.e., creativity, leadership, etc.). In general, teacher-rating scales have been found to be highly accurate when rating specific behaviors associated with giftedness (Borland, 1978). There are a number of gifted rating scales currently on the market, however, these assessments are global in nature and have been found to be lacking in areas such as;

standardization, normative sampling, reliability, and content validity (Jarosewich, Pfeiffer, & Morris, 2002).

An additional method of identifying students for comprehensive gifted assessment is through the use of group-administered intelligence tests. Group-administered intelligence tests can provide a time and resource-efficient method of determining which students warrant further assessment. Within these tests lie many of the same drawbacks as individually-administered tests, that of potential cultural and linguistic bias. Hence, the use of nonverbal, group-administered intelligence tests is becoming common. However, a note of caution, currently there is no group-administered nonverbal intelligence test on the market that shows good validity and predictive ability. A more accurate referral and identification method is needed, one that includes those abilities that are intellectual and nonintellectual and demonstrates high reliability and predictability to achievement.

Statement of Purpose

The previously discussed nonverbal intellectual assessments all have similar limitations. Each of the tests includes administration instructions that are primarily given verbally and none measure both reasoning and memory. The UNIT-GAT was developed to address these limitations. Currently the UNIT-GAT is in the experimental stage and validity of the instrument needs to be determined. The primary purpose of this study is to examine the concurrent validity of the Universal Nonverbal Intelligence Test-Group Ability Test (UNIT-GAT; Bracken & McCallum, in press) by comparing it to an existing standard, the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997a). A secondary purpose is to determine the extent to which both measures predict teacher-reported cognitive, general academic, language arts, math, and reading aptitude as measured by

the Universal Nonverbal Intelligence Test- Gifted Screening Scale (UNIT-GSS; McCallum & Bracken, in press). Within the context of this study, concurrent validity is defined as the comparison of a student's scores on two instruments that are similar in construct and purpose and administered within a relatively short time period to each other. The UNIT-GAT and NNAT, although administered several days apart for some students, both purport to measure nonverbal intelligence, thus meeting the concurrent validity definition. Concurrent validity is generally measured through correlation coefficient analysis. Predictive validity refers to an instruments ability to predict scores on a different instrument and is generally measured through the use of multiple regression analysis. In this study, the predictive ability of both the UNIT-GAT and NNAT will be examined.

Research Questions.

1. Are there significant relationships (i.e., correlation coefficients) between the raw scores of the UNIT-GAT scales, specifically Memory, Reasoning, Symbolic, and Nonsymbolic, and the NNAT Nonverbal Ability Index?
2. Is there a significant relationship (i.e., correlation coefficient) between the total raw scores of the UNIT-GAT and the raw score of the NNAT?
3. To what extent do the raw scores of the UNIT-GAT scales Memory, Reasoning, Symbolic, and Nonsymbolic correlate with the General Academic Aptitude cluster (comprised of the Cognitive Aptitude, Creative Aptitude, Emotional Aptitude, and Leadership Aptitude scales) of the UNIT-GSS?
4. To what extent does the total raw score of the UNIT-GAT correlate with the General Academic Aptitude cluster (comprised of the Cognitive Aptitude,

Creative Aptitude, Emotional Aptitude, and Leadership Aptitude scales), of the UNIT-GSS?

5. To what extent do the raw scores of the UNIT-GAT scales (Memory, Reasoning, Symbolic, and Nonsymbolic) correlate with the Specific Academic Aptitude cluster (comprised of the Language Arts Aptitude, Math Aptitude, Reading Aptitude, and Science Aptitude scales) of the UNIT-GSS?
6. To what extent does the total raw score of the UNIT-GAT correlate with the Specific Academic Aptitude cluster (comprised of the Language Arts Aptitude, Math Aptitude, Reading Aptitude, and Science Aptitude scales) of the UNIT-GSS?
7. What is the relative predictive efficiency of the UNIT-GAT total raw score and the raw score of the NNAT when the General Aptitudes Composite of the UNIT-GSS is the criterion?
8. What is the relative predictive efficiency of the UNIT-GAT total raw score and the raw score of the NNAT when the Specific Academic Aptitude Composite of the UNIT-GSS is the criterion?

Anticipated Results.

The UNIT-GAT is a multi-faceted test in that it measures both memory and reasoning, using nonsymbolic and symbolic items. The NNAT, on the other hand, is a unidimensional test, using only nonsymbolic items to measure the examinee's reasoning skills. Therefore, it is expected that the UNIT-GAT and NNAT will show moderate to strong correlations between the scales of the UNIT-GAT, Reasoning and Nonsymbolic, that measure similar constructs as the NNAT. The two remaining scales of the UNIT-

GAT, Memory and Symbolic, should show small correlations. Given an expected high correlation between all four scales of the UNIT-GAT to the UNIT-GAT total score, the total score of the UNIT-GAT should show a moderate correlation to the NNAT. The UNIT-GAT was developed with similar theoretical grounding as the original UNIT (Bracken & McCallum, 1998) which was shown to have moderate to strong correlations with measures of academic achievement (Williams, 1995; Hooper, 2003). Therefore, it is also expected that the UNIT-GAT will show a moderate correlation with and predictive ability to the composites and scales of the UNIT-GSS, a measure which has been shown to correlate moderately with achievement (Gray, 2006). Currently there is a shortage of predictive validity studies using the NNAT (Maller & Mowery, 2000) but the unidimensional nature of the instrument leads to an expectation of a low correlation with the UNIT-GSS.

Chapter 2

Methods

Participants

Data for this study were obtained from an existing data set provided by Riverside Publishing Company established to gather data for the purpose of developing normative standards for the UNIT-GAT and UNIT-GSS. The current data set contains thirty-two examinees in the second grade, thirty examinees in the fifth grade, and thirty-one examinees in the ninth grade. All students were administered the UNIT-GAT and NNAT. The regular classroom teachers (two second-grade, two fifth-grade, and one ninth-grade) of the students randomly chose 15 students in each grade and completed a UNIT-GSS. Participants were from an elementary and high school in a low socioeconomic school district. The elementary school has a population of 600 students, with 80% of the students receiving free or reduced lunch. The ethnic population of the elementary school at the time of data collection was 77.5% Caucasian, 1.2 % African American, and 21.4 % Hispanic. The ethnic diversity of the sample tested was 76% Caucasian and 24% Hispanic or African American. The high school has a population of 1100 students, with 33.2% receiving free or reduced lunch. The ethnic diversity of the high school at the time of data collection was 91.5% Caucasian, 1.0% African American, and 6.9% Hispanic. Ethnicity of the sample tested was 81% Caucasian and 19% Hispanic, Asian, or African American. Informed consent was obtained from the school administration by Riverside Publishing Company. No identifying information was contained in the data set analyzed.

Instruments

The instruments used in this study were the UNIT-GAT, NNAT, and UNIT-GSS. Concurrent validity was examined by administration of the UNIT-GAT and NNAT in counter-balanced order. As previously discussed, the NNAT is a group-administered nonverbal intelligence test that primarily measures reasoning ability and is currently the most commonly used group-administered nonverbal test. The NNAT has high reliability and validity studies found small differences between various populations (Naglieri & Ronning, 2000) as described below.

Universal Nonverbal Intelligence Test-Group Ability Test. The Universal Nonverbal Intelligence Test-Group Ability Test (UNIT-GAT; McCallum & Bracken, in press) is a group-administered, nonverbal test designed to measure the abilities of children and adolescents in a language-free fashion. The UNIT-GAT is intended to be used as a screening instrument to efficiently identify those students who are in need of more comprehensive evaluation for possible giftedness or developmental delay. While the UNIT-GAT does include the use of several words during administration (i.e., look, study, stop), these words do not convey information about the nature of the test nor how the test questions are to be answered. The UNIT-GAT is appropriate to use with children who are culturally different or have sensory limitations (i.e., deafness), learning disabilities, and various language-impairing neurological disorders. The test is divided into two primary scales, Memory and Reasoning, and two secondary scales, Symbolic and Nonsymbolic.

The Memory scale items require the examinee to study a series of related paired pictures. After a short time delay, the examinee must then pick the missing picture of the

pair from four possible responses. Half of the item pairs consist of pictures, whereas the other half consist of geometric patterns. These items require the examinee to determine relationships between objects. The items become progressively harder as the examinee moves through the test (e.g., pictures contain more details; distractor responses look more similar to the correct response).

The Reasoning scale questions require the examinee to look at two rows of boxes. The examinee must recognize the relationship between the items as they move across the row. In the bottom row, the final box contains a question mark. The examinee must choose the correct picture to complete the analogy from the four possible responses listed below the item. The items become more difficult as the examinee progresses (i.e., contains more detail, requires the examinee to pay attention to more than one dimension). This task requires attention to the orientation and details of the picture. Half of the items contain pictures and symbols to create the analogy, while the other half uses geometric patterns.

The Symbolic scale is comprised of items that use pictures and other concrete representations, while the Nonsymbolic scale contains items that use geometric patterns. Half of the items in the Symbolic scale are from the Memory scale and half are from the Reasoning scale. Similarly, half of the items in the Nonsymbolic scale are from the Memory scale and half from the Reasoning scale. The UNIT-GAT includes demonstration and sample items in both of the primary scales (Memory and Reasoning) to ensure that the examinee fully understands the task before moving to items scored for credit. The UNIT-GAT is administered through the use of universal signs and one-word verbal directions (e.g., stop, look, watch) by the examiner.

Internal consistency calculations for the second-grade sample obtained for this study show Cronbach alphas of .86 for the Memory Scale, .78 for the Reasoning Scale, .82 for the Symbolic Scale, and .72 for the Nonsymbolic Scale. The fifth-grade showed Cronbach alphas of .77 for the Memory Scale, .80 for the Reasoning Scale, .43 for the Symbolic Scale, and .53 for the Nonsymbolic Scale. The ninth-grade sample yielded Cronbach alphas of .41 for the Memory Scale, .50 for the Reasoning Scale, .14 for the Symbolic Scale, and .18 for the Nonsymbolic Scale.

Naglieri Nonverbal Ability Test. The Naglieri Nonverbal Ability Test (NNAT; Naglieri, 2003) is a nonverbal general ability measure for children and adolescents ages 5 to 17 years. The test has 2 forms, A and B, both of which have 72 items. Each form yields a total standard score (mean = 100, SD = 15). The NNAT is comprised of four types of question formats; Pattern Completion, Reasoning by Analogy, Serial Reasoning, and Spatial Visualization.

Pattern Completion questions require an examinee to look at a pattern design with a piece missing and determine which of five choices complete the pattern. The examinee must extend the potential answers to the original pattern to be able to complete the answer. This task requires considerable attention to both the details of the pattern and the general orientation. These items are generally found in the levels intended for elementary students.

Reasoning by Analogy questions require the examinee to look at a two columns of boxes with geometric shapes in them. The examinee must recognize the relationship between the boxes as they move down the column and across the row. The examinee must pay attention to many different details of the design, including shading, orientation

of the figure, and the change in the overall design. These items become more difficult in complexity of design and the number of dimensions as the examinee advances. Serial Reasoning items are constructed of a series of shapes that change as they move across the rows and down the columns. Typically the items change position in each row. These items require the examinee to recognize the sequence of the shapes, even with a varying starting shape. The items become more difficult as the examinee progresses through the test. For example, items may progress from a simple shape sequence to a shape sequence and color sequence in the same item.

Spatial visualization questions are made up of a series of boxes that contain geometric figures. The examinee is asked to visualize what two or more designs in a row would look like if combined. Additionally, the designs in the columns can be combined to make the design in the bottommost box of each column. This requires the examinee to recognize that the shapes can be combined in different ways. The NNAT is administered by the examiner verbally reading the directions and two sample items. Examinees then complete the remainder of the test individually within the given time frame.

Psychometric properties of the test are determined through reliability and validity studies reported in the Technical Manual. Kuder-Richardson Formula #20 reliability coefficients for the full score Nonverbal Ability Index are generally high, with all falling above .80 (Naglieri, 1997b). Kuder-Richardson Formula #21 reliability coefficients of cluster scores were lower, as expected. However, some cluster scores were as low as .25 (i.e., Spatial Visualization for grade 2).

Data from validity studies show that the NNAT produced very small differences between males and females, white (mean = 99.3) and African-American (mean = 95.1)

students, white (mean = 101.4) and Hispanic (mean = 98.6) students, white (mean = 103.6) and Asian (mean = 103.9) students, (Naglieri & Ronning, 2000), Hispanic students with limited-English proficiency (mean = 98.0) and Hispanic students without limited-English proficiency (mean = 96.7) (Naglieri, Booth, & Winsler, 2004), and Native American students and white students (Kaufman & Naglieri, 2002). Researchers found similar correlations between the ability of the NNAT to predict achievement for white, black, and Asian groups (Naglieri, 1985; Naglieri & Ronning, 2000). Recently attempts have been made to extend the NNAT to the purpose of gifted screening. According to Naglieri and Ford (2003) the NNAT identified similar percentages of white, black, and Asian students. These results differ from previous studies showing that the Raven Progressive Matrices identified more minority and economically-disadvantaged students as needing further assessment for giftedness than the NNAT (Stephens, Kiger, Karnes, & Whorton, 1999). According to Maller & Mowery (2000), there remains a shortage of validity studies on the NNAT, its most serious weakness.

UNIT-Gifted Screening Scales. The Universal Nonverbal Intelligence Test- Gifted Screening Scales (UNIT-GSS; McCallum & Bracken, in press) is a screening scale completed by classroom teachers designed to quickly identify those students who are in need of additional testing for gifted services. The scale is comprised of two clusters and eight scales. The General Aptitude cluster consists of four scales: Cognitive Aptitude, Creative Aptitude, Emotional Aptitude, and Leadership Aptitude. The Cognitive Aptitude Scale assesses abstract and logical reasoning, problem-solving ability, memory, cognitive speed, and quantitative facility. The Creative Aptitude Scale assesses the ability to produce useful and novel solutions to problems through divergent thinking. The

Emotional Aptitude Scale assesses the ability to get along with peers, recognize one's own and other's emotions, and manage emotions. The Leadership Aptitude Scale measures the examinee's ability to inspire confidence in others, successfully lead and positively influence group behavior.

The Specific Academic Aptitude Cluster consists of four scales; Language Arts Aptitude, Math Aptitude, Reading Aptitude, and Science Aptitude. The Language Arts Aptitude Scale assesses the student's ability to use written and spoken language. The Math Aptitude Scale measures the examinee's ability to use numbers, solve mathematical problems, and understand numerical relationships. The Reading Aptitude Scale measures the ability to read fluently, prosodically, and with comprehension. The Science Aptitude Scale measures interest and abilities used in the process of analyzing the relationships found in nature and the experimental investigation of phenomena.

The UNIT-GSS is intended to be used by the teachers of students aged 5 through 18 years. Teachers are instructed to rate all statements based on their knowledge of the examinee and relative to his or her same-aged peers in the local environment. Teachers are instructed to take the native language of the examinee into account, and to focus on the examinee's communication ability, regardless of the language or medium used. Each scale is comprised of 15 questions rated with a numerical ranking system ranging from 1 (well below average) to 5 (well above average). A rating of 2 indicates below average performance, a rating of 3 indicates average performance, and a rating of 4 indicates above average performance.

The UNIT-GSS is constructed to allow raters to compare the examinee to other peers in the local environment. Standardization data are used primarily to establish

variability in the population for future comparison. Reliability information indicates that the UNIT-GSS scales are highly reliable, with no scale falling below .95 in a recent study (Gray, 2006). Correlations between the UNIT-GSS and Terra Nova Comprehensive Test of Basic Skills (CTBS; CTB, 1996) test scores indicate that the UNIT-GSS significantly correlated with student's performance in math, reading/language arts, and science with all correlations falling above .54 and most above .60. Internal consistency calculations for this sample show Cronbach alphas ranging from .96 to .98 for the total sample.

Procedures

Data for this study was collected by Riverside Publishing Company at an elementary and high school in Southeast Tennessee. The UNIT-GAT and the NNAT were administered to 95 examinees in groups of approximately 15 students, thirty-three examinees in the second grade, thirty examinees in the fifth grade, and thirty-two examinees in the ninth grade. The examiner received training on the administration of the instruments before data collection began. The tests were administered in counterbalanced order (i.e., approximately 15 students in each grade were administered the UNIT-GAT first and the other 15 were administered the NNAT first) to minimize the effects of test administration order. The UNIT-GSS was provided to the primary teacher of each grade. The teacher was instructed in scoring procedures and asked to randomly choose fifteen students in their grade and complete the UNIT-GSS. The data set also contains demographic information on the school district, including ethnicity, Title I status, and socioeconomic level, completed by the principal of each school.

During administration of the UNIT-GAT, the examiner stood in front of the classroom with all the students sitting at desks facing her. The Memory scale was always

administered first. The examiner held the administration demonstration card next to her so that all children could see it. She then pointed to the first set of demonstration items, said the word "look," tapped her temple, and said "study." She then nodded, indicating a relationship between the items. She then pointed to the second set of demonstration items and nodded to indicate a relationship between the items. The students were given ten seconds to look at the paired pictures. The demonstration card was turned over to show the paired items, with one item replaced with a question mark, and four possible options to go in the question mark. The examiner then pointed to the first item, then the question mark and made a shrugging motion to the students. She then pointed to each possible option indicating whether the option was correct by shrugging or making a "thumbs up" sign, ending with the correct option. The examiner then demonstrated filling in a circle below the correct option with a magic marker. The procedures were repeated using the administration sample card except the examiner did not indicate which of the four possible responses was correct, instead pointing to the students and to the response books to indicate that students were to choose the correct response in their test booklets. During the scored test, students were given one minute to examine twenty pairs of items. They then flipped to a blank page for 10 seconds, then turned to the answer page. They had one minute to choose their correct responses. After one minute, they were told "stop."

The Reasoning scale was administered directly after the Memory scale. The examiner held the Reasoning demonstration card up. On each side, the card contained two rows of two boxes with a question mark in the bottom right box and four possible responses at the bottom. The examiner said "look," pointed to the top two boxes and nodded, indicating their relationship, and then pointed to the bottom row. She pointed to

the first picture, then to the question mark and shrugged. She then pointed to each of the four possible responses, indicating a correct or incorrect choice with a shrug or thumbs up sign. She then filled in the circle under the correct answer with a magic marker. The examiner then flipped the demonstration card over and repeated the procedures for the second demonstration question. Administration of the sample items also followed the same procedures, except the examiner did not indicate which of the possible responses was correct, instead pointing to the students and to the response booklets to indicate that they should choose the correct response (in their test booklets). When the demonstration and sample items were completed, students turned to the scored test item page. They then were told "begin," and given thirty minutes to complete the thirty-six items. At the end of the thirty minutes, students were told to "stop."

Administration of the NNAT consisted of the examiner reading a paragraph explaining the test to the students. The students were then timed for thirty minutes. The UNIT-GSS was administered by the examiner explaining the directions of the test to the teachers. The teachers also had a copy of the directions and the examiners contact information they could consult. Each of the UNIT-GSS scales contained fifteen questions on which the teacher rated the student, using a Likert scale of one through five, with one indicating a well below average skill level and five meaning a well above average skill level.

Inter-rater Reliability.

Inter-rater reliability was assessed by having an independent second rater score a photocopy of 10 UNIT-GAT, 10 NNAT protocols from each grade and 5 UNIT-GSS protocols from each grade. Reliability percentage was determined by adding the number

of agreements and disagreements and then dividing by the number of agreements. In the instance of a disagreement, a third rater examined both protocols to ensure that the correct score for each student was recorded. Reliability ranged from 90% to 100% with an average reliability of 93%.

Chapter 3

Results

The purpose of this study was to determine the concurrent validity of the UNIT-GAT. Validity was assessed by the administration of the UNIT-GAT, NNAT, and UNIT-GSS to a sample of ninety-three students in the second, fifth, and ninth grades. Descriptive statistics for the total sample and individual grade levels on each assessment instrument are displayed in Tables 1 through 4. All tables are located in the Appendices. Correlations were classified using Cohen's (1988) ratings. A correlation of below r = .10 was negligible, r = .11 to r = .30 was considered weak, r = .31 to r = .50 was considered moderate, and r = .51 to r = .70 was considered strong. Correlations above r = .71 were considered very strong. It is important to note that the UNIT-GAT is currently in the standardization phase of production; therefore, standard scores are unavailable and raw scores were used for all analyses.

Relationship between the UNIT-Group Ability Test and the Naglieri Nonverbal Ability Test.

Correlations between the UNIT-GAT and NNAT are listed in Tables 5 through 8. For the total sample, the four scales of the UNIT-GAT correlated with the total raw score of the NNAT at r = .25 ($p < .05$) for the UNIT-GAT Memory Scale, r = .29 ($p < .01$) for the UNIT-GAT Reasoning Scale, r = .17 for the UNIT-GAT Symbolic Scale, and r = .50 ($p < .01$) for the UNIT-GAT Nonsymbolic Scale, respectively. The total raw score of the UNIT-GAT correlated with the total raw score of the NNAT at an r = .35 ($p < .01$) level.

Because there was some variability in the correlations as a function of class, data from the three classes are reported. Second-grade student's UNIT-GAT total raw score

correlated with the NNAT r = .74 ($p < .01$) and had UNIT-GAT scale score correlations ranging from r = .61 (Memory Scale) to r = .79 (Nonsymbolic Scale). All of the correlations for the second grade were significant at the $p < .01$ level. Fifth-grade student's UNIT-GAT total raw scores correlated with the NNAT at an r = .30 level. Their scale score correlations ranged from r = -.07 (Memory Scale) to an r = .33 (Reasoning Scale). None of the correlations between the UNIT-GAT total or scale scores and the NNAT were significant. Ninth-grade students showed similar correlations to the fifth-grade, with the UNIT-GAT total raw score correlating at an r = .29 level and the UNIT-GAT scale score correlations ranging from r = -.12 (Symbolic Scale) to r = .49 (Nonsymbolic Scale; $p < .01$).

Relationship between the UNIT-Group Ability Test and the UNIT-Gifted Screening Scales.

Correlations between the UNIT-GAT total and scale scores and UNIT-GSS composite and scale scores for the total sample are shown in Table 9. In general, the sample showed negligible to weak correlations with the composites and scales of the UNIT-GSS, with only one relationship correlating at a statistically significant level (between the UNIT-GAT Memory Scale and the UNIT-GSS Emotional Scale, r = -.31; $p < .05$). The correlations of the four scales of the UNIT-GAT and the UNIT-GSS General Aptitude Composite ranged from r = -.03 (Memory Scale) to r = .21 (Reasoning Scale). The total raw score of the UNIT-GAT correlated weakly and nonsignificantly at r = .13. Examination of the correlations between the UNIT-GAT scales and the scales comprising the General Aptitude Composite (Cognitive, Creative, Emotional, and Leadership), reveals correlations ranging from r = -.31 (between the Memory Scale of the UNIT-GAT

and the Emotional Scale of the UNIT-GSS) to r = .21 (between the Nonsymbolic Scale of the UNIT-GAT and the Creative Scale of the UNIT-GSS). The total raw score of the UNIT-GAT correlated with the scales of the UNIT-GSS at levels ranging from r = -.06 (Emotional Scale) to r = .19 (Creative Scale). Correlation coefficients between the scale scores of the UNIT-GAT and UNIT-GSS Specific Academic Aptitudes Composite ranged from r = -.03 (Symbolic Scale) to r = .17 (Nonsymbolic Scale). The total raw score of the UNIT-GAT correlated with the UNIT-GSS Specific Academic Aptitudes Composite at a level of r = .06. The scales of the UNIT-GAT and the scales of the GSS correlated at levels ranging from r = -.14 (Memory Scale of the UNIT-GAT and Science Scale of the UNIT-GSS) to r = .22 (Nonsymbolic Scale of the UNIT-GAT to the Reading Scale of the UNIT-GSS). The total raw score of the UNIT-GAT correlated with the UNIT-GSS scales from r = -.06 with the Science Scale to r = .12 with the Reading Scale.

Relationship Between the UNIT-Group Ability Test and the UNIT-Gifted Screening Scales by Grade.

Correlations between the UNIT-GAT and the UNIT-GSS by grade level are shown in Table 10 through 12. The second-grade students showed correlations between the UNIT-GAT and UNIT-GSS ranging from weak to moderate, with none of the correlations statistically significant. Correlation coefficients between the UNIT-GAT scale scores and the UNIT-GSS General Aptitudes Composite ranged from r = .08 with the Memory Scale to r = .20 with the Nonsymbolic Scale. The total raw score of the UNIT-GAT correlated at a level of r = .14. The correlations between the scale scores of the UNIT-GAT and the scale scores of the UNIT-GSS ranged from r = -.27 between the Symbolic Scale of the UNIT-GAT and Emotional Scale of the UNIT-GSS to r = .48

between the Nonsymbolic Scale of the UNIT-GAT and the Cognitive Scale of the UNIT-GSS. The total raw score of the UNIT-GAT correlations ranged from r = -.22 with the Emotional Scale to r = .35 with the Cognitive Scale. Correlation coefficients between the scales of the UNIT-GAT and the Specific Academic Aptitudes Composite score ranged from r = .12 with the Memory Scale to r = .28 with the Nonsymbolic Scale. The total raw score of the UNIT-GAT correlated at a rate of r = .21. The scales of the UNIT-GAT and UNIT-GSS had correlations ranging from r = -.21, between the Symbolic Scale of the UNIT-GAT and Science Scale of the UNIT-GSS, to r = .43, between the Nonsymbolic Scale of the UNIT-GAT and the Language Arts Scale of the UNIT-GSS. The total raw score also showed a negative correlation to the Science scale (r = -.10) and a moderate correlation with the Language Arts Scale (r = .35).

The fifth-grade students generally showed correlations in the weak to moderate/strong range, with several reaching statistical significance. Between the scales of the UNIT-GAT and the General Aptitudes Composite of the UNIT-GSS, the students had correlations of r = -.19 for the UNIT-GAT Memory Scale, r = .27 with the UNIT-GAT Symbolic Scale and the UNIT-GSS General Aptitude Composite, r = .47 with the UNIT-GAT Nonsymbolic Scale and the UNIT-GSS General Aptitude Composite, and r = .53 between the UNIT-GAT Reasoning Scale and the UNIT-GSS General Aptitude Composite. The UNIT-GAT total raw score for the fifth-grade students correlated at r = .44 with the UNIT-GAT General Aptitude Composite. Correlations between the scales of the UNIT-GAT and the scales of the UNIT-GSS ranged from r = -.52 between the Memory Scale of the UNIT-GAT and the Emotional Scale of the UNIT-GSS and r = .51 ($p < .05$) between the Reasoning Scale of the UNIT-GAT and the Emotional Scale of the

UNIT-GSS. The total raw score of the UNIT-GAT showed correlations of r = .12 with both the Emotional and Leadership Scales of the UNIT-GSS, r = .41 with the Creative Scale of the UNIT-GSS, and r = .44 with the Cognitive Scale of the UNIT-GSS. Correlations between the Specific Academic Aptitudes Composite and the scales of the UNIT-GAT varied from r = .28 with both the Memory and Reasoning Scales to r = .63 ($p < .05$) with the Nonsymbolic Scale. The total raw score correlated at r = .57 ($p < .05$) level. The correlations between the scale scores of the UNIT-GAT and the scales of the UNIT-GSS ranged from r = .07, between the UNIT-GAT Memory Scale and the UNIT-GSS Science Scale, and r = .66 ($p<.05$) between the UNIT-GAT Nonsymbolic Scale and the UNIT-GSS Reasoning Scale. The total raw score correlated at an r = .46 level with the Math Scale, r = .51level with the Science Scale, r = .57 ($p <. 05$) level with the Language Arts Scale, and r = .62 ($p < .05$) with the Reading Scale.

Correlations between the scales of the UNIT-GAT and the General Aptitudes Composite for the ninth-grade students ranged from r = -.13 on the Nonsymbolic Scale to r = .10 on the Symbolic Scale. The total UNIT-GAT raw score correlated at r = -.02. The correlations between the scales of the UNIT-GAT and UNIT-GSS ranged from r = -.32 between the Nonsymbolic Scale and the Cognitive Scale to r = .29 between the UNIT-GAT Reasoning Scale and UNIT-GSS Leadership Scale. The UNIT-GAT total raw score correlations ranged from r = -.26 on the UNIT-GSS Cognitive Scale to r = .26 on the UNIT-GSS Leadership Scale. Correlations between the UNIT-GAT scales and the Specific Academic Aptitudes Composite varied between r = -.31 for the UNIT-GAT Memory Scale and the UNIT-GSS Specific Academic Aptitude scores and r = .28 for the UNIT-GAT Reasoning Scale and the UNIT-GSS Specific Academic Aptitude Scale. All

correlations with the Specific Academic Aptitudes Composite were negative for the ninth-grade sample. The UNIT-GAT total raw score correlation was r = -.06. Correlations between the scales of the UNIT-GAT and the scales of the UNIT-GSS Specific Academic Aptitude scales ranged between r = -.24 (between the Symbolic Scale and Language Arts Scale) and r = .45 (between the Nonsymbolic Scale and Science Scale). The UNIT-GAT total raw score correlations ranged between r = -.15 on the UNIT-GSS Language Arts Scale and r = .37on the UNIT-GSS Science Scale.

Relationship Between the Naglieri Nonverbal Ability Test and the UNIT-Gifted Screening Scales.

Correlations between the NNAT and the UNIT-GSS are shown in Table 13. In general the NNAT shows small/moderate correlations to the composites and scales of the UNIT-GSS, although several of the relationships for specific grades do reach the moderate range and statistical significance (e.g., second grade Creative Scale, r = .52, $p < .05$). For the most part, correlations between the NNAT and the UNIT-GSS General Aptitude Composite are of a lesser magnitude than correlations between the UNIT-GAT total raw score and the UNIT-GSS General Aptitude Composite, ranging from r = -.21 between the ninth-grade sample and the UNIT-GSS General Aptitude Composite to r = .10 between the second-grade sample and the UNIT-GSS General Aptitude Composite. The NNAT total sample raw score correlated r = -.10 with the UNIT-GSS General Aptitude Composite. For the Specific Academic Aptitude Composite, the NNAT total raw score showed stronger correlations for the total sample and the ninth-grade sample than the UNIT-GAT total raw score. The second grade correlated at r = .29 between the NNAT total raw score and the UNIT-GSS Specific Academic Aptitude Composite, the

fifth grade correlated at r = .43, and the ninth grade correlated at r = -.08. The total NNAT sample showed a correlation of r = .19 to the UNIT-GSS Specific Academic Aptitude Composite.

Relative Predictive Efficiency of the UNIT-Group Ability Test and Naglieri Nonverbal Ability Test to the UNIT-Gifted Screening Scales.

The capability of the UNIT-GAT total raw score and NNAT to predict teacher-completed ratings of the General Aptitude Composite of the UNIT-GSS was determined through stepwise multiple regression analysis. Results are displayed in Table 14. In the first step of the model, the UNIT-GAT total raw score was entered, based on its stronger correlation with the General Aptitude Composite than the NNAT total raw score. In the first model, the UNIT-GAT was not found to be significant (R^2= .02, p = .43), only accounting for 2% of the variance. In the second step, the NNAT total raw score was entered (R^2 = .05, p = .41). This accounted for an additional 3% of the variance in the scores of the General Aptitude Composite and was also nonsignificant.

The ability of the UNIT-GAT total raw score and the NNAT total raw score to predict teacher ratings on the UNIT-GSS Specific Academic Aptitude Composite was examined through stepwise multiple regression. Results are shown in Table 15. In the first step, the NNAT was entered, accounting for 4% of the variance at a nonsignificant level (R^2 = .04; p = .23). In the second step, the UNIT-GAT total raw score was added, but did not explain any additional variance (R^2 = .04; p = .49) and still was not significant.

Chapter 4

Discussion

The primary goal of this study was to evaluate the concurrent validity qualities of the UNIT-GAT and the NNAT and their ability to predict achievement as measured by the UNIT-GSS. In general, results of the analysis indicated that the UNIT-GAT has moderate concurrent validity with the NNAT and variable predictive validity with the UNIT-GSS, ranging from small correlations in the total sample to moderate and strong correlations in the fifth-grade sample. These results indicate that the UNIT-GAT may be an acceptable measure of nonverbal intelligence. More predictive studies will need to be conducted to fully determine its ability to predict achievement.

The UNIT- Group Ability Test.

Correlations between the total and scale scores of the UNIT-GAT and NNAT were not unexpected. First the overall correlation coefficient is lower than the coefficient between the UNIT-GAT Reasoning Scale and NNAT. As a matrix analogies test measuring a student's ability to reason using items that are nonsymbolic in nature, the NNAT should correlate better with the Reasoning and Nonsymbolic Scales of the UNIT-GAT than both the UNIT-GAT Memory and Symbolic Scales. This pattern occurred, but these correlations are still only moderately strong. The magnitude of the correlations between the UNIT-GAT Reasoning and the NNAT may have been limited as the UNIT-GAT Reasoning Scale contains both symbolic and nonsymbolic items, whereas the NNAT contains only nonsymbolic. Thus, the criterion variable (NNAT) may be more limited than the predictor.

As previously noted, the total sample total score of the UNIT-GAT and NNAT correlated r = .35 (p < .01) but the UNIT-GAT Nonsymbolic Scale and NNAT correlated r = .50 (p < .01). Perhaps this pattern is possible because the UNIT-GAT overall score is assessing cognitive components above those assessed by the NNAT. The most obvious (added) component is memory, which is not tapped by the NNAT. Additionally, the modest UNIT-GAT Symbolic Scale and NNAT total score correlation could be limited by the unidimensional nature of the criterion variable (NNAT). Overall these scores indicate that the UNIT-GAT shows fair to good concurrent validity with the NNAT, as the most similar scales have correlations in the moderate and strong range. Across the three grades, there is a noticeable decrease in the strength of the correlations. This decrease in correlational strength is most likely attributable to error, indicated by the decrease in internal consistency of the UNIT-GAT scales across age.

The UNIT- Group Ability Test and UNITT- Gifted Screening Scales.

Results of the correlations between the UNIT-GAT and the UNIT-GSS were lower than anticipated. The total sample showed correlations in the negligible to weak range, with only one correlation (between the Memory Scale of the UNIT-GSS and Emotional Scale of the UNIT-GAT, r = -.31, p < .05) reaching statistical significance, but in a negative direction. In general, Nonsymbolic Scale scores correlated more strongly and positively with measures of the Specific Academic Aptitude Composite of the UNIT-GSS then did other scales of the UNIT-GAT. Perhaps indicating that the Nonsymbolic items are stronger psychometrically. The second-grade students showed small correlations with the Language Arts Scale reaching a moderate correlation with the total raw score, Reasoning Scale, and Nonsymbolic Scale. The fifth-grade students showed

small to strong correlations, with their UNIT-GAT total raw score significantly correlating ($p < .05$) with both composites and almost all scales. Additionally, the Nonsymbolic Scale showed moderately strong and statistically significant ($p < .05$) correlations with both composites and several scales in the Specific Academic Aptitude Composite. The ninth-grade students correlated negatively with both of the composites and half of the UNIT-GSS scales. These results indicate that the UNIT-GAT may not be related to teacher-perceived abilities and actual ability for very young students and for older students who do not spend a significant portion of their day with one instructor, than for those in middle school. It should be noted that, due to the small sample size, the results of the overall correlations between the UNIT-GAT and total UNIT-GSS sample are heavily influenced by the ninth-grade results.

The UNIT-GSS ratings of the ninth-grade students, who transition to a different instructor for each class, were considerably lower than the correlations of the second and fifth grades. Of all groups, the fifth-grade students showed the highest correlations between their UNIT-GSS ratings and performance on the two intelligence measures. There are several possible reasons for this occurrence. First, as mentioned above, teachers of high school students may not know their students as well as teachers of younger students who stay with students all day. Second, while the UNIT-GSS requires teachers to be familiar with the student, many teachers will not be knowledgeable in all areas assessed, particularly those who teach high school students, resulting in an uninformed and inaccurate assessment of the student's abilities. Third, the stronger correlations of the fifth-grade may have been produced because the teachers of these students could have more experience in attempting to develop many of the skills rated then is the case for

teachers of younger or older students. For example, in the Specific Academic Aptitudes Composite, the second-grade students correlated negatively (r = -09) with the Science Scale of the UNIT-GSS, while the fifth-grade showed a strong correlation of .52 ($p < .05$). Several of the questions contained in the Science scale are related to topics not generally addressed systematically in second-grade curriculum (e.g., question 13, "understand scientific concepts") but are topics taught in the fifth-grade. Additionally, many of the qualities measured by the UNIT-GSS may not be developmentally sensitive for children as young as the second grade (e.g., Question 12 of the Emotional Scale, "is diplomatic in confrontational situations"), resulting in lower correlations between their ratings and performance on the intelligence measures.

The results of the correlations between the UNIT-GAT and UNIT-GSS should be investigated further. The correlations indicate that a relationship between the intellectual abilities of the students and their perceived ability by their teachers may exist. A previous study investigated the relationship between the UNIT-GSS and measures of achievement, intelligence, and emotional stability (Gray, 2006). Gray found that, in a sample of 106 students, the UNIT-GSS Cognitive Aptitude Scale showed significant correlations ($r = .85$, $p < .01$) with the intellectual measure of the Gifted Rating Scales (Pfieffer & Janoseqich, 2003). Correlations between the Math Aptitude of the UNIT-GSS and the Math (r = .60), Reading/Language Arts (r = .63), and Science (r = .63). Composite scores of the CTBS (CTB, 1996) were moderately strong. The Language Arts Aptitude of the UNIT-GSS correlated r = .57 with the CTBS Math Composite, r = .64 with the CTBS Reading/Language Arts Composite, and r = .60 with the CTBS Science Composite. The UNIT-GSS Reading Aptitude correlated r = .54 with the CTBS Math Composite, r = .64

with the CTBS Reading/Language Arts Composite, and r = .62 with the CTBS Science Composite. The UNIT-GSS Science Aptitude correlated r = .56 with the CTBS Math Composite, r = .60 with the Reading/Language Arts Composite, and r = .62 with the CTBS Science Composite. All of the correlations (with the exception of UNIT-GSS Reading Aptitude and CTBS Reading/Language Arts Composite) were significant at the $p < .01$ level.

The results obtained by Gray (2006) showed the UNIT-GSS to have a moderate to strong relationship with standardized achievement scores. The low correlations between the composite and scale scores of the UNIT-GSS and the UNIT-GAT indicate a low ability by the UNIT-GAT to predict achievement scores. However, the correlations are highly variable by grade level. The UNIT-GAT should be directly correlated to standardized achievement measures to further investigate this relationship. Additionally, to further strengthen the research base on the UNIT-GSS, studies should be conducted to determine the relationship between the UNIT-GSS and established verbal intellectual measures.

Relative Predictive Ability of the UNIT- Group Ability Test and Naglieri Nonverbal Ability Test.

Tables 14 and 15 show the ability of the UNIT-GAT total raw score and NNAT total score to predict the General Aptitude and Specific Academic Aptitude Composite scores of the UNIT-GSS. Examination of the tables reveals that none of the variables significantly predicts either the General Aptitude or the Specific Academic Aptitude Composite. These results are inconsistent with previous studies that have found the NNAT to show moderate to strong and statistically significant correlations with measures

of mathematics and reading achievement (Naglieri, & Ronning, 2000; Naglieri, Booth, & Winsler, 2004). These discrepancies could be due to the previously discussed difficulties with teacher-ratings and should be investigated further.

Limitations of the Current Study and Future Directions.

There are several limitations in the current study. The first is that the standardization phase of the UNIT-GAT has not yet been completed and all analyses were conducted with raw scores. While using raw scores does not affect the strength or direction of the relationships, standard scores should be obtained to make mean difference comparisons with other standardized instruments such as the NNAT. Other studies should focus on determining relationships between the UNIT-GAT and other instruments. In addition, although the UNIT-GSS has been found to correlate strongly with end of year achievement scores in a previous study (Gray, 2006), these results indicate that more research should be conducted.

Examination of the testing environment during administration of the instruments reveals several variables that may help to explain the low correlations between the Memory Scale of the UNIT-GAT and the UNIT-GAT total raw score for the fifth and ninth grade students. Due to the availability of rooms, the second grade students completed each test in their regular classroom, sitting at their desk with the teacher present. The fifth grade students transitioned to the room for testing. The ninth grade students completed the two instruments in their regular homeroom, which also served as the school's band practice room. Both the fifth-grade and ninth-grade teachers left the room during administration. While these changes are small, they may have resulted in some distraction or confusion and negatively impacted student performance on the

UNIT-GAT Memory scale which was always administered first. Additionally, although the two instruments were administered according to the standardized instructions, there were no fidelity checks completed. In the future, researchers should be careful to include fidelity checks to ensure adherence to standardization procedures.

The sample size of the current study included students who were from a low socio-economic, rural area. The ethnic diversity of the school district was small at 13% Hispanic or Asian and 87% White. Future researchers should be careful to ensure a population that is more representative of the U.S. school population. Additionally, validity studies investigating the identification rates for specific populations should be conducted. The increasing attention given to the lack of diversity and under representation of minorities in gifted programs and the overrepresentation of minority groups in special education (Donovan & Cross, 2002, Fuhrman, 2005) has made the accurate identification of these groups a primary concern of educators. Central to this issue is the ability of the screening measures to predict end of year grades or standardized achievement scores. The UNIT-GAT did not predict UNIT-GSS ratings for the students in the current study, and its predictive ability to the student's grades, achievement scores, and other test scores is needed. For example, while the UNIT-GSS has been shown to strongly correlate with standardized achievement scores (Gray, 2006), a direct prediction between the UNIT-GAT and standardized achievement is needed. If the UNIT-GAT is found to have low correlations with other measures of achievement, then this would provide evidence that the UNIT- GAT may not be a useful measure for screening purposes. If the UNIT-GAT shows moderate to high correlations with measures of achievement, then the low correlation between the UNIT-GAT and UNIT-GSS

Composites and scales may be an indication of the UNIT-GSS' low ability to predict intelligence.

The increasing diversity in U.S. schools has led to increased scrutiny in nonverbal assessment. The ability of nonverbal assessment to accurately measure the abilities of diverse groups of students has made it a natural choice when attempting to determine potential giftedness. The UNIT-GAT will become the only multi-faceted, nonverbal, group-administered intellectual assessment on the market. Currently in the standardization phase of development, the test shows some promise for efficiently screening large groups of students and may be helpful in this manner.

References

Abdel, K., & Ahmed, M. (1988). Egyptian results on the Standard progressive Matrices. *Personality and Individual Differences, 9*, 193-195.

Abdel, K., & Ahmed, M. (2005). Reliability and factorial validity of the Standard Progressive Matrices among Kuwaiti children ages 8 to 15 years. *Perceptual and Motor Skills, 101,* 409-412.

Alvino, J., McDonnel, T. C., & Richert, S. (1981). National survey of identification practices in gifted and talented education. *Exceptional Children, 48,* 124-132.

Anastasi, A., & Urbina, S. (1997). *Psychological Testing.* Upper Saddle River, NJ: Prentice Hall.

Angelini, A. L., Alves, I. C. B., Cutodio, E. M., & Duarte, W. F. (1989). The Sao Paulo norms for the Raven's Coloured Progressive Matrices. *Psychological Test Bulletin, 2*(2), 46-49.

Barnabas, I. P., Kapur, M., & Rao, S. (1995). Norm development and reliability of coloured progressive matrices test. *Journal of Personality and Clinical Studies, 11*, 17-22.

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies*. New York, NY: Cambridge University Press.

Bellah, C. G. (2001). [Review of the Beta III.] In J.C. Impara & B.S. Plake, (Eds.), *The sixteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Benbow, C. P., & Minor, L. L. (1990). Cognitive profiles of verbally and mathematically precocious students: Implications for identification of the gifted. *Gifted Child Quarterly, 34,* 21-26.

Bhogle, S., & Prakash, I. (1992). Performance of Indian children on the Coloured Progressive Matrices. *Psychological Studies, 37*(2-3), 171-181.

Boake, C. (2002). From the Binet-Simon to the Weschler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24,* 383-405.

Boatman, T. A., Davis, K. G., & Benbow, C. P. (1995). Best practices in gifted education. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 1083-1095). Washington, DC: National Association of School Psychologists.

Bourdier, G. (1964). Utilization et nouvel etalonnage du PM 47. *Bulletin de Psychologie, 18*(18)*,* 29-34.

Bracken, B. A., & McCallum, R. S. (in press). *Universal Nonverbal Intelligence Test-Group Ability Test*. Itasca, IL: Riverside.

Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside.

Bracken, B. A., & Naglieri, J. A. (2003). Assessing diverse populations with nonverbal tests. In C. R. Reynolds & R. W. Kamphaus (Eds.) *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement, second edition.* New York: Guildford Press.

Braden, J. P. (1992). Intellectual assessment of deaf and hard-of-hearing people: A quantitative and qualitative research synthesis. *School Psychology Review, 21,* 82-95.

Carney, R. N. (2001). [Review of the InView.] In J. C. Impara & B. S. Plake, (Eds.), *The*

sixteenth mental measurements yearbook (pp.478-480). Lincoln, NE: Buros Institute of Mental Measurements.

Cattell, J. M., & Farrand, L. (1896). Physical and mental measurements of the students of Columbia University. *Psychological Review, 3,* 618-648.

Chan, J. (1989). The use of Raven's Progressive Matrices in Hong Kong: A critical review. *Psychological Test Bulletin, 2*(2), 40-45.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences, second edition. Hillsdale, NJ: LEA.

Coleman, M. R., & Gallagher, J. J. (1995). State identification policies: Gifted students from special populations. *Roeper Review, 17*, 268-275.

Costenbader, V., & Ngari, S. M. (2001). A Kenya standardization of the Raven's Coloured Progressive Matrices. *School Psychology International, 22*, 258-268.

Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R., & Crewther, S. G. (2005). A normative and reliability study for the Raven's Coloured Progressive Matrices for primary school aged children from Victoria, Australia. *Personality and Individual Differences, 39,* 647-659.

CTB. (1996). Terra Nova Comprehensive Test of Basic Skills (5th ed.). Monterey, CA: CTB/McGraw-Hill.

CTB-McGraw Hill. (2000). *InView*. Monterey, CA: Author.

Day, H. E., Fusfield, I. S., & Pintner, R. (1928). *A survey of American schools for the deaf.* Washington, DC: National Research Council.

DeStefano, L. (1999). [Review of the Otis-Lennon School Ability Test, Seventh Edition.]

In J.C. Impara & B.S. Plake, (Eds.), *The fourteenth mental measurements yearbook* (pp.875-879). Lincoln, NE: Buros Institute of Mental Measurements.

Donovan, S. M., & Cross, C. T. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.

Drever, J., & Collins, M. (1928). *Performance tests of intelligence.* Edinburgh: Oliver & Boyd.

DuBois, P. H. (1970). *A history of psychological testing.* Boston: Allyn & Bacon.

Fuhrman, A. S. (2005). The squeaky wheel gets the oil: Why special education classrooms are filled with boys. Unpublished doctoral dissertation, Widener University.

Gray, R. (2006). Concurrent Validity of the Gifted Screening Scale. Unpublished doctoral dissertation, The University of Tennessee, Knoxville.

Gudjonsson, G. H. (1995). Raven's norms on the SPM revisited: A reply to Raven. *Personality and Individual Differences, 18*, 447.

Harcourt Brace Educational Measurement (1996). *Stanford Achievement Test, Ninth Edition,* San Antonio, TX: Author.

Hooper, V. S. (2003). Concurrent and predictive validity of the Universal Nonverbal Intelligence Test and the Leiter International Performance Scale-revised. Unpublished doctoral dissertation, The University of Tennessee, Knoxville.

Jarosewich, T., Pfeiffer, S. I., & Morris, J. (2002). Identifying gifted students using teacher rating scales: A review of existing instruments. *Journal of Psychoeducational Assessment, 20*, 322-336.

Kaufman, J. C., & Naglieri, J. A. (2002). Comparison of Native American children to

other ethnicities on the Naglieri Nonverbal Ability Test.

Kazlauskaite, V., & Lynn, R. (2002). Two year test-retest reliability of the Colored Progressive Matrices. *Perceptual and Motor Skills, 95,* 354.

Kellogg, C., & Morton, N. W. (1999). *Beta III.* San Antonio, TX: PsychoCorp/Harcourt Assessment.

Lohman, D. F., & Haggen, E. P. (2001). *Cognitive Abilities Test- Form 6*. Itasca, IL: Riverside Publishing.

Maller, S. J., & Mowery, D. (2000). Naglieri Nonverbal Ability Test review. *Journal of Psychoeducational Assessment, 18,* 275-280.

Marco, G. L. (1999). [Review of the Leiter International Performance Scale-Revised.] 1 In J. C. Impara & B. S. Plake, (Eds.), *The fourteenth mental measurements yearbook* (pp. 683-687). Lincoln, NE: Buros Institute of Mental Measurements.

McCallum, R. S. (2003). Context of nonverbal assessment of intelligence and related abilities. In R.S. McCallum (Ed.), *Handbook of Nonverbal Assessment* (pp. 3-21). New York: Kluwer/Plenum Press.

McCallum, R. S., & Bracken, B. (in press a). *Gifted Screening Scale.* Itasca, IL: Riverside.

McCallum, R. S., & Bracken, B. (in press b) Universal Nonverbal Intelligence Test – Group Ability Test. Itasca, IL: Riverside.

McCallum, R. S., Bracken, B., & Wasserman, J. (2001). *Essentials of Nonverbal Assessment.* New York: Wiley.

Mills, C. J., & Tissot, S. L. (1995). Identifying academic potential in students from

under-represented populations: Is using the Ravens Progressive Matrices a good idea? *Gifted Child Quarterly, 39,* 209-217.

Naglieri, J. A. (1985). *Matrix Analogies Test-Short form.* San Antonio, TX: The Psychological Corporation.

Naglieri, J. A. (1997a). *Naglieri Nonverbal Ability Test.* San Antonio, TX: The Psychological Corporation.

Naglieri, J. A. (1997b). *Naglieri Nonverbal Ability Test Technical Manual.* San Antonio, TX: The Psychological Corporation.

Naglieri, J. A. (2000). Weschler intelligence tests. In A.E. Kazdin (Ed.) *Encyclopedia of Psychology, Vol. 8.* (pp. 239-240). Washington, DC: American Psychological Association.

Naglieri, J. A., Booth, A. L., & Winsler, A. (2004). Comparison of Hispanic children with and without limited English proficiency on the Naglieri Nonverbal Ability Test. *Psychological Assessment, 16,* 81-84.

Naglieri, J. A., & Ford, D. (2003). Addressing under-representation of gifted minority children using the Naglieri Nonverbal Ability Test {NNAT}. *Gifted Child Quarterly, 47,* 155-160

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, NY: Guildford.

Naglieri, J. A., & Ronning, M. E. (2000). Comparison of white, African-American,

Hispanic and Asian children on the Naglieri Nonverbal Ability Test. *Psychological Assessment, 12,* 328-334.

Otis, A. S., & Lennon, R. T. (1996). *Otis-Lennon School Ability Test, Seventh Edition.* San Antonio, TX: The Psychological Corporation.

Owen, K. (1991). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences, 13,* 149-159.

Pfeiffer, S. I. (2001). Professional psychology and the gifted: Emerging practice opportunities. *Professional Psychology: Research and Practice, 32*, 175-180.

Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted Rating Scales*. San Antonio, TX: The Psychological Corporation.

Pintner, R. (1924). Results obtained with the non-language group tests. *Journal of Educational Psychology, 15,* 473-483.

Pintner, R., & Patterson, D. G. (1915). The Binet Scale and the deaf child. *Journal of Educational Psychology, 6,* 201-210.

Pintner, R., & Patterson, D. G. (1921). Learning tasks for deaf children. *Psychological Monographs, 29,* 132-136.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41,* 1-48.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven's Progressive Matrices.* Oxford, UK: Oxford Psychologists Press.

Reamer, J. C. (1921). Mental and educational measurements of the deaf. *Psychological Monographs, 29*(3), 1-130.

Roid, G. H. (2003). *Stanford Binet Intelligence Scales, fifth edition.* Scarborough, OH:

Thomas Nelson.

Roid, G. H., & Miller, L. J. (1997). Leiter International Performance Scale-Revised: Examiner's Manual. In G.H. Roid and L.J. Miller, *Leiter International Performance Scale-Revised*. Wood Dale, IL: Stoelting Co.

Rojahn, J., & Naglieri, J. A. (2006). Developmental gender differences on the Naglieri Nonverbal Ability Test in a nationally normed sample of 5-17 year olds. *Intelligence, 34,* 253-260.

Stephens, K., Kiger, L., Karnes, F. A., & Whorton, J. E. (1999). Use of nonverbal measures of intelligence in identification of culturally diverse gifted students in rural areas. *Perceptual and Motor Skills, 88,* 793-796.

Stinnett, T. A., (1999). [Review of the Naglieri Nonverbal Ability Test –Revised.] In J.C. Impara & B.S. Plake, (Eds.), *The fourteenth mental measurements yearbook* (pp. 687-692). Lincoln, NE: Buros Institute of Mental Measurements.

Thompson, B. (2001). [Review of the InView.] In J.C. Impara & B.S. Plake, (Eds.), *The sixteenth mental measurements yearbook* (pp. 480-481). Lincoln, NE: Buros Institute of Mental Measurements.

Trevisan, M. S. (1999). [Review of the Naglieri Nonverbal Ability Test –Revised.] In J.C. Impara & B.S. Plake, (Eds.), *The fourteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

U.S. Bureau of the Census (2000). Language use [Online]. Retrieved October 1, 2006 from www.census.org.

U.S. Department of Education, Office of Educational Research and Improvement (1993).

National excellence: A case for developing America's talent. Washington, DC: Author.

Valencia, R. R. (1984). Reliability of the Raven Coloured Progressive Matrices for anglo and for Mexican-American children. *Psychology in the Schools, 21*, 49-52.

Van der Ven, A. H. G. S., & Ellis, J. L. (2000). A Rasch analysis of Raven's Standard Progressive Matrices. *Personality and Individual Differences, 29,* 45-64.

Vernon, M. (1968). Fifty years of research on the intelligence of deaf and hard-of-hearing children: A review of literature and discussion of implications. *Journal of Rehabilitation of the Deaf, 1,* 1-11. Reprinted in 2005, *Journal of Deaf Studies and Deaf Education, 10,* 225-231.

Vodegel-Matzen, L. B. L., van der Molen, M. W., & Dudink, A. C. M. (1994). Error analysis of Raven test performance. *Personality and Individual Differences, 16,* 433-445.

Wasserman, J. D., & Tulsky, D. S. (2005). A History of Intelligence Assessment. In D.P. Flanagan and P.L. Harrison (Eds.) *Contemporary Intellectual Assessment.* New York: Guilford.

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

Williams, P. C. (1995). Predictive validity of the intellectual constructs identified on the Universal Nonverbal Intelligence Test and the Woodcock-Johnson Psychoeducational Battery –revised with achievement. Unpublished doctoral dissertation, The University of Tennessee.

Yerkes, R. M., Bridges, J. W., & Hardwick, R. S. (1915). *A Point scale for measuring*

mental ability. Baltimore: Warwick and York.

Yeudall, L. T., Fromm, D., Reddon, J. R., & Stefanyk, W. O. (1986). Normative data stratified by age and sex for 12 neuropsychological tests. *Journal of Clinical psychology, 42*(6), 918-946.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army mental tests.* New York: Holt.

Appendices

Table 1.

Descriptive Statistics of UNIT-Group Ability Test (UNIT-GAT) and Naglieri Nonverbal Ability Test Raw Scores for the Total Sample (n=93).

	Mean	Standard Deviation
UNIT-GAT Total Raw Score	30.38	8.10
UNIT-GAT Memory Scale	8.95	4.19
UNIT-GAT Reasoning Scale	21.43	6.06
UNIT-GAT Symbolic Scale	15.71	5.12
UNIT-GAT Nonsymbolic Scale	14.67	3.85
Naglieri Nonverbal Ability Test	17.39	6.30

Table 2.

Means and (Standard Deviations) of UNIT-Group Ability Test (UNIT-GAT) and Naglieri Nonverbal Ability Test Raw Scores for Second, Fifth, and Ninth-Grade Students.

	Second Grade n = 32	Fifth Grade n = 30	Ninth Grade n = 31
UNIT-GAT Total Raw Score	36.91 (9.17)	27.37 (5.73)	26.55 (3.62)
UNIT-GAT Memory Scale	11.19 (4.95)	7.23 (3.69)	8.29 (2.60)
UNIT-GAT Reasoning Scale	25.72 (5.53)	20.13 (6.00)	18.26 (3.79)
UNIT-GAT Symbolic Scale	20.50 (5.13)	13.63 (3.07)	12.77 (2.46)
UNIT-GAT Nonsymbolic Scale	16.41 (4.49)	13.73 (3.55)	13.77 (2.73)
Naglieri Nonverbal Ability Test	16.81 (6.91)	14.87 (5.28)	20.42 (5.40)

Table 3.

Descriptive Statistics of UNIT – Gifted Screening Scales Composite and Scale Scores for Total Sample (n = 93).

	Mean	Standard Deviation
General Aptitudes Composite	3.11	.49
Cognitive Aptitude Scale	3.01	.76
Creative Arts Aptitude Scale	3.13	.54
Emotional Aptitude Scale	3.25	.77
Leadership Aptitude Scale	3.07	.72
Academic Aptitudes Composite	3.11	.62
Language Arts Aptitude Scale	3.09	.78
Math Aptitude Scale	3.12	.76
Reading Aptitude Scale	3.13	.63
Science Aptitude Scale	3.05	.51

Table 4.

Means and (Standard Deviations) of UNIT – Gifted Screening Scales Composite and Scale Scores for Each Grade.

	Second Grade n = 15	Fifth Grade n = 15	Ninth Grade n = 11
General Aptitudes Composite	3.11 (.41)	3.30 (.52)	2.89 (.49)
Cognitive Aptitude Scale	3.00 (.54)	3.36 (.88)	2.63 (.67)
Creative Arts Aptitude Scale	3.15 (.35)	3.33 (.75)	2.87 (.34)
Emotional Aptitude Scale	3.21 (.63)	3.34 (1.11)	3.18 (.41)
Leadership Aptitude Scale	3.06 (.54)	3.24 (.86)	2.87 (.75)
Academic Aptitudes Composite	2.96 (.43)	3.30 (.90)	3.05 (.20)
Language Arts Aptitude Scale	2.88 (.56)	3.32 (1.14)	3.07 (.30)
Math Aptitude Scale	3.01 (.49)	3.41 (1.02)	2.91 (.60)
Reading Aptitude Scale	3.05 (.62)	3.22 (.86)	3.12 (.17)
Science Aptitude Scale	2.87 (.31)	3.24 (.77)	3.01 (.03)

Table 5.

Correlations Between the Total and Scale Raw Scores of UNIT-Group Ability Test (UNIT-GAT) and the Naglieri Nonverbal Ability Test for Total Sample (n = 93).

	UNIT-GAT Total	UNIT-GAT Memory	UNIT-GAT Reasoning	UNIT-GAT Symbolic	UNIT-GAT Nonsymbolic
UNIT-GAT Total raw score					
UNIT-GAT Memory Scale	.69**				
UNIT-GAT Reasoning Scale	.86**	.22*			
UNIT-GAT Symbolic Scale	.93**	.66**	.79**		
UNIT-GAT Nonsymbolic Scale	.87**	.57**	.77**	.63**	
Naglieri Nonverbal Ability Test	.35**	.25*	.29**	.17	.50**

*p <.05
**p <.01

Table 6.

Correlations Between the Total and Scale Raw Scores of UNIT-Group Ability Test (UNIT-GAT) and the Naglieri Nonverbal Ability Test for Second-Grade Students (n = 32).

	UNIT-GAT Total	UNIT-GAT Memory	UNIT-GAT Reasoning	UNIT-GAT Symbolic	UNIT-GAT Nonsymbolic
UNIT-GAT Total raw score					
UNIT-GAT Memory Scale	.86**				
UNIT-GAT Reasoning Scale	.89**	.53**			
UNIT-GAT Symbolic Scale	.96**	.79**	.88**		
UNIT-GAT Nonsymbolic Scale	.95**	.84**	.81**	.81**	
Naglieri Nonverbal Ability Test	.74**	.61**	.68**	.62**	.79**

**All correlations significant at the p <.01 level.

Table 7.

Correlations Between the Total and Scale Raw Scores of UNIT-Group Ability Test (UNIT-GAT) and the Naglieri Nonverbal Ability Test for Fifth-Grade Students (n = 30).

	UNIT-GAT Total	UNIT-GAT Memory	UNIT-GAT Reasoning	UNIT-GAT Symbolic	UNIT-GAT Nonsymbolic
UNIT-GAT Total raw score					
UNIT-GAT Memory Scale	.25				
UNIT-GAT Reasoning Scale	.80**	-.38*			
UNIT-GAT Symbolic Scale	.84**	.28	.63**		
UNIT-GAT Nonsymbolic Scale	.89**	.16	.75**	.50**	
Naglieri Nonverbal Ability Test	.30	-.07	.33	.18	.32

*p <.05
**p <.01

Table 8.

Correlations Between the Total and Scale Raw Scores of UNIT-Group Ability Test (UNIT-GAT) and the Naglieri Nonverbal Ability Test for Ninth-Grade Students (n = 31).

	UNIT-GAT Total	UNIT-GAT Memory	UNIT-GAT Reasoning	UNIT-GAT Symbolic	UNIT-GAT Nonsymbolic
UNIT-GAT Total raw score					
UNIT-GAT Memory Scale	.29				
UNIT-GAT Reasoning Scale	.76**	-.41*			
UNIT-GAT Symbolic Scale	.66**	.36	.37*		
UNIT-GAT Nonsymbolic Scale	.74**	.07	.65**	-.03	
Naglieri Nonverbal Ability Test	.29	-.08	.34	-.12	.49**

*p < .05
**p <.01

Table 9.

Correlations Between UNIT-Group Ability Test Total and Scale Raw Scores and UNIT-Gifted Screening Scales (UNIT-GSS) Composites and Scales for Total Sample.

	UNIT-Group Ability Test				
UNIT-GSS	Total raw score	Memory Scale	Reasoning Scale	Symbolic Scale	Nonsymbolic Scale
General Aptitudes Composite	.13	-.03	.21	.11	.14
Cognitive Scale	.15	.19	.08	.12	.18
Creative Scale	.19	.11	.20	.16	.21
Emotional Scale	-.06	-.31*	.15	-.05	-.05
Leadership Scale	.08	-.09	.17	.07	.08
Specific Academic Aptitudes Composite	.06	.05	.05	-.03	.17
Language Arts Scale	.07	.12	.01	-.03	.20
Math Scale	.08	.07	.07	.03	.15
Reading Scale	.12	.11	.10	.04	.22
Science Scale	-.06	-.14	.01	-.13	.05

*p <.05

Table 10.

Correlations Between UNIT-Group Ability Test Total and Scale Raw Scores and UNIT-Gifted Screening Scales (UNIT-GSS) Composites and Scales for Second-Grade Students.

	UNIT-Group Ability Test				
UNIT-GSS	Total raw score	Memory Scale	Reasoning Scale	Symbolic Scale	Nonsymbolic Scale
General Aptitudes Composite	.14	.08	.18	.09	.20
Cognitive Scale	.35	.24	.39	.21	.48
Creative Scale	.29	.17	.35	.26	.29
Emotional Scale	-.22	-.22	-.17	-.27	-.13
Leadership Scale	.11	.13	.07	.15	.05
Specific Academic Aptitudes Composite	.21	.12	.25	.13	.28
Language Arts Scale	.35	.24	.39	.25	.43
Math Scale	.22	.21	.18	.19	.22
Reading Scale	.14	.05	.21	.08	.20
Science Scale	-.10	-.17	.01	-.21	.06

Table 11.

Correlations Between UNIT-Group Ability Test Total and Scale Raw Scores and UNIT-Gifted Screening Scales Composites (UNIT-GSS) and Scales for Fifth-Grade Students.

	UNIT-Group Ability Test				
UNIT-GSS	Total raw score	Memory Scale	Reasoning Scale	Symbolic Scale	Nonsymbolic Scale
General Aptitudes Composite	.44	-.19	.53	.27	.47
Cognitive Scale	.44	.45	.04	.25	.48
Creative Scale	.41	.07	.30	.27	.41
Emotional Scale	.12	-.52	.51	.12	.08
Leadership Scale	.12	-.35	.38	.03	.17
Specific Academic Aptitudes Composite	.57*	.28	.28	.33	.63*
Language Arts Scale	.57*	.39	.20	.33	.64*
Math Scale	.46	.22	.24	.23	.55*
Reading Scale	.62*	.35	.26	.37	.66*
Science Scale	.51	.07	.39	.35	.51

*p <.05

Table 12.

Correlations Between UNIT-Group Ability Test Total and Scale Raw Scores and UNIT-Gifted Screening Scales (UNIT-GSS) Composites and Scales for Ninth-Grade Students.

	UNIT-Group Ability Test				
UNIT-GSS	Total raw score	Memory Scale	Reasoning Scale	Symbolic Scale	Nonsymbolic Scale
General Aptitudes Composite	-.02	-.05	.01	.10	-.13
Cognitive Scale	-.26	-.28	-.07	.00	-.32
Creative Scale	.04	.28	-.22	-.01	.06
Emotional Scale	-.10	-.01	-.11	.12	-.25
Leadership Scale	.26	.03	.30	.20	.15
Specific Academic Aptitudes Composite	-.06	-.31	.28	-.06	-.03
Language Arts Scale	-.15	-.18	-.03	-.24	-.01
Math Scale	-.01	-.15	.13	.08	-.08
Reading Scale	-.05	-.13	.06	.09	-.14
Science Scale	.37	.28	.24	.00	.45

Table 13.

Correlations Between Naglieri Nonverbal Ability Test Total Score for Total Sample, Second-Grade, Fifth-Grade, and Ninth-Grade Students and UNIT-Gifted Screening Scales (UNIT-GSS) Composites and Scales.

	Naglieri Nonverbal Ability Test			
UNIT-GSS	Total Sample	Second Grade	Fifth Grade	Ninth Grade
General Aptitudes Composite	-.10	.10	.02	-.21
Cognitive Scale	.03	.34	.32	-.23
Creative Scale	.18	.52*	.45	-.18
Emotional Scale	-.22	-.22	-.25	-.16
Leadership Scale	-.28	-.12	-.43	-.15
Specific Academic Aptitudes Composite	.19	.29	.43	-.08
Language Arts Scale	.19	.36	.34	.19
Math Scale	.07	.36	.32	-.24
Reading Scale	.23	.18	.47	-.00
Science Scale	.19	.10	.56*	.15

*p <.05

Table 14.

Stepwise Multiple Regression Analysis Predicting UNIT-Gifted Screening Scales General Aptitude Composite from UNIT-Group Ability Test and Naglieri Nonverbal Ability Test for Total Sample (n = 40).

Prediction models	Beta	R^2	ΔR^2	Significance level
Model 1		.02	.02	.43
UNIT-GAT total raw score	.13			.43
Model 2		.05	.03	.41
UNIT-GAT total raw score	.21			.24
NNAT total raw score	-.19			.29

Table 15.

Stepwise Multiple Regression Analysis Predicting UNIT-Gifted Screening Scales Specific Academic Aptitude Composite from UNIT-Group Ability Test and Naglieri Nonverbal Ability Test for Total Sample (n = 40).

Prediction models	Beta	R^2	ΔR^2	Significance level
Model 1		.04	.04	.23
NNAT total raw score	.19			.23
Model 2		.04	.00	.49
NNAT total raw score	.20			.26
UNIT-GAT total raw score	-.02			.89

Vita

Stacy Lynn Bliss was born in Hebron, Nebraska, on July 11th, 1979. She was raised in Hebron and matriculated through the Hebron public school system, graduating in 1997. She attended Peru State College and graduated with a B.S. double major in Psychology/Sociology and Social Sciences in 2001. Stacy worked as a Treatment Specialist with the Outreach and Intensive Treatment Services department of Beatrice State Developmental Center for several years before returning to school for her advanced degrees. She graduated from the University of Tennessee, Knoxville with her Master of Science in Applied Educational Psychology in August of 2007.

Stacy is currently completing her doctoral studies in education with an emphasis in School Psychology at the University of Tennessee, Knoxville. She will be an intern at the Munroe Meyer Institute in Omaha, Neb., beginning July 2007.