CONSTRUCT AND CRITERION-RELATED VALIDITY OF THE
DRAW A PERSON: A QUANTITATIVE SCORING SYSTEM FOR NORMAL, READING DISABLED,
AND DEVELOPMENTALLY HANDICAPPED CHILDREN

A Thesis
Presented in Partial Fulfillment of the Requirements for
the degree Master of Arts in the Graduate School of
The Ohio State University

by
Jean Marie Smith, B.B.A.

* * * * *

The Ohio State University
1987

Master's Examination Committee:
Jack A. Naglieri
Judy L. Genshaft

Approved by

Adviser
College of Education
To Danny
I would like to express my appreciation to Dr. Jack A. Naglieri for his support and guidance throughout the completion of this thesis. I would also like to thank him for teaching me the importance of and stimulating my interest in the field of research. In addition, I would like to extend my appreciation to Dr. Judy L. Genshaft for the direction and encouragement she provided in completing this project.

I want to thank Sean Reardon, Susan Johnston, and Joe Cantrell for their help scoring the individual drawings. Special thanks goes to Nancy Boltz for the extra time and effort she spent scoring numerous drawings. My sincere appreciation also goes to my family and friends who continually provided support and encouragement throughout the research.
VITA

June 4, 1958 ....................... Born - Kent, Ohio

1980 ............................. B.B.A. in Accounting
    Kent State University, Kent, Ohio

1980 - 1985 ....................... Auditor
    BancOhio Corporation

FIELD OF STUDY

Major Field: School Psychology
# TABLE OF CONTENTS

DEDICTION ...................................................... ii  
ACKNOWLEDGEMENTS ........................................ iii  
VITA ............................................................ iv  
LIST OF TABLES ................................................ vii  

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
</tbody>
</table>
| II. LITERATURE REVIEW | 5  
- Goodenough-Harris Drawing Test | 5  
- Draw A Person: A Quantitative Scoring System | 10  
- Kaufman Test of Educational Achievement | 13  
- Purpose of Study | 15 |
| III. METHOD | 20  
- Subjects | 20  
- Instruments | 22  
- Draw A Person: A Quantitative Scoring System | 22  
- Goodenough-Harris Drawing Test | 23  
- Kaufman Test of Educational Achievement | 24  
- Procedure | 25 |
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Age, Sex, and MAT-SF Means and Standard Deviations for the Normal, Reading Disabled (RD), and Developmentally Handicapped (DH) Samples</td>
<td>22</td>
</tr>
<tr>
<td>3.</td>
<td>Means, Standard Deviations, and t Test Values for the Draw A Person and the Goodenough-Harris for Reading Disabled</td>
<td>29</td>
</tr>
<tr>
<td>5.</td>
<td>Correlations Between the Draw A Person and the Goodenough-Harris</td>
<td>35</td>
</tr>
<tr>
<td>6.</td>
<td>Means and Standard Deviations of the K-TEA by Sample</td>
<td>36</td>
</tr>
<tr>
<td>7.</td>
<td>Correlations Between the Draw A Person and the K-TEA</td>
<td>37</td>
</tr>
<tr>
<td>8.</td>
<td>Correlations Between the Goodenough-Harris and the K-TEA</td>
<td>39</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Studies relating children's drawings to developmental progress have been investigated since the late 1800s. Many of the early studies provided insightful information regarding developmental aspects of drawings that generally, has remained unaltered to this day. The evidence of developmental processes in drawing set the stage for utilizing drawings as measures of intelligence (Harris, 1963).

Based on this notion that children's drawings reflect intellectual development, Florence Goodenough developed a simple, useful, and comprehensive method for interpreting children's drawings of a man. The Goodenough Draw-A-Man Test (Goodenough, 1926) consistently ranked among the most popular tests in clinical services in the United States from its inception in 1926 (Sundberg, 1961).

In 1963, Dale Harris revised the Goodenough Draw-A-Man Test to include updated procedures, norms, and scoring systems. The Harris revision provided a more extensive, presumably more objective scoring system (Dunn, 1967a). In addition, standard scores (mean of 100 and standard deviation of 15) were substituted for the IQ measures previously used and the woman and self drawings were included as possible alternate forms to the man drawing (Scott, 1981).
Anastasi and Dunn commended the Goodenough-Harris Drawing Test (Harris, 1963) and recommended its use to others (Buros, 1972). Numerous studies involving the Goodenough-Harris Drawing Test have been published since its development in 1963. Many of these investigations have addressed the test's validity and reliability. Harris' revision has been found to be a stable and reliable measure that does in fact discriminate between children's performance from one age level to the next (Scott, 1981). It also has been shown to correlate with intelligence tests such as the Binet, Wechsler, and McCarthy Scales (Dunn, 1967b; Naglieri & Maxwell, 1981; Scott, 1981; Sutter & Bishop, 1986; Tramill, Edwards, & Tramill, 1980; White, 1979) and has demonstrated good inter- and intrarater reliability (Anastasi, 1982; Gottling, 1985; Naglieri & Maxwell, 1981; Sattler, 1982; Scott, 1981).

Although the Goodenough-Harris test has demonstrated good reliability and validity, it has recently been criticized for its limitations. One limitation of the Goodenough-Harris is item ambiguity (Scott, 1981). In examining scorer's disagreements item by item, systematic errors were found indicating ambiguity in the scoring of 12 items (Phillips, Smith, & Broadhurst, 1973). Another limiting factor is that Harris did not include normative data for the use of the self drawing or a total score comprised of all three drawing scores. It was also suggested that the norms be refined into half-year or quarter-year intervals in order to provide more accurate standard scores (Scott, 1981). Additionally, a more recent standardization sample that is more representative of the U.S. population Census data is needed to replace Harris' norm sample obtained during the 1950s (Naglieri, 1987).
The Draw A Person: A Quantitative Scoring System was developed by Naglieri (1987) to overcome these limitations of the Goodenough-Harris drawing system. As identified in the manual for the Draw A Person test, the major goals of Naglieri's scoring system are to provide a system that has been normed on the most recent census data, reduce subjective scoring, establish norms with half- and quarter-year intervals, provide an objective and efficient scoring system, and provide a composite standard score of all three drawings.

In the development of any new testing instrument, it is important to determine what the test measures and how well it does so. This is achieved by establishing its validity, which can be defined only through examining objective sources of information and utilizing empirical operations (Anastasi, 1982). There are several types of validity, all of which concern the relationship between performance on the test and other independently observable facts related to the characteristics being considered. Three main categories of validity are content, criterion-related, and construct.

Content validity involves the systematic examination of the test content to determine whether it covers the domain to be measured (Anastasi, 1982). The domain must be analyzed to ensure that all major aspects are covered by the test items. Typically, content validity is built into a test when it is initially developed through the choice of appropriate items.

Criterion-related validity helps establish the effectiveness of a test in predicting an individual's behavior in specific situations (Anastasi, 1982). This type of validity is determined by correlating scores on one test to an independent measure of what the test is designed to measure, for example, correlating intelligence tests and achievement tests.
The third type of validity is that of construct validity. Construct validity is the extent to which the test measures a theoretical construct or trait (Anastasi, 1982). In other words, construct validation involves the gathering of data to support the contention that the test is actually a reflection of the construct or attribute it is designed to reflect. Each of these types of validity is important to the usefulness of a test.

In the development of any new test, it is important that its validity be established to ensure that the new test actually measures what it was intended to measure. Due to the need to establish the validity of any new test, this study investigated both the criterion-related validity and the construct validity of the Draw A Person test. Specifically, criterion-related validity is demonstrated by the use of contrasted groups. This type of validity refers to the extent to which test scores differentiate between groups of people (Walsh & Betz, 1985). For this study, the total population of subjects is comprised of a group of normal, reading disabled, and developmentally handicapped children. This sample of three distinct groups of children was utilized to provide evidence of criterion-related validity by the method of contrasted groups.

As stated above, construct validity is the extent to which the test is said to measure a theoretical construct or trait (Anastasi, 1982). This study investigated the construct validity of Naglieri's Draw A Person test by correlating it with the Goodenough-Harris Drawing Test. This method of validity involves the correlation between a new test and a similar earlier test. A second form of construct validity known as convergent validity, is demonstrated by showing that a test correlates highly with other variables with which it should theoretically correlate (Anastasi, 1982). This investigation studies convergent validity by correlating the Draw A Person test, which is intended to measure overall ability, with a measure of academic achievement (Kaufman Test of Educational Achievement).
CHAPTER II

LITERATURE REVIEW

Because the Goodenough-Harris Drawing Test is a forerunner to Naglieri's Draw A Person test, research involving the Goodenough-Harris was found to be especially pertinent to the present study. It is also important because the Goodenough-Harris is being correlated with the Draw A Person to investigate the test's construct validity. Additionally, the present study utilizes a sample of normal, reading disabled, and developmentally handicapped (mentally retarded) children, and therefore, research with these special populations was also considered to be important to the present study. Literature related to the Kaufman Test of Educational Achievement (K-TEA; Kaufman & Kaufman, 1985) is included because it is used to test the Draw A Person's convergent validity. All of the research found to be relevant to this study is described below.

Goodenough-Harris Drawing Test

The Goodenough-Harris Drawing Test (Harris, 1963) is essentially a revision of the Goodenough Draw-A-Man Test. The standardization sample for the Goodenough-Harris scoring system consisted of 2,975 children who were distributed among four geographic areas and were representative of the occupational distribution of the U.S. in 1950. A total of
75 children were included at each of the age levels from 5 to 15 years. Because both Goodenough and Harris found that girls consistently scored slightly higher than boys, separate raw-to-standard score conversion tables are provided for boys and girls. The standard scores have a mean of 100 and a standard deviation of 15.

In an attempt to establish adequate criteria for adolescent drawings, Harris' revision involved the addition of 22 scoring items to the original 51 items on the Goodenough scale, resulting in a total of 73 scoring points for the drawing of a man. In addition to the scoring system for the man, Harris included woman and self drawings as possible alternate forms. The scoring system for the woman consists of 71 individual items. The self drawing is scored using the point scale of the appropriate sex.

With regard to the individual items comprising the Goodenough—Harris scoring system, a systematic examination of differences between scorers revealed ambiguities in definitions of 12 items (Phillips, Smith, & Broadhurst, 1973). Examples of items causing difficulty in scoring were "wrist or ankle shown", "attachment of arms and legs", and "sketching technique". Additionally, it is pointed out that the test manual provides only one illustration for some items, and six items are not illustrated at all. Therefore, even with training and practice, the scorer remains an important variable because of the evidence of ambiguity of specific items.

The relation between the Goodenough Draw-A-Man Test and Harris' revision was first investigated by Harris (1963) using a sample of 6- to 14-year old Canadian Indian children. Harris found a high correlation between the two tests in that they ranged from .91 to .98. Subsequent studies by other researchers relating the Goodenough Draw-A-Man Test
and the Goodenough-Harris Drawing Test, reported generally high correlations, but not as high as those reported by Harris. The median correlation score of the Man scale for the two systems was .86, and the median score for the Woman scale was .76.

As of 1977, the Goodenough-Harris had been related to tests of intelligence in more than 40 studies (Scott, 1981). The overall results of these studies through 1977 showed correlations using the Man scale ranging from .24 to .83 (median of .49). As might be expected, the strongest correlations were found to be with other drawing tests (McCarthy Drawing scales and the Rutgers Drawing tests). The correlation coefficients with these drawing tests ranged from .62 to .83. In line with the overall correlations with intelligence tests, similar correlations were reported when the Goodenough-Harris was correlated with major individual tests of intelligence such as the Stanford-Binet (Terman & Merrill, 1960) and the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974). The median correlation for 15 studies correlating the Goodenough-Harris and the Stanford-Binet Intelligence Scale was .51. In 14 separate studies, the Goodenough-Harris was correlated with the WISC-R which resulted in a median correlation of .50. A few studies investigated the relationship of both the Goodenough Draw-A-Man and the Goodenough-Harris to the Stanford-Binet and the WISC-R. These studies revealed a weaker correlation with Goodenough's Draw-A-Man and intelligence tests than the Goodenough-Harris and intelligence tests.

In a more recent study by Naglieri and Maxwell (1981), the standard scores for the Goodenough-Harris and the McCarthy Draw-A-Child (McCarthy, 1972) were found to be significantly and positively correlated to intelligence for a sample of 60 children, including 20 learning disabled, 20 mentally retarded, and 20 normal children. In a separate study by
White (1979), Goodenough-Harris standard scores were found to correlate moderately with Peabody Individual Achievement Test (PIAT; Dunn & Markwardt, 1970) and WISC-R scores. The Goodenough-Harris correlated .47 with the PIAT and .55 with the WISC-R Full Scale IQ. Both of these correlations are significant (p < .01). The results of this study were duplicated by Sutter & Bishop (1986) using a larger sample of 360 subjects in comparison to the 30 subjects used in White’s study. All resulting correlations in this study were statistically significant (p < .01). The Goodenough-Harris correlated .29 with the PIAT Total Test scores, and .38 with the WISC-R Full Scale IQ.

In studies relating the Goodenough-Harris to global ratings of academic achievement, correlations were moderate and yielded coefficients of approximately .30 (Dudeck, Goldberg, Lester, & Harris, 1969; Pihl & Nimrod, 1976). For the most part, correlations with academic achievement were found to be lower than those with tests of intelligence. In specific areas of reading-related skills and arithmetic skills, correlations were generally low. The Goodenough-Harris correlations with reading-related skills ranged from -.09 to .56, and correlations with arithmetic skills ranged from -.06 to .40.

Because the present study utilizes a sample of normal, reading disabled, and developmentally handicapped children, the following Goodenough-Harris research involves special populations, specifically, learning disabled and mentally retarded subjects.

It was reported by Scott (1981), that a major drawback in studies involving the learning disabled, is that the majority of studies lack normal control subjects as a part of their sample. As of 1977, only two of ten studies completed had utilized normal controls. Those studies with normal controls reported the following results. Goodenough-Harris
scores were found to be significantly lower for a group of low-achieving boys, despite the fact that their California Test of Mental Maturity scores were similar to that of their normal counterparts (Colligan, 1967). A second study with normal controls reported that Goodenough-Harris scores appeared to be comparable for normal and learning disabled children (Hartman, 1972). Overall, the studies involving learning disabled subjects were not comparable due to methodological problems (i.e. lack of normal controls) and a lack of precise definition of variables and samples (Scott, 1981).

Many of the studies involving mentally retarded subjects were qualitative studies rather than quantitative. Harris (1963) reviewed qualitative studies and reported that drawings by mentally retarded subjects consistently include more details and body parts than normal children of the same mental age, but retardates appear to have difficulty with organization, proportion, and dimension. More recent studies involving mentally retarded children support this idea that their drawings include more detail than those of normal children (Scott, 1981). One study matched noninstitutionalized retardates with normal children by mental age as determined by the Stanford-Binet. No significant differences were found in the Goodenough-Harris scores between the two groups. It was concluded that because the drawing of the retardates resembled the work of a chronologically younger child, that the retarded child is not cognitively deficient in an absolute sense, and instead possesses essentially normal abilities that appear slowly.
Draw A Person: A Quantitative Scoring System

Due to the recent development of Naglieri's Draw A Person test, very few studies have been conducted investigating the test's reliability and validity. The investigations that have been conducted to date are described in the test's manual.

The standardization data for the Draw A Person scoring system was collected in the fall of 1984. Data was collected in a group setting from 4,468 individuals ranging in age from 5 to 17 years. Of this data, 2,622 students were selected in order to obtain a sample that was representative of the U.S. population 1980 Census data according to age, sex, race, geographic region, and ethnic group. Additional considerations in selecting the sample included socioeconomic status and community size.

A series of steps were followed in developing the standard scores for the Draw A Person test (mean of 100 and standard deviation of 15). The first step in the development of Total Test standard scores involved a comparison of the Total Test raw score means and standard deviations by sex. In contrast to Harris' separate conversion tables for males and females, it was determined that separate norms by sex were not necessary.

Naglieri's Draw A Person scoring system is organized into four categories, 14 criteria, and 64 individual items. The Draw A Person utilizes the same 64-item scoring system for each of the three drawings (man, woman, and self), unlike the Goodenough-Harris system that has two separate scoring systems for the man and woman. Additionally, Naglieri's system provides a composite standard score which combines the data for all three drawings.

A number of studies are presented in the Draw A Person manual to provide evidence of the test's reliability. Cronbach's alpha was used to calculate the internal consistency.
coefficients on the standardization sample. The Draw A Person reliability coefficients ranged from .83 to .89 (median of .86) indicating good internal reliability. The Draw A Person man, woman, and self drawings were also found to have good internal consistency in that the reliability coefficients ranged from .56 to .78 (median of .70).

In order to compute the test-retest reliability, 112 students from the standardization sample were retested four weeks after the initial testing. The Draw A Person total score retest reliability coefficients ranged from .60 to .89 (median of .73). The test-retest reliability coefficients for the man, woman, and self were consistent with the results for the total score, thus providing support for the test-retest reliability of the Draw A Person over a four-week interval.

Interrater and intrarater reliability was studied using a sample of 45 elementary school children and 44 junior high students (Gottling, 1985). Each drawing by the students was scored by two examiners using the Draw A Person and Goodenough-Harris scoring systems in counterbalanced order. The interrater reliabilities were found to be similar for the Draw A Person and Goodenough-Harris scoring systems ranging from .86 to .95, thereby providing evidence of interrater reliability. These drawings were scored a second time by the same scorer in order to test for intrarater reliability. The correlation coefficients between a rater's initial and second scoring ranged from .89 to .97, and provided evidence of intrarater reliability of the Draw A Person and the Goodenough-Harris scoring systems.

The Draw A Person manual also provides evidence of various forms of validity. According to Anastasi (1982), one method used to examine construct validity is analyzing
developmental changes in mean scores. An analysis of the separate drawing scores and the total test scores indicated a clear increase in raw score as age increased. This evidence of age differentiation provides support of the test’s construct validity. Construct validity was also demonstrated through two different factor analyses with the norm sample. A factor analysis was performed on the composite criterion scores and on the individual item scores. Results of these analyses provided additional evidence of construct validity.

Concurrent validity was investigated by correlating the Draw A Person with measures of nonverbal ability, the Matrix Analogies Test—Short Form (MAT–SF; Naglieri, 1985) and achievement, the Multilevel Academic Survey Test (MAST; Howell, Zucker, & Morehead, 1985) using 1,328 students in grades 4 through 12. Analysis of the Draw A Person and the MAT–SF indicated that the two tests yielded similar mean scores. It was also found that the standard scores for the two tests correlated significantly, ranging from .19 to .27. The Draw A Person standard scores also correlated significantly with the MAST Reading and Math standard scores. Correlations between the Draw A Person Man, Woman, Self, and Total Test standard scores and the MAST Math and Reading standard scores ranged from .19 to .24. The Draw A Person Total Test correlated .24 with the MAST Reading and .21 with the MAST Math, both of which are significant (p < .01). The Draw A Person correlated higher with the measure of nonverbal ability (MAT–SF) than with the measure of achievement (MAST), providing evidence of convergent validity.

Concurrent validity was also investigated by comparing the Draw A Person scoring system with that of the Goodenough–Harris. The drawings for 89 students were scored by one examiner using the two different scoring systems. The two scoring systems were found to yield highly correlated standard scores and correlation coefficients ranging from .75 to
.84, thus supporting the test's concurrent validity.

Additional studies analyzing sex, race, and hispanic/nonhispanic differences are also included in the Draw A Person manual. The results of these individual studies indicate that the Draw A Person standard scores appear not to vary according to sex, race, or Spanish origin.

**Kaufman Test of Educational Achievement**

The Kaufman Test of Educational Achievement (K-TEA; Kaufman & Kaufman, 1985) provides standard scores (mean of 100 and standard deviation of 15) by age (three month increments from 6-0 to 18-11), by grade (one year intervals), and are reported separately for spring testing (February-July) and fall testing (August-January). Standard scores are provided for each of the five subtests and for the Reading Composite, Mathematics Composite, and Battery Composite scores.

Separate standardization programs were conducted in order to provide separate norms for the spring and fall. The spring program included more than 1,400 students and was conducted between April and June of 1983. The fall program included more than 1,000 students and was conducted between September and December of 1983. The samples were stratified by sex, geographic region, socioeconomic status, and race or ethnic group, and was based on the U.S. Census Bureau population reports of 1983 and 1984.

Various reliability and validity studies are included in the K-TEA manual. Split-half reliability is reported by age and by grade for each subtest and for the composite scores. The mean values of split-half reliability coefficients by grade ranged from .90 to .98, while
the mean values by age ranged from .92 to .98.

Test-retest reliability coefficients are reported for each subtest and composite score based on data from 172 students organized into two separate groups: grades 1–6 and grades 7–12. The correlation coefficients for grades 1–6 ranged from .83 to .97, and from .90 to .97 for grades 7–12.

The K-TEA has two separate forms, the Comprehensive Form and the Brief Form. Because of the similarity between the two tests, correlations were obtained between the Comprehensive and Brief forms. The reliability coefficients ranged from .87 to .96 across grade levels, and .90 to .97 across age levels.

Construct validity is demonstrated by the steady increase in subtest scores from grades 1–12 and ages 6–17. Only 18-year-olds did not show a consistent change. Internal consistency was evaluated by correlating subtest scores with the Battery Composite. The mean values obtained ranged from .77 to .85 across grade level and from .82 to .88 across age.

Concurrent validity was demonstrated by correlating the K-TEA with the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983), the Wide Range Achievement Test (WRAT; Jastak & Jastak, 1978), the PIAT, and the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981). The two K-TEA Reading subtests, Reading Composite and Spelling correlated with the WRAT. However, the Mathematics subtests are limited to grades 7–12. The K-TEA Battery Composite correlated .86 with the PIAT Total Test, Reading Composite correlated .82 with each PIAT Reading Subtest, Mathematics Composite correlated .75 with PIAT Mathematics, and Spelling correlated .78.
with PIAT Spelling. The correlational analysis between the K-TEA and the K-ABC provides additional evidence of the concurrent validity of the K-TEA. With regard to the PPVT-R, the Reading Comprehension subtest on the K-TEA consistently had the highest correlation with the PPVT-R.

Purpose of the Study

The studies presented above support the need for research investigating the validity of Naglieri's Draw A Person test. First, and foremost, is the need to establish the validity of a newly developed test, since, the only validity studies to date for the Draw A Person are those described in the test's manual. In addition, no previous studies have been conducted with the Draw A Person using special populations of reading disabled and developmentally handicapped children. Therefore, there is a particular need for a study that includes these special populations. Based on this information, it is evident that this validation study will provide important information pertinent to the establishment of the Draw A Person as a useful nonverbal measure of ability, particularly for special populations.

The present study investigates several forms of validity by correlating the Draw A Person, the Goodenough-Harris, and the K-TEA with three matched samples of normal, reading disabled, and developmentally handicapped children. The following hypotheses are investigated in this study:

1. a. Is there a significant difference between Naglieri's Draw A Person Total Test mean and the Goodenough-Harris Total Test mean for the normal, reading disabled, and developmentally handicapped samples?
b. Is there a significant difference between the Draw A Person Man mean and the Goodenough-Harris Man mean for the normal, reading disabled, and developmentally handicapped samples?

c. Is there a significant difference between the Draw A Person Woman mean and the Goodenough-Harris Woman mean for the normal, reading disabled, and developmentally handicapped samples?

d. Is there a significant difference between the Draw A Person Self mean and the Goodenough-Harris Self mean for the normal, reading disabled, and developmentally handicapped samples?

2. a. Is there a significant difference between the Draw A Person Total Test standard scores between normal, reading disabled, and developmentally handicapped groups?

b. Is there a significant difference between the Draw A Person Man standard scores between normal, reading disabled, and developmentally handicapped groups?

c. Is there a significant difference between the Draw A Person Woman standard scores between normal, reading disabled, and developmentally handicapped groups?

d. Is there a significant difference between the Draw A Person Self standard scores between normal, reading disabled, and developmentally handicapped groups?
3. a. Is there a significant difference between Goodenough-Harris Total Test standard scores between normal, reading disabled, and developmentally handicapped samples?

b. Is there a significant difference between Goodenough-Harris Man standard scores between normal, reading disabled, and developmentally handicapped samples?

c. Is there a significant difference between Goodenough-Harris Woman standard scores between normal, reading disabled, and developmentally handicapped samples?

d. Is there a significant difference between Goodenough-Harris Self standard scores between normal, reading disabled, and developmentally handicapped samples?

4. a. Is there a significant correlation between Draw A Person Total Test standard scores and Goodenough-Harris Total Test standard scores for the normal, reading disabled, and developmentally handicapped samples?

b. Is there a significant correlation between Draw A Person Man standard scores and Goodenough-Harris Man standard scores for the normal, reading disabled, and developmentally handicapped samples?

c. Is there a significant correlation between Draw A Person Woman standard scores and Goodenough-Harris Woman standard scores for the normal, reading disabled, and developmentally handicapped samples?
d. Is there a significant correlation between Draw A Person Self standard scores and Goodenough-Harris Self standard scores for the normal, reading disabled, and developmentally handicapped samples?

5. a. Is there a significant correlation between standard scores for the Draw A Person Total Test and the K-TEA Total Test for the normal, reading disabled, and developmentally handicapped samples?

b. Is there a significant correlation between standard scores for the Draw A Person Man and the K-TEA Total Test for the normal, reading disabled, and developmentally handicapped samples?

c. Is there a significant correlation between standard scores for the Draw A Person Woman and the K-TEA Total Test for the normal, reading disabled, and developmentally handicapped samples?

d. Is there a significant correlation between standard scores for the Draw A Person Self and the K-TEA Total Test for the normal, reading disabled, and developmentally handicapped samples?

6. a. Is there a significant correlation between the standard scores for the Goodenough-Harris Total Test and the K-TEA Total Test for the normal, reading disabled, and developmentally handicapped samples?

b. Is there a significant correlation between the standard scores for the Goodenough-Harris Man and the K-TEA Total Test for the normal, reading
disabled, and developmentally handicapped samples?
c. Is there a significant correlation between the standard scores for the
   Goodenough-Harris Woman and the K-TEA for the normal, reading disabled, and
developmentally handicapped samples?
d. Is there a significant correlation between the standard scores for the
   Goodenough-Harris Self and the K-TEA Total Test for the normal, reading
disabled, and developmentally handicapped samples?

7. a. Is there a significant difference between the Draw A Person Total Test and K-TEA
correlation and the Goodenough-Harris Total Test and K-TEA correlation for each
   of the separate samples (normal, reading disabled, and developmentally
   handicapped)?
b. Is there a significant difference between the Draw A Person Man and K-TEA
correlation and the Goodenough-Harris Man and K-TEA correlation for each of the
   separate samples (normal, reading disabled, and developmentally handicapped)?
c. Is there a significant difference between the Draw A Person Woman and K-TEA
correlation and the Goodenough-Harris Woman and K-TEA correlation for each of the
   separate samples (normal, reading disabled, and developmentally
   handicapped)?
d. Is there a significant difference between the Draw A Person Self and K-TEA
correlation and the Goodenough-Harris Self and K-TEA correlation for each of the
   separate samples (normal, reading disabled, and developmentally handicapped)?
CHAPTER III
METHOD

Subjects

The total sample for this study was comprised of 120 children between the ages of 9 years, 1 month and 12 years, 3 months from a suburban school district with a population of 15,900 students in a Midwestern community. This sample of students was further organized by the school district into groups of reading disabled, developmentally handicapped, and normal children.

A sample of 40 students had been identified as reading disabled by a multidisciplinary team comprised of school personnel. The students were identified as having a two z-score unit difference between their intellectual ability measure and their individual reading achievement test score. That is, they evidenced specifically a reading deficiency as opposed to a deficiency in math or other achievement-related area.

The second group of 40 students had been identified as developmentally handicapped (mentally retarded) by a multidisciplinary team comprised of school personnel. According to Ohio guidelines, developmentally handicapped means “significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior manifested during the developmental period, which adversely affects a child’s performance” (Ohio
Department of Education, 1982, p.5). The rules further require that a developmentally handicapped child shall have an overall intelligence quotient of 80 or below.

The normal subjects consisted of 112 children selected on the basis of sex and age from 10 regular fourth and fifth grade classrooms in two elementary schools for completion of a larger study. Of these 112 students, 55 were administered the K-TEA achievement measure. The normal sample for this study was comprised of 40 students who were randomly selected from the 55 students who had been administered the K-TEA.

The parents of children in the participating classrooms were notified of the study and were given the opportunity to decline participation. The students were then selected for each of the three groups so that the reading disabled, developmentally handicapped, and normal groups were matched as closely as possible on the basis of sex and age. A description of each of the three groups is presented in Table 1. Each of the students were individually administered the K-TEA, while the Draw A Person and the MAT-SF were group administered. The MAT-SF, which measures nonverbal ability, was scored by a doctoral student in school psychology. The results of the MAT-SF and other descriptive variables are provided in Table 1 to further describe and help differentiate between each of the individual groups. These results indicate that the normal sample obtained a MAT-SF mean score of 98 and the reading disabled students earned a MAT-SF mean score of 90, both of which fall in the Average range of intellectual functioning. In contrast, the developmentally handicapped sample obtained a MAT-SF mean score of 76 which falls in the Borderline range of intellectual functioning.
Table 1  
Age, Sex, and MAT-SF Means and Standard Deviations for the Normal, Reading Disabled (RD), and Developmentally Handicapped (DH) Samples

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Sex</th>
<th></th>
<th></th>
<th>MAT-SF</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>Males</td>
<td>Females</td>
<td>Mean</td>
</tr>
<tr>
<td>NORMAL</td>
<td>40</td>
<td>10-6</td>
<td>8 mos.</td>
<td>48%</td>
<td>52%</td>
<td>98</td>
</tr>
<tr>
<td>RD</td>
<td>40</td>
<td>10-6</td>
<td>9 mos.</td>
<td>75%</td>
<td>25%</td>
<td>90</td>
</tr>
<tr>
<td>DH</td>
<td>40</td>
<td>10-9</td>
<td>8 mos.</td>
<td>48%</td>
<td>52%</td>
<td>76</td>
</tr>
</tbody>
</table>

Instruments

**Draw A Person: A Quantitative Scoring System.**

The Draw A Person: A Quantitative Scoring System (Naglieri, 1987) can be administered to a group or individually. The test was intended for use as a nonverbal measure of ability and may be used as part of a test battery or as a screening device. For the Draw A Person test, the students are required to draw three individual pictures (a man, a woman, and self) using a pencil and three separate pieces of paper. The students are allowed a maximum of five minutes to complete each drawing.

The Draw A Person scoring system is organized into four categories, 14 criteria, and 64 individual items. The four categories which are applied as appropriate to the different criteria are presence, detail, proportion, and bonus. The 14 criteria generally relate to
specific body parts such as arms, eyes, fingers, and mouth. Each criterion is comprised of individual items for which the child can earn points. The same 64-item scoring system is utilized for each of the three drawings (man, woman, and self). Additionally, Naglieri's system provides a composite standard score which compiles the data for all three drawings.

Draw A Person standard scores are based on a mean of 100 and standard deviation of 15. Studies providing evidence of the test's reliability and validity are presented in the Draw A Person manual. The Draw A Person internal consistency reliability coefficients ranged from .83 to .89 (median of .86) indicating very good internal consistency. The Draw A Person total score retest reliability coefficients ranged from .60 to .89 (median of .73), while the interrater reliability coefficients ranged for .86 to .95. The manual also includes research providing evidence of the test's construct and concurrent validity.

Goodenough–Harris Drawing Test.

The Goodenough Harris Drawing Test (Harris, 1963), like the Draw A Person test, can be administered individually or in a group setting. The Goodenough–Harris scoring system for the drawing of a man includes 73 items. Harris included woman and self drawings as possible alternate forms to the man drawing. The scoring system for the woman consists of 71 individual items. The self drawing is scored using the point scale of the appropriate sex. According to Harris (1963), the score yielded from the self drawing must be considered as a tentative measure of maturity because it has not been standardized.

The reliability of the Goodenough–Harris Drawing Test has been investigated by numerous researchers using a variety of procedures (Anastasi, 1982). Test-retest reliability coefficients fall in the .50s to .70s, while alternate-form (man/woman) reliability ranges from a low of .72 to a high of .90. Interrater reliabilities fall in the .80s
and .90s (Sattler, 1982).

Sattler (1982) reported that the reliabilities are good, but the validity is not at the same high level. When the Goodenough-Harris test is correlated with various intelligence tests for Anglo groups, the correlations range from .24 to .88 (median of .49), while the correlations range from .22 to .85 (median of .57) for ethnic minority groups. The Goodenough-Harris test was also correlated with various perceptual-motor tests, which resulted in correlations from .39 to .83 (median of .46). Correlations with academic performance and teacher ratings ranged from .03 to .57 (median of .36).

Kaufman Test of Educational Achievement.

The Kaufman Test of Educational Achievement (K-TEA; Kaufman & Kaufman, 1985) is an individually administered achievement test that assesses reading, mathematics, and spelling skills for grades 1 through 12. The K-TEA has two forms, a Brief Form and a Comprehensive Form. The K-TEA Comprehensive Form, which was utilized for this study, is comprised of five subtests which include Reading Decoding, Reading Comprehension, Mathematics Applications, Mathematics Computation, and Spelling. The Comprehensive Form takes about 30 to 60 minutes for grades 1 to 3 and 60 to 75 minutes for grades 4 to 12.

Mathematics Applications, the first subtest, is comprised of 60 items and measures mathematics concepts, applications of mathematical skills, and problem solving. An easel format is utilized whereby the examiner reads the problem and the child is required to solve it using the information provided on the child's side of the easel. The Reading Decoding subtest measures letter and word recognition and also has 60 items. The child is asked to
read aloud letters and words presented on the easel. For the Spelling subtest, the examiner says the word to be spelled and the child writes it. The Reading Comprehension subtest has 50 items that measure single word, sentence, and paragraph comprehension. The fifth subtest is Mathematics Computation, which has 60 items and measures pencil and paper computational skills from basic operations through algebra.

Various reliability and validity studies are included in the K-TEA manual. The mean values of split-half reliability coefficients ranged from .90 to .98, while the test-retest reliability coefficients ranged from .83 to .97. Correlations between the Comprehensive and Brief forms of the test ranged from .90 to .97 across age levels.

Procedure

The man, woman, and self drawings were obtained in a group setting during the same session in which the MAT-SF was administered. The tests were administered by three graduate students (1 female, 2 males) in school psychology who had been trained in the administration procedures as described in the test manuals. In order to counterbalance the administration of the K-TEA, Draw A Person, and MAT-SF, the group administration of the Draw A Person and the MAT-SF was given at a point when half of the class had been administered the K-TEA. As a result, half the class was administered the K-TEA first, and the other half was administered the Draw A Person and MAT-SF first.

The K-TEA was individually administered to each child by five graduate students (1 female, 4 males) in school psychology or a related field of study. The test was administered using standardized procedures as described in the test manual. The test record forms were scored by the examiners and subsequently checked for accuracy by a doctoral student in school psychology.
The drawings were scored by five school psychology graduate students (3 females, 2 males) using both Naglieri's Draw A Person scoring system and the Goodenough-Harris scoring system. The same rater did not score the same drawing using both the Naglieri and Goodenough-Harris scoring systems. Each drawing was scored by two different raters (one scored the drawing using Naglieri's system, and a second rater scored the drawing using the Goodenough-Harris drawing system). Prior to scoring, each of the raters was trained in the scoring procedures as described in the test manuals. After all of the drawings were initially scored by the raters, the computation of the total scores was verified by the author of this study. Additionally, 20% of the forms were rechecked for scoring accuracy.

The Goodenough-Harris scoring system does not provide for a composite score of all three drawings (man, woman, and self). In order to provide for comparability with the Draw A Person Total Test score, a Goodenough-Harris total test score was derived by calculating the mean of the Goodenough-Harris Man, Woman, and Self standard scores for each of the subjects.
CHAPTER IV

RESULTS

The results of the statistical analyses are presented in response to each of the individual questions. The first question looked at mean differences between the Draw A Person and the Goodenough-Harris. Next, the between-group differences for both the Draw A Person and Goodenough-Harris systems are analyzed. Correlational analyses were performed for the remaining questions. The correlations between the Draw A Person and the Goodenough-Harris scoring systems are presented, as well as the correlations between achievement and the two scoring systems (Draw A Person and Goodenough-Harris). Finally, it is determined whether one of the two scoring systems correlates better than the other with achievement.

**Draw A Person and Goodenough-Harris Mean Score Differences**

Means and standard deviations for the Draw A Person and the Goodenough-Harris Man, Woman, Self, and Total Test standard scores are presented in Tables 2, 3, and 4 by sample (normal, reading disabled, and developmentally handicapped). The Bonferroni correction was used to determine the significance (p < .05) of t values between mean differences of the Draw A Person and the Goodenough-Harris. A t value of 2.4 was needed for significance.
### Table 2

Means, Standard Deviations, and t Test Values for the Draw A Person (DAP) and Goodenough-Harris (G-H) for Normals

<table>
<thead>
<tr>
<th></th>
<th>Normals (N = 40)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Difference</td>
<td>t Value</td>
<td></td>
</tr>
<tr>
<td>TOTAL TEST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>104.9</td>
<td>11.8</td>
<td>3.0</td>
<td>1.201</td>
<td></td>
</tr>
<tr>
<td>G-H</td>
<td>101.9</td>
<td>10.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>106.3</td>
<td>12.6</td>
<td>2.8</td>
<td>.978</td>
<td></td>
</tr>
<tr>
<td>G-H</td>
<td>103.5</td>
<td>12.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>103.3</td>
<td>12.6</td>
<td>1.7</td>
<td>.655</td>
<td></td>
</tr>
<tr>
<td>G-H</td>
<td>101.6</td>
<td>10.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>105.7</td>
<td>13.6</td>
<td>4.9</td>
<td>1.815</td>
<td></td>
</tr>
<tr>
<td>G-H</td>
<td>100.8</td>
<td>10.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: t values are Draw A Person/Goodenough-Harris comparisons.

*Difference = Mean score difference between Draw A Person and Goodenough-Harris.

* p < .05 (using Bonferroni correction).
Table 3

Means, Standard Deviations, and t Test Values for the Draw A Person (DAP) and Goodenough-Harris (G-H) for Reading Disabled

<table>
<thead>
<tr>
<th>Reading Disabled (N = 40)</th>
<th>Mean</th>
<th>SD</th>
<th>Difference</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL TEST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>93.6</td>
<td>14.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G-H</td>
<td>88.3</td>
<td>10.8</td>
<td>5.3</td>
<td>1.837</td>
</tr>
<tr>
<td>MAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>96.1</td>
<td>18.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G-H</td>
<td>92.0</td>
<td>12.2</td>
<td>4.1</td>
<td>1.160</td>
</tr>
<tr>
<td>WOMAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>93.8</td>
<td>14.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G-H</td>
<td>86.2</td>
<td>11.8</td>
<td>7.6</td>
<td>2.592*</td>
</tr>
<tr>
<td>SELF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>91.9</td>
<td>15.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G-H</td>
<td>86.8</td>
<td>11.4</td>
<td>5.1</td>
<td>1.661</td>
</tr>
</tbody>
</table>

Note: *values are Draw A Person/Goodenough-Harris comparisons.
*Differences = Mean score difference between Draw A Person and Goodenough-Harris.
* p < .05 (using Bonferroni correction).
Table 4

Means, Standard Deviations, and t Test Values for the Draw A Person (DAP) and Goodenough-Harris (G-H) for Developmentally Handicapped

<table>
<thead>
<tr>
<th></th>
<th>Developmentally Handicapped (N = 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td><strong>TOTAL TEST</strong></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>81.4</td>
</tr>
<tr>
<td>G-H</td>
<td>76.4</td>
</tr>
<tr>
<td><strong>MAN</strong></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>83.3</td>
</tr>
<tr>
<td>G-H</td>
<td>79.6</td>
</tr>
<tr>
<td><strong>WOMAN</strong></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>80.2</td>
</tr>
<tr>
<td>G-H</td>
<td>74.4</td>
</tr>
<tr>
<td><strong>SELF</strong></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>81.9</td>
</tr>
<tr>
<td>G-H</td>
<td>75.2</td>
</tr>
</tbody>
</table>

Note: <sup>6</sup> values are Draw A Person/Goodenough-Harris comparisons.

<sup>6</sup>Difference = Mean score difference between Draw A Person and Goodenough-Harris.

* p < .05 (using Bonferroni correction).
These $t$ values are included in Tables 2, 3, and 4. Only one mean difference was found to be statistically significant. These results indicate that the Goodenough-Harris and Draw A Person scores differed slightly, but not significantly in each group.

The Bonferroni $t$ statistic was utilized because several pairs of scores were being compared which increases the probability of drawing an erroneous conclusion (Silverstein, 1982). When a single pair of scores is being compared, it may be determined that the difference is significant, for example at the .05 level, and the probability that this conclusion was in error would be .05. However, when several pairs of scores are being compared, the probability of an erroneous conclusion is much greater than .05. For example, when four pairs of scores are being compared, the probability level for the entire process is raised to .20 (.05 x 4 comparisons). Silverstein (1982) suggests that the Bonferroni $t$ statistic be utilized to compensate for this discrepancy. Bonferroni $t$ statistics are utilized by dividing the level of probability used to compute the differences required for significance by the number of comparisons. In this study four comparisons are performed for each of the samples (normal, reading disabled, and developmentally handicapped). Therefore, the significance levels for all $t$-tests are determined by dividing .05 by four (number of comparisons), resulting in a .01 significance level. The .01 level of significance is used to achieve a .05 level of significance for the entire process.

Between-group (normal, reading disabled, and developmentally handicapped) differences on Draw A Person standard scores were analyzed using one-way analysis of variance (ANOVA) procedures. An ANOVA was computed separately for the Total Test, Man, Woman, and Self standard scores. Results revealed that the three groups differed significantly on the Draw A Person Total Test, $F(2, 117) = 26.35, p < .001$; Man, $F(2, 117) = 19.12, p < .001$; Woman, $F(2, 117) = 23.21, p < .001$; and Self,
F (2, 117) = 23.67, p < .001.

The ANOVA results indicate that there is at least one significant difference between the normal, reading disabled, and developmentally handicapped samples for each of the Man, Woman, Self, and Total Test standard scores. The Scheffé's test was conducted in order to determine which group means are significantly different from the other group means. These post hoc tests (Scheffé's test) revealed that all groups differed significantly (p = .05) from one another in each of the separate comparisons (Man, Woman, Self, and Total Test). For example, the Draw A Person Man scores differed significantly between the normal and reading disabled sample, between the normal and developmentally handicapped sample, and between the reading disabled and developmentally handicapped sample.

Between-group (normal, reading disabled, and developmentally handicapped) differences for the Goodenough–Harris standard scores were also analyzed using ANOVA procedures. ANOVAs were computed separately for Man, Woman, Self, and Total Test scores. It was found that the standard score means differed significantly between the normal, reading disabled, and developmentally handicapped samples for each of the areas that were analyzed (Man, Woman, Self, and Total Test). The three groups differed significantly on the Goodenough–Harris Total Test, F (2, 117) = 57.3, p < .001; Man, F (2, 117) = 40.09, p < .001; Woman, F (2, 117) = 54.42, p < .001; and Self, F (2, 117) = 52.88, p < .001.

As indicated in Question 2, the ANOVA results indicate that there is at least one difference between the normal, reading disabled, and developmentally handicapped samples for each of the Man, Woman, Self, and Total Test standard scores. Therefore, Scheffé's post-test analyses were conducted to determine which group means differed from the others.
These post hoc tests indicated that all groups differed significantly (p = .05) from one another in each of the separate comparisons (Man, Woman, Self, and Total Test).

**Draw A Person and Goodenough-Harris Correlations**

Pearson product moment correlations were computed to determine the relationship between the Draw A Person and the Goodenough-Harris scoring system for each of the normal, reading disabled, and developmentally handicapped samples. The correlation coefficients are presented in Table 5. A correlation of .36 was needed for significance (p < .01). Each of the correlations between the Draw A Person and the Goodenough-Harris which range from a low of .634 to a high of .891 were found to be statistically significant (See Table 5). These high correlations provide evidence of the construct validity of Naglieri's Draw A Person test.

The correlations between the Draw A Person and Goodenough-Harris were found to be even higher when corrected for "restriction of range". Restriction of range refers to test scores for a particular group being concentrated in a small portion of the possible range of scores (Kaufman, 1972). In other words, group scores have smaller standard deviations than those that are not restricted. "When a group's scores on a test are restricted in range, correlations between the test and other measures will be lower for that group whose scores are not restricted" (Kaufman, 1972, p.2).

In this study, some scores were found to be restricted and others enlarged. In order to compensate for these restricted and expanded standard deviations, a correction formula (Thorndike, 1949, p. 173) is utilized. The formula to correct Pearson product-moment correlations for restriction of range is as follows:
\[ R_{12} = \frac{r_{12}(s_1 / s_1)}{\sqrt{1 - r_{12}^2 + r_{12}^2 (s_1^2 / s_1^2)}} \]  

Because the standard deviations for both tests being correlated were found to be restricted or enlarged, the formula was performed twice for each final correlation. This formula accounted for restricted or expanded standard deviations in each variable and resulted in a final corrected correlation.

The corrected correlations for the Draw A Person and Goodenough-Harris correlations are presented in Table 5. The corrected correlations range from a low of .732 to a high of .917, and provide evidence of the Draw A Person's construct validity.
<table>
<thead>
<tr>
<th></th>
<th>Obtained r</th>
<th>Corrected r&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NORMAL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Test</td>
<td>.688*</td>
<td>.871*</td>
</tr>
<tr>
<td>Man</td>
<td>.681*</td>
<td>.792*</td>
</tr>
<tr>
<td>Woman</td>
<td>.634*</td>
<td>.813*</td>
</tr>
<tr>
<td>Self</td>
<td>.711*</td>
<td>.858*</td>
</tr>
<tr>
<td><strong>READING DISABLED</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Test</td>
<td>.756*</td>
<td>.850*</td>
</tr>
<tr>
<td>Man</td>
<td>.778*</td>
<td>.775*</td>
</tr>
<tr>
<td>Woman</td>
<td>.636*</td>
<td>.732*</td>
</tr>
<tr>
<td>Self</td>
<td>.721*</td>
<td>.794*</td>
</tr>
<tr>
<td><strong>DEVELOPMENTALLY HANDICAPPED</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Test</td>
<td>.881*</td>
<td>.917*</td>
</tr>
<tr>
<td>Man</td>
<td>.891*</td>
<td>.913*</td>
</tr>
<tr>
<td>Woman</td>
<td>.833*</td>
<td>.829*</td>
</tr>
<tr>
<td>Self</td>
<td>.796*</td>
<td>.824*</td>
</tr>
</tbody>
</table>

<sup>a</sup>Corrected correlations account for restricted or expanded standard deviations

* p < .01
Draw A Person and Goodenough-Harris Correlated with Achievement

The K-TEA standard score means and standard deviations for the individual samples are presented in Table 6. Pearsonian correlations between the K-TEA and the Draw A Person (Man, Woman, Self, and Total Test) are summarized in Table 7. These correlations range from a low of .065 to a high of .409. A correlation coefficient of .26 was necessary to achieve significance (p < .05). The correlations between the K-TEA and the Draw A Person for the normal and reading disabled samples were insignificant. The correlations between the K-TEA and the Draw A Person Man, Woman, Self, and Total Test scores for the developmentally handicapped group were significant and ranged from .295 to .409.

Table 6

Means and Standard Deviations of the K-TEA by Sample

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>104.5</td>
<td>11.1</td>
</tr>
<tr>
<td>Reading Disabled</td>
<td>80.0</td>
<td>8.2</td>
</tr>
<tr>
<td>DH</td>
<td>69.6</td>
<td>6.8</td>
</tr>
</tbody>
</table>
Table 7

Correlations Between the Draw-A-Person and the K-TEA

<table>
<thead>
<tr>
<th></th>
<th>Obtained $r$</th>
<th>Corrected $r^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Test</td>
<td>.193</td>
<td>.320*</td>
</tr>
<tr>
<td>Man</td>
<td>.084</td>
<td>.134</td>
</tr>
<tr>
<td>Woman</td>
<td>.220</td>
<td>.341*</td>
</tr>
<tr>
<td>Self</td>
<td>.208</td>
<td>.302*</td>
</tr>
<tr>
<td>READING DISABLED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Test</td>
<td>.154</td>
<td>.276*</td>
</tr>
<tr>
<td>Man</td>
<td>.235</td>
<td>.336*</td>
</tr>
<tr>
<td>Woman</td>
<td>.065</td>
<td>.122</td>
</tr>
<tr>
<td>Self</td>
<td>.073</td>
<td>.127</td>
</tr>
<tr>
<td>DEVELOPMENTALLY HANDICAPPED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Test</td>
<td>.366*</td>
<td>.622*</td>
</tr>
<tr>
<td>Man</td>
<td>.335*</td>
<td>.545*</td>
</tr>
<tr>
<td>Woman</td>
<td>.295*</td>
<td>.494*</td>
</tr>
<tr>
<td>Self</td>
<td>.409*</td>
<td>.655*</td>
</tr>
</tbody>
</table>

$^a$Corrected correlations account for restricted or expanded standard deviations.

$^* p < .05.$
These correlations were also influenced by both restricted and enlarged ranges. Therefore, the formula (1) was utilized to obtain corrected correlations which are presented in Table 7. The corrected correlations range from a low of .122 to a high of .655. The Draw A Person Total Test correlated significantly (p < .05) with the K-TEA for each of the normal, reading disabled, and developmentally handicapped samples.

Table 8 contains the Pearson product moment correlations between the K-TEA and the Goodenough-Harris Drawing test for each of the normal, reading disabled, and developmentally handicapped samples. A correlation of .26 was needed to achieve significance (p < .05). None of the correlations were significant except the K-TEA and Goodenough-Harris Man correlation for the reading disabled sample (.272). The corrected correlations between the Goodenough-Harris and the K-TEA are also presented in Table 8. The resulting correlations were found to be significant for the reading disabled and developmentally handicapped groups only.

A t-test for the difference between correlated correlations (Gullford & Fruchtner, 1978, p. 164) was conducted in order to examine whether there was a significant difference between the Draw A Person and K-TEA correlations and the Goodenough-Harris and K-TEA correlations. This analysis was performed for each of the different Man, Woman, Self, and Total Test scores by sample (normal, reading disabled, and developmentally handicapped). A total of 12 computations were completed, four for each of the individual samples. A t value of 2.4 was necessary to achieve significance (p < .05) using the Bonferroni correction. The t values that resulted ranged from -1.11 to 1.66, none of which were significant.

The corrected correlations obtained in Questions 6 and 7 were used to recalculate the difference between correlated correlations. The resulting t values ranged from a low of
Table 8

Correlations Between the Goodenough-Harris and the K-TEA

<table>
<thead>
<tr>
<th></th>
<th>Obtained r</th>
<th>Corrected r&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Test</td>
<td>.046</td>
<td>.091</td>
</tr>
<tr>
<td>Man</td>
<td>.014</td>
<td>.022</td>
</tr>
<tr>
<td>Woman</td>
<td>.103</td>
<td>.196</td>
</tr>
<tr>
<td>Self</td>
<td>.052</td>
<td>.105</td>
</tr>
<tr>
<td>READING DISABLED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Test</td>
<td>.230</td>
<td>.515*</td>
</tr>
<tr>
<td>Man</td>
<td>.272*</td>
<td>.537*</td>
</tr>
<tr>
<td>Woman</td>
<td>.217</td>
<td>.459*</td>
</tr>
<tr>
<td>Self</td>
<td>.136</td>
<td>.314*</td>
</tr>
<tr>
<td>DEVELOPMENTALLY HANDICAPPED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Test</td>
<td>.258</td>
<td>.623*</td>
</tr>
<tr>
<td>Man</td>
<td>.222</td>
<td>.569*</td>
</tr>
<tr>
<td>Woman</td>
<td>.252</td>
<td>.561*</td>
</tr>
<tr>
<td>Self</td>
<td>.251</td>
<td>.583*</td>
</tr>
</tbody>
</table>

<sup>a</sup>Corrected correlations account for restricted or expanded standard deviations

* p < .05
0.02 to a high of 3.37. Of the twelve resulting $t$ values, three were found to be significant ($p < .05$). The Draw A Person Total Test correlated better with the K-TEA than the Goodenough-Harris Total Test for the normal sample ($t = 3.16$), and the Goodenough-Harris Total Test and Woman correlated better with the K-TEA than the corresponding Draw A Person scores for the reading disabled samples (Total Test, $t = 3.32$; Woman, $t = 3.37$). These inconsistent and, for the most part, insignificant results indicate that neither of the tests (Draw A Person or Goodenough-Harris) correlate better with achievement (K-TEA).
CHAPTER V

DISCUSSION

The purpose of this study was to investigate the construct and criterion-related validity of the Draw A Person: A Quantitative Scoring System. One uniqueness of this study was the examination of criterion-related validity by using the method of contrasted groups (normal, reading disabled, and developmentally handicapped). The construct validity of the test was examined by correlating the Draw A Person with the Goodenough-Harris and with the K-TEA. Results of these statistical procedures provided evidence of both construct and criterion-related validity of the Draw A Person.

Criterion-Related Validity

The first question examined the standard score means between the Draw A Person and the Goodenough-Harris. Only one of the mean score differences was found to be statistically significant, yet, the Draw A Person mean scores were consistently higher than the Goodenough-Harris mean scores. Typically, it is expected that a newer test would yield lower scores than an earlier developed test.

One reason for lower scores with the Goodenough-Harris scoring system, may be due to specific item content. The Goodenough-Harris test retains many outdated clothing
criteria, particularly in the items for the drawing of a woman. For example, points are awarded for a "modeled" skirt indicating pleats, feminine garb, and a clearly feminine shoe. The criteria on Naglieri's test is more general in nature and appears resistant to change in fashion or culture.

Another possible explanation for these mean differences is item ambiguity. Phillips, Smith, and Broadhurst (1973), performed a systematic examination of differences between scorers, and found ambiguities in definitions of scoring for 12 items on the Goodenough-Harris. Examples of items causing difficulty in scoring were "brow or lashes shown", "details of fingers shown", and "modeling technique". It was also pointed out that some items are illustrated only once in the manual, and six items are not illustrated at all. This lack of sufficient scoring guidelines could influence the Goodenough-Harris scores and possibly cause lower mean scores.

A third reason for the higher scores on the Draw A Person was suggested by Naglieri (1987). In a study investigating the relation of the Draw A Person with the Goodenough-Harris using 89 students, 45 in grades 1, 2, and 3, and 44 in grades 5, 6, and 7, standard scores for students in grades 1–3 were consistent with past research in that the new test (Draw A Person), yielded lower scores than the older test (Goodenough-Harris). However, the scores for the older group indicated that the Draw A Person may yield somewhat higher scores than the Goodenough-Harris system at the upper ages. The results of the present study, which includes fourth and fifth graders, produced similar results in that the Draw A Person scoring system yielded higher scores than the Goodenough-Harris scoring system. Naglieri suggested that older students may have a greater opportunity to earn higher scores because of the way the Draw A Person system was developed, for example, the inclusion of bonus items. Credit for bonus items is given if all items for a specific criterion are scored
as correct (Naglieri, 1987). For example, in order to earn credit for the "Arms Bonus" item, the "Arms Present", "Arms Detail 1", "Arms Detail 2", and "Arms Proportion" items must all be scored as correct. That is, additional bonus credit is earned when no items are failed within a criterion. Bonus items were included to provide additional points at the upper end of the distribution of scores.

Criterion-related validity is demonstrated by the test's ability to differentiate between groups of people. The present ANOVA results indicated a significant difference between the three groups for each of the areas tested (Man, Woman, Self, and Total Test) for both the Draw A Person and the Goodenough-Harris scoring systems. This suggests differences between all three groups for each of the Man, Woman, Self, and Total Test comparisons. The standard scores for the normal sample were higher than those for the reading disabled, while the reading disabled subjects scored higher than the developmentally handicapped children. By definition, it is expected that the developmentally handicapped subjects would score lower than the other two groups, and the reading disabled sample would be expected to yield scores similar to those of the normal sample.

It is difficult to explain why the reading disabled sample did not score as high as the normal sample. One reason may be that the Draw A Person may involve specific cognitive components that are not currently included in intelligence tests such as the WISC-R and the Stanford-Binet. Das (1984) has suggested that current intelligence tests do not include two of the four components in his cognitive processing model. The two missing components are planning and arousal. If the Draw A Person includes either or both of these two components, it could effect the resultant scores by producing lower scores for the reading disabled sample. The fact that these components may be involved, and could potentially help identify reading disabled children, is of critical importance and requires further investigation.
Construct Validity

Naglieri's Draw A Person test was found to demonstrate good construct validity. The Draw A Person correlated significantly with a similar intelligence test (Goodenough-Harris) and achievement (K-TEA), thus providing evidence of the test's construct validity.

Practical Considerations

This study provides evidence of both construct and criterion-related validity of Naglieri's Draw A Person test. Although these statistical qualities are important to the test's usefulness, there are other critical factors that the practitioner should consider in determining whether to utilize the Draw A Person scoring system in place of the Goodenough-Harris scoring system.

Unlike the Goodenough-Harris scoring system, the Draw A Person combines all three of the Man, Woman, and Self standard scores into one Total Test Draw A Person score. This Total Test score provides for a score that is much more reliable than any of the individual scores. Therefore, this overall score is more advantageous than the individual drawing scores in that there is a marked increase in reliability.

The scoring time for the Goodenough-Harris system took approximately twice as long as the time required using Naglieri's system. Average scoring time for a drawing set (man, woman, and self) using the Goodenough-Harris system ranged from 15-20 minutes, whereas, using the Draw A Person system ranged from 8-10 minutes. The fact that Naglieri's test took approximately half the time of the Goodenough-Harris to score, is beneficial to practitioners.

Another advantage of the Draw A Person is its recent development which has resulted
in a more modern, updated test. The Goodenough-Harris system includes many outdated items, especially with regard to the clothing criteria for the women. Although, many of the criteria for dress and hairstyling were appropriate in the 1950’s, they no longer are appropriate to today’s styles and fashion. In contrast, Naglieri’s test is comprised of 14 general criteria areas that appear more resistant to change in fashion or culture. In fact, the criteria are general enough that the same 64 items are utilized for scoring of the man, woman, and self.

In addition, there are norm differences between the two tests. Harris’ test was standardized almost thirty years ago using information based on the 1950 U.S. Census. The Draw A Person standardization sample was representative of the 1980 U.S. Census and the data was collected in 1984, thereby providing more recently updated norms.

When considering an human figure drawing scoring system to utilize, it is important that the practitioner consider not only the validity and reliability of that system, but also those practical factors that are pertinent to the use of any new test.

Strengths and Limitations of the Study

A major strength of this study is the sample which is comprised of 120 students organized into three separate groups (40 normals, 40 reading disabled, and 40 developmentally handicapped). The Draw A Person test has been stated as being useful with special populations because of the test’s nonverbal and nonacademic content and ease of administration (Naglieri, 1987). Therefore, the use of special populations in this study provides particularly useful information. Additionally, a major drawback of prior Goodenough-Harris studies using special populations is that they did not include normal controls. Therefore, the inclusion of normal controls in this study provides for a better
designed research study.

A limitation of the study was the limited age range of the subjects making it difficult to generalize the results to other age groups. Although, the fact that the subjects ages ranged from 9-1 to 12-3 provided for a more homogeneous sample. The study was also restricted with regard to geographic location. The population was obtained in the Midwest which limits the generalizability of the results to the Midwest area.

Further Research Directions

This study looks at the construct and criterion-related validity of the Naglieri Draw A Person test using a sample of normal, reading disabled and developmentally handicapped children. There are various forms of validity that need yet to be investigated, both with normal and special populations.

The results of this study revealed, although insignificant, mean score differences between the Goodenough-Harris and the Draw A Person. The Goodenough-Harris mean scores were consistently lower than the Draw A Person mean scores. An item analysis of the two systems may provide insightful information into these mean score differences.

It was also determined that there were significant mean differences between the normal, reading disabled, and developmentally handicapped samples. By definition, reading disabled children do not have lower intelligence scores than those of a normal population. Therefore, research is needed investigating these differences between the three groups.

The studies described above would provide additional, pertinent information to the establishment of the Draw A Person as a useful instrument that provides an estimate of intelligence. However, because of the test's recent development and the varying populations with which it can be used, there is an unlimited amount of research yet to be done.
LIST OF REFERENCES


