THE DRAW-A-PERSON TEST AS A MEASURE OF SELF-ESTEEM
AND SELF-CONCEPT IN OUTPATIENT FEMALES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Julie A. Ganley, B.A., M.S.

* * * * *
The Ohio State University
1989

Dissertation Committee
James V. Wigtil, Ed.D.
Richard C. Kelsey, Ph.D.
James L. Collins, Ph.D.
W. Bruce Walsh, Ph.D.

Approved by
James V. Wigtil
Advisor
College of Education
ACKNOWLEDGEMENTS

While indeed one of the most solitary of all projects, the completion of a dissertation does require support and assistance. To those individuals who offered such aid, I sincerely express my appreciation.

I thank the members of my Committee who provided valuable comments and direction as well as support. I would especially like to express my appreciation to Dr. James Collins; his continuing support and encouragement made the completion of this project possible. I am also grateful to Dr. Richard Kelsey for his support and assistance. My appreciation also goes to the psychologists who acted as raters for the scoring system used in this study and the intern who assisted in compiling the data. Many thanks to the therapists who referred their clients as subjects, as well as to the community mental health center which permitted me to use their facility as a research site.

I am especially appreciative of family and friends who provided emotional support and encouragement. Many thanks to C.C. who typed and provided editorial assistance with much patience and endurance.

Lastly, a special thanks to S.P. who made all the difference.
VITA

February 25, 1952 . . . . Born - Appleton, Wisconsin

1974. . . . . . . . . . . . . . . B.A., Social Sciences  
Wright State University  
Dayton, Ohio

1978. . . . . . . . . . . . . . . M.S., Counseling  
Troy State University  
Troy, Alabama

1978-1983 . . . . . . . . Counselor, Outpatient Clinic  
Greene County Mental Health Center  
Xenia, Ohio

1983-1988 . . . . . . . . Psychology Intern, Outpatient  
Clinic, South Community Inc.  
Kettering, Ohio

1988-Present. . . . . . . Adjunct Faculty, Human Services  
Department, Antioch University  
Yellow Springs, Ohio

1988-Present. . . . . . . Adjunct Faculty, Counselor  
Education Department,  
University of Dayton,  
Dayton, Ohio

January, 1989-Present . . . Counselor, University of Dayton  
Counseling Center,  
Dayton, Ohio

FIELDS OF STUDY

Major Field: Department of Educational Services and Research
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ................................................. ii

VITA ................................................................. iii

LIST OF TABLES ...................................................... vii

CHAPTER  

<table>
<thead>
<tr>
<th>I. INTRODUCTION. .................................................. 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background of the Study ....................................... 1</td>
</tr>
<tr>
<td>Statement of the Problem ..................................... 3</td>
</tr>
<tr>
<td>Purpose of the Study .......................................... 8</td>
</tr>
<tr>
<td>Significance of the Study .................................... 10</td>
</tr>
<tr>
<td>Research Questions. ............................................ 12</td>
</tr>
<tr>
<td>Definitions of Terms. ......................................... 12</td>
</tr>
<tr>
<td>Assumptions .................................................... 14</td>
</tr>
<tr>
<td>Limitations ..................................................... 15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. REVIEW OF LITERATURE. ...................................... 17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projective Tests ................................................ 17</td>
</tr>
<tr>
<td>Human Figure Drawings ......................................... 21</td>
</tr>
<tr>
<td>Projective use of drawing tests ................................ 23</td>
</tr>
<tr>
<td>Reliability and validity of HFD's .............................. 28</td>
</tr>
<tr>
<td>HFD's as expression of self ..................................... 30</td>
</tr>
<tr>
<td>Self-Esteem And Self-Concept. .................................. 33</td>
</tr>
<tr>
<td>Gender Differences in Self-Esteem and Self-Concept .......... 41</td>
</tr>
<tr>
<td>Measurement of Self-Esteem and Self-Concept ................. 45</td>
</tr>
<tr>
<td>Female Adult Development ....................................... 48</td>
</tr>
<tr>
<td>Summary .......................................................... 53</td>
</tr>
</tbody>
</table>

iv
<table>
<thead>
<tr>
<th>APPENDICES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Memo to Clinicians</td>
<td>142</td>
</tr>
<tr>
<td>B. Script to Subjects</td>
<td>144</td>
</tr>
<tr>
<td>C. Subject Consent Form</td>
<td>146</td>
</tr>
<tr>
<td>D. Human Subjects Review Approval</td>
<td>148</td>
</tr>
<tr>
<td>E. Rosenberg Self-Esteem Scale.</td>
<td>150</td>
</tr>
<tr>
<td>F. Instrument Administration Procedures By Instrument</td>
<td>152</td>
</tr>
<tr>
<td>G. Criterion Summary Sheet.</td>
<td>155</td>
</tr>
<tr>
<td>H. Instructions to Raters</td>
<td>157</td>
</tr>
<tr>
<td>I. DAP Summary Sheet</td>
<td>161</td>
</tr>
<tr>
<td>J. Scoring System</td>
<td>163</td>
</tr>
<tr>
<td>K. Sample Protocols</td>
<td>170</td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>188</td>
</tr>
</tbody>
</table>

vi
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>DEVELOPMENTAL TASKS, PSYCHOSOCIAL CRIZES AND PROCESSES FOR RESOLUTION OF CONFLICT FOR EARLY AND MIDDLE ADULTHOOD</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pilot Study and Research Study Demographic Data.</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>Pilot Study Criterion Score Means, Standard Deviations, Ranges</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>Demographic Distribution of Pilot Study Sample By Frequency and Percent</td>
<td>76</td>
</tr>
<tr>
<td>4</td>
<td>Demographic Distribution of Research Study Sample By Frequency and Percent</td>
<td>77</td>
</tr>
<tr>
<td>5</td>
<td>Demographic Data of Pilot Sample By Criterion and Level</td>
<td>78</td>
</tr>
<tr>
<td>6</td>
<td>Demographic Data of Research Study Sample By Criterion Group and Level</td>
<td>79</td>
</tr>
<tr>
<td>7</td>
<td>Biserial Correlation of DAP Self-Concept Drawing Variable Scores and Pearson's Product Moment Correlation of DAP Self-Concept Composite Scores with Criterion Measure (Miskimins SGODS) by Judge</td>
<td>86</td>
</tr>
<tr>
<td>8</td>
<td>Biserial Correlation of DAP Self-Esteem Drawing Variables and Pearson's Product Moment Correlation of DAP Self-Esteem Composite Score With Criterion Measure (Rosenberg SEI) by Judge</td>
<td>87</td>
</tr>
<tr>
<td>9</td>
<td>Criterion Measure Score Ranges, Means, Standard Deviations</td>
<td>89</td>
</tr>
</tbody>
</table>
TABLE | PAGE
--- | ---
11. Biserial Correlation of DAP Self-Concept Drawing Variable Scores and Pearson's Product Moment Correlation of DAP Self-Concept Composite Score With Low Criterion Group Scores by Judge | 89
12. Biserial Correlation of DAP Self-Concept Drawing Variable Scores and Pearson's Product Moment Correlation of DAP Self-Concept Composite Scores with Medium Criterion Group by Judge | 90
13. Biserial Correlation of DAP Self-Concept Drawing Variable Scores and Pearson's Product Moment Correlation of DAP Self-Concept Composite Scores with High Criterion Group by Judge | 90
16. Biserial Correlation of DAP Self-Esteem Drawing Variable Scores and DAP Self-Esteem Composite Scores with High Criterion Group Scores By Judge | 93
17. DAP Self-Concept Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Size Variable | 95
18. DAP Self-Concept Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Placement Variable | 95
19. DAP Self-Concept Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Gender Identification Variable | 96
20. DAP Self-Concept Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Affect Variable | 96
<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>DAP Self-Concept Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Age Variable</td>
</tr>
<tr>
<td>22</td>
<td>DAP Self-Concept Inter-Rater Reliability (Pearson's Product Moment Correlation) By Judge, DAP Composite Score</td>
</tr>
<tr>
<td>23</td>
<td>DAP Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Size Variable</td>
</tr>
<tr>
<td>24</td>
<td>DAP Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Quality Variable</td>
</tr>
<tr>
<td>25</td>
<td>DAP Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Gender Identification Variable</td>
</tr>
<tr>
<td>26</td>
<td>DAP Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Placement Variable</td>
</tr>
<tr>
<td>27</td>
<td>DAP Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) By Judge, DAP Omissions Variable</td>
</tr>
<tr>
<td>28</td>
<td>DAP Self-Esteem Inter-Rater Reliability (Pearson's Product Moment Correlation By Judge, DAP Composite Score</td>
</tr>
<tr>
<td>29</td>
<td>Comparison of DAP Self-Concept and Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) By Scoring Variable</td>
</tr>
<tr>
<td>30</td>
<td>DAP Self-Concept Inter-Rater Reliability (Tetrachoric Correlation) For Low Self-Concept Criterion Group By DAP Variable and Judge</td>
</tr>
<tr>
<td>31</td>
<td>DAP Self-Concept Inter-Rater Reliability (Tetrachoric Correlation) For Medium Self-Concept Criterion Group By DAP Variable and Judge</td>
</tr>
<tr>
<td>32</td>
<td>DAP Self-Concept Inter-Rater Reliability (Tetrachoric Correlation) For High Self-Concept Criterion Group By DAP Variable and Judge</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>33</td>
<td>DAP Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) For Low Self-Esteem Criterion Group By DAP Variable and Judge.</td>
</tr>
<tr>
<td>34</td>
<td>DAP Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) For Medium Self-Esteem Criterion Group By DAP Variable and Judge.</td>
</tr>
<tr>
<td>35</td>
<td>DAP Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) For High Self-Esteem Criterion Group By DAP Variable and Judge.</td>
</tr>
<tr>
<td>36</td>
<td>DAP Self-Concept Inter-Rater Reliability By Three Levels of Criterion Group and Variable.</td>
</tr>
<tr>
<td>37</td>
<td>DAP Self-Esteem Inter-Rater Reliability By Three Levels of Criterion Group and Variable.</td>
</tr>
<tr>
<td>38</td>
<td>Fisher's Z(r) Transformation Test For Differences Among DAP Intec-Rater Reliability Correlation Means For DAP Composite Scores</td>
</tr>
<tr>
<td>39</td>
<td>Self-Esteem Scoring Procedure By Variable</td>
</tr>
<tr>
<td>40</td>
<td>Self-Concept Scoring Procedure By Variable</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Background of the Study

It has long been held that all art is symbolically reflective of the artist's personality, lifestyle and general approach to the world. Leonardo Da Vinci, centuries ago, noted that people have a strong tendency to impose their own ideas and interpretations upon unstructured stimuli. Da Vinci stated in his Introduction to the Painter that he could make associative connections as a result of seeing a blot made by a sponge upon the wall "...various experiences can be seen in such a blot, providing one wants to find them in it -- human heads, various animals, battles, cliffs, seas, clouds or forests and other things..." (Zubin, et al, 1965, p.167).

The projective aspects of painting and drawing have been explored for centuries from an artistic angle; a more formal, psychological interest in drawings as a means of gathering information about the personality, emotional status and intellectual abilities of the artist has been developing for nearly 100 years. The clinical use of drawings, like most psychological tests at the turn of the century, primarily focused upon measuring intelligence and scholastic achievement.
A major problem existed in that the field of psychology had developed a very limited concept or theory of personality. It was not until the early twentieth century, when Freud developed his notions about personality functioning via psychoanalysis that projective methods of assessment were able to flourish based on a conceptual framework (Rabin, 1968).

At the same time that Rorschach (1921) was working with inkblots and Murray (1938) with pictures as projective devices, Schilder (1935) suggested that an individual's self-concept was represented or projected in body image, and that the body image drawn by the individual is a configuration, or gestalt, composed of many physiological and psychological sensations experienced within one's body. If the human figure drawing could be considered the self-image, consciously or unconsciously projected, the analysis of drawings would have great significance. Distortions in the drawing could be symbolic representation of inadequacies or distortions in the artist's self-image. Such was the theoretical assumptions of the first theorists to utilize the human figure drawing.

Goodenough (1926), who initially used children's drawing tests as a way to assess intellectual development, later came to see that the drawings could be used interpretatively in the study of personality. Machover (1949) hypothesized that the self-image is projected into the drawing of the human figure. Based upon this intriguing theoretical base, rather than an empirical one,
Machover developed the **Draw-A-Person Test** (DAP) in which the client is asked to draw three separate drawings: "Draw a person"; "draw a person of the opposite sex"; and "draw the self".

The DAP developed by Machover to assess a person's self-image has been widely used in clinical practice. In fact, the DAP has been one of the most widely used tests in the United States. In 1971, Lubin, et al (1971) found that the DAP was the fourth most commonly used testing instrument among clinicians in over 250 clinical settings. Additionally, 79% of these clinicians reported that they used the DAP on a regular basis. The most recent estimate (Piotroski, 1985) ranked the DAP as the sixth most frequently used psychological test in the United States.

**Statement of the Problem**

Several problems exist in human figure drawing practice and research. First, despite the DAP's popularity in clinical use, it ranks much lower in terms of the most frequently researched tests. Buros (1978) indicates that the Goodenough and Harris Draw-A-Person is only the 29th most frequently researched assessment instrument, while the Machover DAP ranks 40th in the area of complete research.

Second, in addition to the paucity of research relative to its usage, the DAP has been the subject of constant controversy. At best the literature can be summarized as inconsistent and in general not supportive of the validity of the DAP (Swensen, 1968; Roback, 1968; Kahill, 1984). Kahill (1984) found the DAP
validation literature mixed, with slightly more negative than positive findings. She stated: "While it is obvious that figure drawings are not meaningless, establishing what it is they mean with any precision or predictability is difficult" (p.288). Many authors (Swensen, 1968; Roback, 1968; Kahill, 1984; Handler, 1985) allowed that the inconsistency of the DAP literature may be due in part to the methodological flaws typically found in the DAP research, specifically problems with internal validity and design selection.

Sims et al (1983) pointed out that the construct validation of the DAP as a measure of self-esteem and self-concept has been hampered by serious problems in design selection, in which investigators avoid the construct validation design paradigm even though it is most methodologically appropriate. For example, studies are more of a "hunting expedition" in which a long list of drawing variables are examined in order to determine which ones might correlate with specific factors such as outside measures of anxiety or self-esteem. Little attention seems to be devoted to developing a viable set of hypotheses designed to focus on the constructs underlying the test. In essence there is a total focus upon the DAP test and or drawing variables rather than on any set of related hypotheses which underlie its interpretation, resulting in studies with a very diffuse theoretical base.
Along with a failure to select an appropriate research paradigm, DAP researchers do not in general appear to take adequate steps to ensure internal validity. For example, specifically problematic in the DAP literature is the failure to control for the influence of confounding variables such as age, sex and race, all of which have been found to have a strong influence upon drawing performance. In general, there appear to be considerable differences in the drawings of adults and children, men and women, and blacks and whites (Kahill, 1984). This study's subjects consisted of white, non-pregnant females, clients in an outpatient mental health clinic.

A second internal validity problem present is the failure to adequately maximize experimental variance by failure to make criterion categories as different as possible. For example, when looking at low and high self-esteem groups and their differential performance on the DAP, these groups must be as unlike as possible in order for differences in the relationship between the variable and DAP performance to have a chance to show itself, so to speak, from the total variance (Kerlinger, 1986). However many of the DAP studies tend to use the mean of criterion scores as the cutoff point between high and low scores resulting in the lack of adequate experimental variance.

A third problem that has been identified in the DAP literature is poor definition or loose operationalization of the criterion (Swenson, 1968; Kahill, 1984). This is especially a
difficulty in the studies designed to examine the DAP as a measure of self-esteem and/or self-concept. The terms "self-esteem" and "self-concept" are in general poorly defined with little reflection of current literature. For example, the two terms are often used interchangeability with no clear distinction between self-esteem and self-concept, when in fact the self-esteem/self-concept literature defines the two terms quite differently. This definition problem results in obscurity of diagnostic labels which has caused inconsistencies in the literature. For example when sorting subjects into high or low categories there appears to be little confidence that those categories accurately reflect the constructs.

Additional problems in the DAP validation research involve failure to include a clearly detailed scoring system such that the meaning of an item is evident. Also, when scoring systems are typically delineated, they tend to be so complex as to prohibit practical clinical use (Handler, 1985). In addition, many Draw-A-Person studies have involved only a very abbreviated version of the Machover original DAP, for example requesting only one drawing rather than three. This practice may result in a failure to gather sufficient data upon which to make sound interpretations.

In part, the contradictory DAP evidence is a result of methodological problems in the research outlined above as well as a reflection of the complexity of the concept of body image.
Hammer (1969) has criticized most researchers for narrowly defining the projection of self, as the person drawn by the client may involve one's actual self or one's idealized self. Indeed, few researchers have incorporated this notion into their research efforts. In addition, no single study has examined both the constructs of self-concept and self-esteem. Another problem with the current DAP research is one of external validity as most of the subjects consist of an inpatient adult psychotic, child or student population. Few studies have been based on a clinical adult outpatient population, the population which is mostly typical in DAP usage.

Several authors (Sims et al 1983; Handler 1985) suggest that future research should examine hypotheses designed to explain drawing performance, the use of clinical judgment in the scoring procedure, and the use of careful construct validation procedures. Both Cronbach (1982) and Anastasi (1982) encourage additional projective validation research to investigate the construct validity of projective instruments by testing specific hypotheses that underlie the use and interpretation of each test.

Wanderer (1969) indicates that it is a major responsibility of clinicians to insure that the instruments they use actually do what they purport to do. In terms of Machover's theory that the human figure drawn by the client is a projection of the self-image, the empirical evidence is mixed, with slightly more findings that fail to support the hypothesis than support it
(Kahill, 1984). However, the literature appears to be fraught with serious methodological flaws which contribute to the conflicting findings. With such an array of contradictory research on the use of human figure drawings as a psychodiagnostic tool, as well as its continued popularity, it is evident that research efforts should be directed towards refining projective drawings as a clinical instrument. Thus, the decision to study the DAP as a measure of self-esteem and self-concept is based upon its widespread use; contradictory conclusions reported in the literature reviews; and failure of the existing body of research to adequately validate Machover's hypothesis that drawings are a reflection of the self.

**Purpose of the Study**

This study was designed to ascertain if self-esteem and self-concept account for variations in drawing performance and do so in a methodologically correct manner avoiding many of the methodological flaws in past research and utilizes a more appropriate design than most previous studies. In addition, it was hoped that the results would be more generalizable to female client populations.

Examining the question of whether, and in what manner, figure drawings express self-esteem and self-concept in outpatient females, a construct validation paradigm was utilized which is methodologically appropriate though seldom used. Construct validation is both a logical (qualitative) and
empirical (quantitative) process, beginning with the formulation of hypotheses about the differential drawing variables of individuals who fall into each of the high or low groups based on the criterion measures. These hypotheses were then used to formulate a scoring system which was empirically tested.

Because reliability of any instrument is an important determination of validity, inter-rater reliability of the DAP scoring system was determined. The validity of the scoring system was determined by its correlation with the self-esteem and self-concept criterion measures, as well as examination of the scoring system's ability to discriminate between various levels of the criterion measures.

This construct validation study is similar to Ex Post Facto/Correlational research in that often a major problem with this type of design is the threat to internal validity by selection of subjects. Because subjects could not be randomly selected in the present study, there exists the possibility that they brought to the study certain traits which confounded drawing performance. Two major strategies often used to control for this threat are to hold these potential threats constant or to build the variables into the research design. Both these strategies were employed in the present study.

Age, gender, race and pregnancy status were held constant and varying levels of self-esteem and self-concept were deliberately incorporated into the study. Because sex
differences exist in drawing behavior (Delatte, 1985; Saarni and Azara, 1977; Handler, 1985; Viney, 1974), sex was held constant. Women were chosen for several reasons. First, females more than males tend to perceive a meaningful relationship between their bodies and life roles, and between their bodies and self-concept (Fisher, 1968). Secondly, there is evidence to suggest that the female drawings of women do indeed represent their self-image (Apfeldorf and Smith, 1966). Only white, non-pregnant women were chosen to participate in the study as the literature suggests that pregnant women tend to represent their body image differently than do non-pregnant women (Viney et al, 1974), and that race has been shown to account for differential drawing performance. It appears that blacks do not project their self-image onto human figure drawings with the same frequency as do whites (Dennis, 1968; Kuhlman, 1979).

**Significance of the Study**

The question of validating the DAP as a measure of self-esteem and self-concept is important not only from a theoretical standpoint but a practical one as well. It was hoped that this study would successfully find additional and improved ways to measure self-esteem and self-concept in women via the DAP. Since the DAP is one of the most frequently used test instruments in the United States, validating it as a measure of self-concept and self-esteem would greatly expand its utility as a measurement instrument with no increased testing costs or time.
While the significance of measuring self-esteem and self-concept may seem evident, a brief discussion along these lines is provided below.

Some authors have maintained that the self-concept is a very sensitive reflector of mental health on both ends of the continuum -- mental illness and mental health (Rogers, 1951; Combs and Snygg, 1959; Pitts, 1972). The way that people evaluate themselves is widely accepted as important by mental health and medical professionals. The expected link between self esteem and self concept and mental illness has been strongly supported in the literature (Shavelson and Bolus, 1982). Clinicians interested in individual clients want to identify those with poor self-esteem and poor self-concept to better understand causes and symptoms in order to develop effective treatment plans.

Traditional measures of self-esteem and self-concept are of the objective self-report type, and as such are limited due to social desirability of items and subject defensiveness (Perkins and Shannon, 1965). It is speculated that this may be more the case with women, and that the DAP, because it is an unobtrusive measure, would be a valuable addition to existing ways of measuring self-concept and self-esteem.

The scoring system developed by this researcher differed substantially from other DAP scoring systems in several ways. First, most other scoring systems have been developed utilizing
student or mentally retarded subjects which greatly limits
generalizability to normal outpatient populations, the target
population of this study. Second, most scoring systems tend to
be extremely complex, prohibiting practical clinical usage. The
scoring system developed for this project is one that can be used
by the counselor or therapist in clinical practice. Third, no
other study to date has simultaneously examined both self-esteem
and self-concept as this study has done. Last, this study was
the first study to examine the relationship between drawings
rather than one drawing in isolation.

**Research Questions**

Four research questions were addressed in this study. They
are as follows:

1. Is there a relationship between the DAP scoring system
and the criterion measures of self-esteem and self-concept?

2. Can the DAP scoring system discriminate between various
levels of the criterion measures?

3. Will the four judges, independently utilizing the
author's rating system, demonstrate agreement?

4. Will judges using the DAP scoring system exhibit
differential scoring agreement based upon level of criterion
measure?

**Definitions of Terms**

Six terms have been defined for the purpose of this study.
A definition of each term is provided below.
1. **Criterion**: A direct and independent measure of that which an instrument is designed to measure. The criterion measures for this study are the Rosenberg Self-Esteem Inventory (SEI) and the Miskimins Self-Goal-Other Discrepancy Scale (MSGO).

2. **Ideal Self**: The concept which persons hold regarding the self as they would like to be, as measured by the MSGO.

3. **Projective Test**: A method of personality assessment in which the client is presented with a relatively unstructured task. The way in which the client interprets or structures the vague or ambiguous testing tasks are seen as a reflection of his or her psychological functioning.

4. **Real Self**: The concept which a person holds of the self as it actually is, as measured by the MSGO.

5. **Self-Concept**: The perception one holds of oneself as a total person in general, and more specifically the individual's picture of real self and ideal self as measured by the MSGO. An individual is said to have a good self-concept when she/he exhibits high real and ideal self congruency. Conversely, a poor self-concept is reflected in an individual who experiences a large discrepancy between real and ideal self.

   a. **Good Self-Concept**: Scores falling one standard deviation above the mean of the MSGO obtained in the pilot study.
b. Poor Self-Concept: Scores falling one standard deviation below the mean of the MSCE obtained in the pilot study.

6. Self-Esteem: Measure of personal worth or value in a global sense, as measured by the SEI.
   a. High Self-Esteem: Scores falling one standard deviation above the mean of the SEI as obtained in the pilot study.
   b. Low Self-Esteem: Scores falling below one standard deviation of the mean of the SEI as obtained in the pilot study.

Assumptions

Several basic assumptions held by the researcher about the DAP were incorporated into this study as part of the hypotheses for drawing variance, thus must be made explicit. These assumptions are adapted from Leland (1984), and are stated below.

1. It is assumed that every aspect of behavior is significant. All client behaviors offer important information to the clinician, including drawing behavior.

2. When the client is asked to draw three separate drawings of a person in the DAP, it is assumed that the first drawing of the DAP represents the drawer's "ideal self"; the second drawing represents the drawer's relationship to the opposite sex; the last drawing represents the drawer's "real self".
3. Lastly, and importantly, rather than examine each drawing in isolation it is necessary to examine the relationships between the figure drawings.

Limitations

Several limitations exist in this study due to the non-experimental nature of this research. First, because direct control of the independent variable is not possible, and neither experimental manipulation nor random assignment can be employed, statements about cause and effect relationships cannot be made.

Second, central to any study utilizing criterion measures are the instruments utilized to measure the criterion variables. Valid and reliable instruments (the Rosenberg SEI and Miskimin's SGODS Scale) were selected which seemed to accurately and adequately reflect the research constructs. However, in spite of careful selection of criterion instruments, it is difficult to ascertain how factorially pure an instrument is and if it measures only that which it is purported to measure.

Third, attempts to strengthen the internal validity of this study unfortunately may have had an impact upon external validity. For example, in this study the ability to generalize to a normally distributed population may be limited due to the selection of subjects. White, non-pregnant female volunteers between the ages of 20-45 in an outpatient setting were utilized, the subjects selected as to their particular possession of various levels of self-esteem and self-concept. Therefore,
results cannot be generalized to males, to women in other age or culturally different groups, to pregnant women or to non-volunteers.

Fourth, research questions two and four examined subgroups of the sample which resulted in N's of less than 20. Small sample size tends to reduce the possibility of finding significant results if they exist.

Last, the criterion subgroups were homogenous groups. This restriction of range tends to produce lower correlations.
CHAPTER II

REVIEW OF LITERATURE

The purpose of this chapter is to present a selective and representative review of the literature most relevant to the present research. Literature on the development and growth of projective tests is reviewed with a specific focus upon human figure drawings. A review of the research relating to self-esteem and self-concept is presented, with special emphasis upon gender differences, and measurement. Last, a section which contains a discussion of adult female development is included followed by a chapter summary.

Projective Tests

Toward the end of the 19th century psychologists began to systematically experiment with the use of unstructured visual stimuli to induce a wide variety of associations and responses via the use of inkblots and pictures. For example, in 1895 Binet and Henri began using inkblots in the investigation of visual imagination, and Binet and Simon in 1905 used pictures to elicit verbal responses in order to assess intellectual development.

The use of these instruments was quite limited in scope in that the focus was primarily upon notions such as imagination,
mental content and individual differences, rather than personality assessment. This occurrence was due in large part to the limited personality theory at that time. Personality simply was not part of the psychology as of yet. It was not until Freud's development of psychoanalysis as a formal personality theory that the various projective methods were given a conceptual base (Zubin, 1965).

Freud (1911) considered projection the main mechanism underlying paranoid disorders, and it was seen as "the tendency on the part of the patient to externalize unacceptable inner drives and other undesirable internal proclivities to the outer world" (Rabin, 1968, p.8.). Later, Freud extended and broadened the meaning of projection which became an extremely significant development in the understanding of projective methods:

...it makes its appearance not only in paranoia, but under certain other psychological conditions as well; in fact it has a regular share assigned to it in our attitude to the external world. For when we refer the causes of certain sensations to the external world, instead of looking for them (as we do in the case of others) inside ourselves this normal proceeding, too, deserves to be called projection (Freud, 1949, p.452).

The narrowly defined term projection was beginning to be widened to mean not only paranoid projection but also the normal process that all individuals utilize in externalizing feelings, thoughts and attitudes to others and the outside world. Freud's expanded definitions fostered a new era of psychology in the 1920's and the true beginnings of projective technique development, primarily by Rorschach and Murray.
In contrast to prior work conducted with inkblots which consisted mainly of studies of imagination and association, Rorschach, strongly influenced by the works of Freud and Jung, emphasized the formal characteristics and determinants of the responses, perceptual modes, and their relationship to personality and psychopathology, first publishing his results in the *Psychodiagnosti* in 1921. In 1935 Morgan and Murray published the Thematic Apperception Test (TAT). In contrast to the Rorschach which was primarily empirically based, the TAT was strongly based on a theoretical framework influenced markedly by psychodynamics. Murray emphasized a dynamic approach to personality and was concerned with the drives, urges, as well as needs of the individual.

Several years later in 1938 Murray wrote his well known *Explorations in Personality*, a seminal work in which Murray first introduced the concept of Projection Tests. He described the methods in the following way: "...in an attempt to discover the covert (inhibited) and unconscious (partial repressed) tendencies of normal persons...procedures are developed which stimulate the imaginative process and facilitate their expression in words or action" (p.8).

At approximately the same time Murray was discussing projection tests, Frank also used the term "projective methods" (Frank, 1939). He described projective techniques as a "method of studying the personality by confronting the subject with a
situation to which he will respond according to what the situation means to him and how he feels when responding" (p.46).

As Rabin (1968) pointed out, understanding the process of projection is necessary but not sufficient for understanding what is meant by projective tests. In a very broad sense an individual projects all the time whenever perceiving and responding to the environment with personal needs, motivations, and unique tendencies. A projective technique, however, provides an individual with a specific stimulus which is responded to in terms of the meaning the particular stimulus has for that individual.

Lindzey (1961) viewed projective tests as instruments that are especially sensitive to covert aspects of behavior, permit a wide variety of subject responses, are multidimensional, and evoke rich and profuse response data with the subject having minimal awareness concerning the purpose of the test. In addition he stated that in projective testing the stimulus material is ambiguous, interpretation is based upon holistic analysis, and there are no right or wrong answers to the test.

Initially, projective tests were hailed as "psychological x-rays", that is, the only instruments that could enter a person's inner world and obtain information that could not otherwise be revealed by the person through other methods (Murstein, 1965). Many feel that the projective test has not lived up to these high expectations; and the projective test has
been strongly criticized for a variety of reasons. Often cited are inadequate administration and scoring standardization; lack of normative data; failure to establish scorer reliability; and importantly, failure to establish validity (Anastasi, 1982).

**Human Figure Drawings**

It seems self-evident that people have a need to communicate a statement about themselves as occupying space through a pictorial representation of the body. Persons from the beginning of time have expressed themselves through the use of pictures. Even early man seems to have been aware of the symbolic nature of drawings. Jaffe (1964) described a Paleolithic cave painting: "...a dancing human being, with antlers, a horse's head, and bear's paws...unquestionably the "Lord of the Animals" (p.235). This painting seems to symbolize the drawer's control over the animals—by representing all their strengths. Indeed, it is from the outgrowth of these pictures that the first crude symbols were devised which eventually became a written language.

One of the earliest formal uses of drawings as clinical instruments began when Simon (1876), in his observations of art drawn by severely disturbed individuals. He was able to observe relationships between features of the drawings and the symptoms of the patients. Like the other projective methods, however, it was not until the early 20th century and the development of personality theory that human figure drawing as a personality test began to flourish. During the modern era drawings were
first used to assess intellectual development. Following the
trend to create objective and standardized tests, detailed
scoring systems were developed to obtain I.Q.'s based on a
person's human figure drawing.

Burt (1921) found that while the sequence of drawing is
quite stable from child to child, the rate of development often
varies with the intellectual abilities of the child. Children
with lower intelligence tend to move through the stages rather
slowly and development stops early, while more intelligent,
gifted children go through the stage comparatively quickly and
continue their development longer.

It was these kinds of observations which led to the idea of
using drawings as a way of assessing children's intellectual
functioning. The most important contributor to this area was
Goodenough (1926) who developed the Draw-A-Man Test, the first
published drawing test to assess a child's intelligence. In her
book *Measurement of Intelligence by Drawings*, she standardized
and validated the Draw-A-Man Test and provided a detailed scoring
system for it. The child was given a blank piece of paper and
pencil and simply told to "draw a person". The child's I.Q. was
determined by several factors: 1) the number of details
included, such as facial features, hands, fingers, 2) correct
proportions between the various body parts, 3) motoric
coordination such as the fluency of the lines and the integration
of the parts. Each drawing was evaluated in regard to 51 items,
with the sum of scores forming a mental age, then an actual I.Q.
Goodenough (1926) described her Draw-A-Man Test as quite a useful
test for determining intellectual development for both the
purpose of making comparisons between groups as well as
supplemental to the usual I.Q. test for individual cases.

Because Goodenough was quite thorough in her presentation,
the test was used for more than 40 years without any attempt at
revision. It was not until 1963 that Harris revised the Draw-A-
Man in the book titled *Childrens Drawings As Measures of
Intellectual Maturity*. He updated the 1926 norms, increased the
number of criteria from 51 to 73 to improve reliability and
validity, and improved the administration of the test.

The Goodenough Harris Draw-A-Man Test has high reliability
and validity. Test-retest reliability is reported as high as
.89, scorer reliability .90 and correlations with other measures
of intelligence are typically over .50.

*Projective Use of Drawing Tests*

In her work with human figure drawing, Goodenough came to
believe that children are influenced by past experiences when
drawing, as they draw more than they actually see. She even
suggested that because drawings are a subjective process,
"drawings made by children might furnish considerable aid in the
early diagnosis of personality disorders and mental adjustments"
(Goodenough, 1926, p.8). In addition to her findings that
drawings could measure intelligence, Goodenough observed
qualitative differences as well as quantitative ones. She classified these differences into four types: 1) The "verbalist" type: drawings with many details, but few ideas 2) The "individual response" type: drawings with traits which are understandable only to the drawer 3) "Flight of ideas" type as when hair is drawn only on one side of the head, or only one ear is drawn 4) "Uneven mental development" type where the drawing consists of both mature and immature characteristics (Goodenough, 1926). Goodenough characterized children with the above drawing styles as exhibiting psychopathology. However, her sample was extremely small, preventing any conclusive generalizations.

Two of the first and most well-known investigators of drawing tests as personality measures were Buck (1948) and Machover (1949). They were the first to provide a thorough theoretical rationale for the projective qualities of human figure drawings. Of the two, Machover's Draw-A-Person Test (DAP) has been the most widely used and has inspired more research, in part due to the provocative research hypotheses that she set forth. Machover's work, strongly based in psychoanalytic theory, posited that when an individual draws herself she is guided by the image she has of herself (the body) and that a drawing is produced that involves both conscious and unconscious attitudes and perceptions. Asking a person to draw a human figure would force the individual to draw a representation of innermost
thoughts, feelings, needs and drives, reflecting significant personality variables and personality style.

Machover's method involves an examination of the drawing (how each part of the body and clothing is drawn); the structural and formal aspects of the drawing (such as size, placement, type of line, shading, symmetry, erasures); overall mood (facial expression and posture); and conflict indicators (differential treatment by omissions, breaks, dimming or reinforcing, erasures or shading). While Machover based her work upon a strong theoretical base and innumerable clinical cases, she supported many of her bold generalizations by case notes, rather than any sound empirical foundation (Zubin, 1965; Rabin, 1968).

Koppitz (1968), dissatisfied with this lack of a data base in Machover's work, attempted to integrate clinical insight with the traditional scientific method. Koppitz disagreed with Machover's psychoanalytic theoretical framework, and instead utilized Sullivan's (1953) Interpersonal Relationship theory. She emphasized ego-psychology and conscious processes, and the drawer's attitude toward self and significant others, rather than unconscious needs, defense mechanisms and psychosexual development. Koppitz rejected the "body image" hypothesis in her work with children. Instead, she viewed children's drawings as a reflection of transient attitudes and emotional indicators, not permanent personality traits. Koppitz's system integrated the
developmental and projective views of drawing tests measuring both mental and socio-emotional maturity.

Koppitz was able to identify 30 specific traits in drawings called emotional indicators which were indicative of emotional problems. For example she found that depressed children had the tendency to draw tiny figures, with no mouth or eyes, while aggressive children drew overly large figures, large ears and hands, with genitals and transparencies (Koppitz, 1966).

The work done by Goodenough, Machover, Buck and Koppitz stimulated a steady flow of research and the development of variations on the original drawing tests. Some of these variations include the Levy Animal-Drawing-Story Technique (LAOS) (Levy and Levy, 1958); the Drawing Completion Test (Kinget, 1952); Draw-A-Person-In-The-Rain (Hammer, 1958); and the popular Kinetic Family Drawing (KFD) (Burns and Kaufman, 1970). Despite the questionable validity of drawing tests, it appears that clinicians utilize the procedure frequently and consider the instruments valuable in clinical diagnosis.

Relatively few studies exist which examine the quality and nature of typical drawings done by adults; most normative studies examine at length the typical drawings done by children. However, the following is a brief summary of the findings pertaining to adult drawings. The typical drawing by an adult consists of a head which is drawn first, facial features (eyes, nose, mouth, ears, hair), legs, feet, arms, hands, fingers, neck,
shoulders and trunk. Also included are such additional details such as a belt and clothing of some type (skirt, dress, trousers, shirt, jacket). The drawing is placed approximately in the middle of the page and is about six to seven inches in size (Handler, 1985). Head, trunk, arms and legs are typically in proportion, with relative symmetry, spontaneity and movement. Line quality is usually consistent. The head tends to be more oval than round, and some attempt is made to draw the facial features in a realistic manner (e.g., eyes are almond shaped, not circles or dots). The body is life-like, with a three dimensional quality. Secondary sexual characteristics are included, with clear distinction between male and female drawings (Jones and Thomas, 1964; Thomas, 1966; Wagner and Schubert, 1955).

The literature supports the existence of gender differences in drawings. Delatte (1985) found that young women with high self-esteem tend to draw more feminine drawings, while men with high self-esteem tend to draw larger figures. He speculated that size may symbolize strength and worth to males, while females may equate attractiveness or femininity with worth.

Saarni and Azara (1977) found that relative to females, males revealed significantly more aggressive-hostile indices (e.g., scars, crossed eyes, gross asymmetry of limbs, genitals present), while females tended to draw more ambiguous, childlike figures when compared to the male DAP. Most normals draw the
same sex first (estimates range from 85 to 95 percent), however this is not as true for women as for men (Handler, 1985). Viney et al. (1974) found that pregnancy status altered women's drawings. They found that pregnant women tended to draw the female figures larger than other women did and that their female figures were considerably larger than the male figures. Pregnant women also tended to draw more distorted figures when compared to other women with gynecological concerns, and showed more signs indicating preoccupation with their bodies. The above data should be evaluated with some caution as an extensive set of age related norms for adults does not exist, such as exists for children.

Reliability and Validity of Human Figure Drawing Tests

The research in the area of projective drawings is contradictory and therefore it is quite difficult to make conclusive comments. One is directed to the complete reviews by Kahill (1984), Molish (1972), Hammer (1969), Roback (1968), and Swenson (1957, 1968). Swensen (1957), after an exhaustive review of the DAP literature between 1945 and 1956, found that Machover's (1948) hypotheses concerning human figure drawings were seldom supported by the literature. Suinn and Oskamp (1969), looking at a smaller sample of studies, also found that the validity of drawing tests were highly tenuous. Swensen and Roback both completed other reviews of the DAP literature in 1968 and found a significant increase in the empirical data supporting the DAP as
a clinical tool but did not see the DAP as able to provide an improvement in clinicians' judgment accuracy. Klopfer and Taulbee (1976), in a three-year survey beginning in 1971, remarked that "it is unfortunate that so much research effort is being expended upon the DAP in view of the generally unencouraging results" (p.1) and recommended that figure drawings be used only as graphic signs that are meaningful only as discussed with the client in the context of other information.

In the most recent comprehensive review of the DAP literature (Kahill, 1984), the reliability of human figure drawing tests appears to have improved considerably. For example, DAP test-retest reliabilities appear acceptable in that they range from .81-.91. Inter-rater reliability of DAP has considerably improved from Swensen's review in 1968. Kahill (1984) found that most of the studies she reviewed yielded correlations over .70, which she attributed to increased attempts to objectify and standardize rating procedures and to adequately train judges.

In terms of validity, Machover's hypotheses regarding both content and structural aspects of drawing behavior have in general not been supported (Swensen, 1968; Kahill, 1984). The validity of global ratings has received more support than sign ratings, with global ratings more able to differentiate groups. Kahill (1984) found that attempts to sort drawings into diagnostic categories have not been encouraging, but that global
ratings and patterns of signs are able to differentiate between groups, or have some relationship to criterion measures more so than individual signs.

**Human Figure Drawings as Expression of Self**

Drawing techniques involve the use of self-drawings to study the personality, finding theoretical justification in self-image psychology, interpersonal and psychoanalytic theory. Schilder (1935) was one of the first to set forth a theoretical base for the drawn "body image" as representative of the person's self-concept, speculating that the body image drawn by an individual is a configuration, or gestalt, composed of many physical, organic and physiological sensations and experiences with one's body. Thus drawings reflect the drawer's perception of self rather than objective reality. Handler (1985), in his clinical work with the DAP, found that people do not represent themselves as they appear physically. He emphasized that to understand how the DAP functions one must understand the principle of symbolic representation, where material is presented in a disguised or symbolically transformed manner. Several major theorists have stressed the importance of understanding the client's perceptive reality (Snygg and Combs, 1949; Rogers 1951). The drawings of an individual are of great importance clinically if they do indeed represent the drawer's conscious or unconscious self-image. Any distortions in the drawing may represent the artist's inadequate or distorted self-image in a symbolic or
literal way. Machover (1949) among others advocated the position that self drawing reflects self-image. She stated:

When an individual attempts to solve the problem of the directive to "draw a person" he is compelled to draw from some source. External figures are too varied in their body attributes to lend themselves to a spontaneous, composite, objective representation of a person. Some process of selection involving identification through projection and introjection enters at some point. The individual must draw consciously, and no doubt unconsciously, his whole system of psychic values. The body, or the self, is the most intimate point of reference in any activity. We have in the course of growth, come to associate various sensations, perceptions and emotions with certain body organs. This investment in body organs, or the perception of the body image as it has developed out of personal experience, must somehow guide the individual who is drawing the specific structure and content which constitutes his offering of a "person" (p.5).

Buck (1948) also theorized that the human figure drawing is a reflection of self-image and even suggested that the drawing of the tree is a projection of the person's adjustment to the natural world and the house drawing is adjustment to the social world. Handler (1985) also saw the human figure drawing as intimately tied to the self. In the process of creating the figure the individual is guided by conscious and unconscious determinants, as well as cultural and social stereotypes which foster body image conception. He felt that certain body images are related to certain personality traits. Hammer (1985) cited as an example Dick Tracy's square jaw which symbolizes strength, virtue and determination, while Elmer Pudd's short, round body symbolizes inadequacy and ineffectiveness. These social images combined with private experiences and the more universal symbols of psychoanalysis and folklore all contribute to the drawer's
complex and subtle projections of self. Thus human beings do not
draw self-pictures that actually represent how they look, but
rather, how they perceive their self-image. Individual when
asked to "draw a person" is forced to draw a figure which
reflects how they feel about themselves, and how they feel about
dealing with the environment.

Hammer (1958) was one of the first to also identify that
individuals may draw the ideal self as well as the perceived real
self. Levy (1950) also posited that the drawing may reflect an
individual's projection of ideal self-image. Koppitz (1968) and
Hammer (1958) also found in their work that the individual
drawings can be either self as experienced or self as ideal.

Several empirically based studies have attempted to relate
the drawn figure to self-concept. Kamano (1960) compared
drawings made by schizophrenics to different aspects of their
self-concepts (actual self, ideal self and least liked self); the
highest correlation was found with actual self-concept. Bodwin &
Bruch (1960) found a correlation of .61 between specific traits
of the drawn figure and judges' ratings of self-concept. In a
number of studies Hunt and Fledman (1960), Silverstein and
Robinson (1961), McHugh (1965) and Bennett (1964,1966) found no
relationship between self-concept and DAP. Van Dyne and
Carskadon (1978) found moderate correlation between undergraduate
subject's self-rating and their ratings of same sex drawings.
It is important, however, to note that generally the studies which yielded no correlation between figure drawing and self-concept utilized a child population, while the studies which yielded positive correlations utilized an adult population. This tends to support the notion that the DAP measures different things for different ages and that a child's self-image is most likely less stable and perhaps expressed via drawings in a different manner than an adult's (Mortensen, 1984).

While there are a number of problems associated with the studies designed to relate self-esteem/self-concept with drawing behavior, one of the primary issues which contributes to research inconsistency is the lack of consensus on the meaning of the words self-concept and self-esteem. Thus a review of the pertinent literature relating to these two concepts are presented below.

**Self-Esteem and Self-Concept**

The study of self-concept represents one of the oldest areas of research in the social sciences. William James' (1890) classic *Principles of Psychology*, the first introductory textbook in psychology, devoted the longest chapter to the self-concept, defined as the "sum total of all that a person can call his" (p.152). While there was an early interest in the self, during the second, third and fourth decades of the twentieth century, constructs concerning the self did not receive much attention from the then dominant behaviorist and functionalist
psychologies. The introspectionists were unable to handle the self and the behaviorists considered such a "mentalistic" construct as self concept unacceptable (Hilgard, 1949).

In the meantime, Freudian and neo-Freudian theorists were formulating psychodynamic postulates which implied a self-referent to make them understandable. However, these theories did not immediately bring the concept of self to the forefront of American psychology for at least two reasons. First, Freud himself, strongly emphasizing the role of the id, did not explicitly formulize a self construct or assign the closely related ego functions much importance. Second, Freud's theory was largely denied or ignored by psychologists who questioned its lack of empirical rigor (Wylie, 1974).

It was not until the 1950's and 1960's that any serious writing and empirical work about the self began. This emergence can be attributed to several influences. In his later writings, Freud (1962) assigned a greater importance to ego development and functioning; the neo-Freudians stressed the importance of the self-picture and the ego-ideal. Concurrently, American psychologists working in clinical areas found the behavioristic models too limited and were ready to entertain psychodynamic ideas. In addition, Gestalt psychologists were injecting their phenomenological methods and theories into the stream of general psychology. Behaviorists began to accept the possibility of complex cognitive and motivational intervening variables
mediating behavior. By the end of the 1960's almost all
personality theorists assigned importance to a phenomenological
and/or nonphenomenological self-concept with cognitive and
motivational attributes (Wylie, 1979).

There are interesting peculiarities concerning self-concept
research. Unlike other areas of research, the study of self-
concept has not occurred within a particular discipline (Marsh,
et al 1983). Also, though many thousands of studies have been
published in this area (Wylie, 1979), only a few researchers have
published a significant number of studies or continued their
research over an extended period of time. In fact, much of the
self-concept research actually focuses upon other theoretical
constructs and the interest in self-concept is secondary, only of
value as it relates to these other constructs (Marsh et al, 1983).

Other significant problems exist in the body of self-concept
research. Most researchers do not distinguish between self-
concept and self-esteem; they often define either of these terms
in quite different ways, and they use different instruments to
measure these constructs (Wylie, 1974). Marsh et al (1983)
commented that in spite of the agreed importance of the issues
and the considerable amount of research in the area of
self-concept, definitions of self-concept are "nonexistent,
imprecise, or contradictory" (p.12), as is the search for the
specific dimensions that underlie self-concept. The constructs
"self-esteem" and "self-concept", like a number of other psychological constructs, appear to suffer from the notion that "everybody knows what it is". Therefore, researchers do not feel compelled to provide any theoretical definition of what it is they are measuring.

In recent years, however, there has been a strong attempt on the part of a number of researchers to more clearly define these two terms as well as to provide theoretical models of the constructs. For example, there now appears to be definitional consensus. A number of authors define self-esteem in similar ways. Wylie (1974) viewed the term as indicative of self-regard, Coopersmith (1967) as a global construct having to do with valuation of one's self, which is compatible with Shavelson, et al (1982) and Rosenberg (1965). These definitions indicate that self-esteem is a subjective, evaluative phenomenon which determines the individual's characteristic perception of personal worth. Self-concept, on the other hand, deals with pure self descriptions, distinguishable from self-esteem which implies self judgments (Fleming and Courtney, 1984).

Most theorists indicate that self-esteem is a part of the self-concept. The self-concept represents a totality of one's perceptions, whereas self-esteem is one dimension of this "totality". Both self-esteem and self-concept are felt to be subjective phenomena which determine the individual's characteristic perception of the self (Battle, 1982).
Because self-esteem is but a portion of self-concept, much of the recent theoretical literature deals with the self-concept. There has, however, been some work regarding self-esteem. A consensus exists that individuals have a need to experience self-esteem and that there is a relationship between the nature and the degree of one's self-esteem and the degree of one's mental health. However, the reasons for needing self-esteem and the underlying conditions contributing to self-esteem are not clear. Also, the nature and causes of the relationship between mental health and self-esteem have not been delineated (Fitts, 1972).

Branden (1969) states that self-esteem has two interrelated aspects: a sense of personal efficacy and a sense of personal worth. Self-esteem is the interrelated sum of self-confidence and self-respect. It is the conviction that "one is competent to live and worth living" (p.110). Self-confidence refers to confidence in one's abilities; self-respect is a moral appraisal based upon one's values. An individual judges the self based upon some standard which determines self-evaluation.

For Rosenberg (1965), high self-esteem as reflected in his Self-Esteem Inventory is the feeling that one is "good enough". The individual feels that she/he is a person of worth, and respects himself/herself for what she/he is. High self-esteem indicates that the individual respects oneself, considers the self worthy; she/he does not necessarily feel better than others but does recognize his/her limitations and expects to grow and
improve. Low self-esteem, on the other hand, implies self-rejection, self-dissatisfaction, self-contempt; the individual lacks respect for the observed self.

In terms of self-concept, there have been recent attempts to provide a comprehensive model which provides an explanation of the construct. A review of literature reveals that Shavelson et al (1976, 1982) is considered to have contributed most significantly to the advancement of knowledge in the field by providing the first comprehensive model of self-concept. Shavelson et al (1976; 1982) reviewed both theoretical and empirical research in this field and used the review as the basis for what has become a widely accepted model of self concept which incorporates aspects from most theoretical positions.

Shavelson et al (1976) defined the self-concept as the individual's perception of the self. Formed through one's experience with and interpretations of the environment, it is influenced especially by reinforcements, interactions with significant others, and attributions of one's own behavior. The construct self-concept is further defined by seven critical features: developmental, organized, multifaceted and hierarchical, stable, descriptive and evaluative, and differentiated.

A number of theorists (Engel, 1959; Sears and Sherman, 1964; Battle, 1982) assumed that self-concept is developmental in nature. The self emerges and takes shape as the child develops
and interacts with significant others in the environment. Although present at birth, the notion of self begins to develop throughout childhood as the child becomes increasingly aware of body image and abilities. While initially a vague, poorly integrated, fragmented phenomenon, the self becomes increasingly more organized, differentiated and integrated as the child matures.

The position that self-concept is organized is based upon the assumption that the individual's experience constitutes the date on which she bases her perceptions of the self. The human organism, in order to reduce the complexity of these experiences, organizes or codes these experiences in a less complex form by placing them into categories. The particular categorical system adopted by the individual is determined by the person's particular cognitive processes particular to culture, but typically revolve around family, friends and work. An individual's descriptive statements about the self will evolve around these categories.

Self-concept is said to be multifacted; many authors (Coopersmith, 1967; Shavelson et al, 1982; Battle, 1982) do not view self-concept as a unitary phenomenon, but rather comprised of a number of interrelated facets. These interrelated factors form a general dimension, or a general view perception of the self. Self-concept is also hierarchical in nature, in that it is composed of various sub-components which may be ranked in a
hierarchical fashion. Self-concept may then be comprised of a hierarchy which ranges from the individual's experiences in a particular situation at the base to general self-concept at its apex. General self-concept is stable, and as one descends the hierarchy, self-concept becomes increasingly multifaceted as the individual develops from infancy to adulthood. It is felt by a number of authors that self-concept becomes fixed in early adulthood (Mortimer, Finch and Kumka, 1981). Once articulated, the forces of cognitive consistency act to maintain the self-concept that the young adult has formed of the self. It is both a descriptive and evaluative dimension such that individuals may describe themselves (e.g., I am happy) and evaluate themselves (e.g., I do well in school). Also, self-concept can be differentiated from other constructs such as academic achievement.

The validity of these assumptions has only been partially confirmed by direct empirical test, although much of the existing evidence may be interpreted as supportive (Shavelson and Bolus, 1982). It is assumed then that a person holds a global or summary self-perception, and that the global self-concept is associated with a correspondingly global judgment of self-worth that is relatively stable. Self-concept is seen to be organized, multi-faceted, hierarchical and stable.
Gender Differences In Self-Esteem and Self-Concept

There are a number of authors who have posited a relationship between self-esteem and gender. These theorists can be placed into one of two categories: Those who attribute the relationship to biological factors and those who define the relationship entirely to social or cultural influences. Freud (1927) associated feelings of superiority with masculinity and inferiority with femininity, attributing the source of this occurrence as stemming from the presence or absence of a penis. The little boy's feeling of superiority originates in his discovery that he possesses something that the little girl lacks, with her sense of inferiority springing from the realization that she has been shortchanged. This insult to her pride sets the girl on a prolonged and difficult developmental course marked by a strong sense of inferiority. Horney (1967) also agreed with Freud's notion of a male genital superiority, but identified as the primary source of female inferiority the young girl is feeling of guilt and anxiety as a result of her libidinal desires for her father. The female child then feels ashamed of her femininity and retreats into a desire to identify with males, damaging her sense of female worth. However, Horney does temper this dismal picture by identifying that the female's capacity for child bearing is a rich source of satisfaction and self-worth. In fact, she suggested that the male tendency to devalue women may stem from their jealousy of the female's life giving power.
A number of theorists have pointed to cultural or social factors as contributing to women's sense of inferiority. Thompson (1950) related the derogatory attitude towards female sexuality prevalent in western culture with women's negative appraisal of themselves. Hacker (1951) pointed to the subordinate role of women in our culture in the social hierarchy and the cultural ideology as the rationale for lowered female self-regard. Chesler (1972) attributed the status of women in a patriarchal society as contributing to a lowered self-esteem. Women have no power in such a society - economic, military or political, and thus have no legacy of female pride.

There are, then, a number of theorists who attested to the notion that women will exhibit a lowered sense of self-worth either due to biological or cultural factors. However, the empirical literature does not seem to support the notion. Wylie's (1979) very extensive review found that whether one studies children, adolescents or adults the findings are the same: Null results prevail. There are several possibilities for the failure to find the hypothesized statistically significant differences in self-esteem between the sexes. While measurement and design problems are often cited as possible artifacts (Wylie, 1979) is it possible that the null results could indeed reflect genuine similarity between the sexes in level of overall self-regard.
While the empirical findings relating to sex differences in self-esteem have been counterintuitive, the findings relating to reported self-concept and sex differences have not. In terms of self-concept it is quite apparent that women and men view themselves in different ways (Maccoby & Jacklin, 1974; O'Leary, 1974). When asked to describe themselves most men and women respond with descriptors that correspond strongly with sexual stereotypes. That is, men consistently describe themselves with such adjectives as resourceful, mature, logical, adventurous, realistic, deliberate, efficient. In contrast, females tend to endorse adjectives such as feminine, emotional, affectionate, pleasant and tempermental (Zahran, 1967). Women score themselves higher on such traits as sociability, dependence, sensitivity and tolerance, whereas men describe themselves as more confident, self-accepting, dominant and realistic (Block, 1973).

These self-ratings tend to coincide strongly with the sex-stereotypes held by both men and women. There is strong evidence that sex-role stereotypes which specify different sets of behaviors expected from men and women are fairly consistent across sex, age and educational levels (Broverman et al, 1972). This evidence indicates that the male role is associated with individual effectiveness, independent competence or agency. On the other hand, the evidence shows that the female role is
associated with interpersonal warmth and expressiveness, or communion (Broverman et al., 1972; Block, 1973).

It is important to note that these sex differences appear only when persons are asked for self-ratings, that is the subjective, phenomenological view of the self, or asked to describe men and women in general. When male and female behaviors are assessed, most studies do not show that the sexes differ on agency or communion behaviors (Broverman et al., 1972; Block, 1973). Although the sexes may differ in the items that they use to define themselves, they do not evaluate the items differently when applying them to self-worth, as indicated in previous discussion that males and females have comparable levels of self-esteem (Maccoby and Jacklin, 1974).

Interestingly, women subjects show significantly lower correlations between mean ideal-self versus mean socially desirable-self, and for mean actual-self versus socially desirable-self than they do mean ideal-self versus mean actual-self. This suggests that females conceive of themselves as living up to their own ideals, but that their personal ideals and conduct are incongruent with what they perceived society desires of them (Block, 1973).

In summary, the majority of studies show that both males and females share the same sex-role stereotypes and tend to apply these stereotypes to themselves. Both men and women tend also to view male sex-role stereotypes as more positive and female
sex-role stereotypes as more negative. However, these perceived sex-role differences are not translated into the self-esteem of the sexes as there consistently appear to be little difference between male and female levels of self-esteem. This later finding appears to be related to the fact that women compare themselves to self expectations rather than societal expectations.

Measurement of Self-Esteem and Self-Concept

The success in any field of research depends in large part on the ability of the researcher to adequately measure the variables of interest. This issue is particularly germane to the study of the self, as the validity of measurement instruments has been the most frequent object of criticism (Wylie, 1974; Shavelson, 1984). Self-concept studies have been criticized from a measurement perspective by a number of authors (Crown and Stephens, 1961; Wylie, 1961; Shavelson et al, 1976).

Attempts to validate measurement instruments are largely missing in the self-concept area. In Wylie's 1974 review of existing self-concept measures she identified several hundred such measures but very few validation attempts. At that time she strongly suggested that more attention be paid to measurement issues versus measurement generation.

Construct validation refers to the procedures and evidence used in support of a construct interpretation of a measurement. Construct validation begins with construct definition (Cronbach,
1982). Because the study of self-concept is at such an immature stage, construct validation has proceeded with only informal definitions. This has caused enormous problems in this field (Shavelson et al, 1976). For example, because there has been a failure to clarify definitions of self-esteem and self-concept, it is difficult to interpret empirical evidence presented in support of the construct validation of a measurement instrument.

Construct validity is critical in the area of self-concept research because of the very nature of the construct. An individual's attitudes and feelings about the self are private and beyond the direct observation of the investigator. In order to obtain ideas about a person's self-concept the examiner must use some form of self-report response. This self-report behavior usually takes the form of a verbal response or choice response. While phenomenologists would like to assume that the subject's self-report responses are determined by his/her phenomenological field, it is obvious that such responses can be influenced by various factors such as attempts to give socially desirable or misleading responses.

It is not clear in many cases exactly what self-concept measures are actually measuring. With any such measure it is possible that subjects may select responses that are socially desirable rather than self-descriptive; individuals may refuse to report their "private" self-concepts (Snygg and Combs, 1949). Crowne and Stephens (1961) and Shavelson et al (1976) concluded
that measures of self-esteem and self-concept are particularly vulnerable to criticism on social desirability grounds. In fact, they suspected that much of the variance among research findings on self-concept may well have been due to the variable of social desirability. In summary, it is important to identify, minimize, and take into account the influence of irrelevant response determiners.

Another difficulty in the interpretation of measures of self-concept is that little data is available regarding the equivalence of various self-concept measurement instruments. As a result it is difficult to generalize interpretations across studies using different instruments. In addition, data exists which suggests that for any given instrument, generalization of construct interpretations across different populations of subjects is questionable (Wylie, 1979).

Shavelson et al (1982) suggested that researchers in this area need to limit the number of self-concept definitions, examine the validity of proposed constructs, and examine rival interpretations of the definitions to ascertain the equivalence of self-concept instruments, so that data collected with one instrument can be related to data collected with another. They cautioned that until some degree of standardization of definition, instrumentation, and interpretation has been accomplished, there can be no fruitful investigation of self-concept. In addition, they criticized the current status of
self-concept research by saying that the bulk of such work has "addressed itself to substantive problems before problems of definition, measurement and interpretation have been resolved" (p.12). Until these problems have been dealt with in a manner made possible by advances in construct validation methodology, the generalizability of self-concept findings will be severely limited, and data on self-concept will continue to be ambiguous.

**Female Adult Development**

This section contains a review of selected aspects of life-span theories of adult development. At this time it appears that most of the research effort in this area does not clearly differentiate between male and female adult development. Therefore, the following discussion will focus upon adult development in general, with inclusion of female developmental theory.

Both psychoanalytic and cognitive-structural theories emphasize the importance of stages in development. The author of the psychoanalytic perspective that has had the greatest impact upon the study of life-span development is Erikson (1968) who proposed that individuals go through eight different stages in the life cycle. Two of these stages focus upon early and middle adulthood - intimacy versus isolation and generativity versus despair. Erikson emphasized that the major developmental theme of early adulthood involves a mutually satisfying and intimate relationship with another individual. In middle adulthood
individuals need to assist the younger generation in developing and leading useful lives, such as the successful rearing of children.

Nonstaged theorists such as the behavioral social learning and humanistic approaches stress that change in adulthood occurs in a continuous, nonstaged manner. For example, in the social learning perspective (Santrock, 1981), change in adulthood comes about through the manner in which the adult processes information about the environment. In contrast, the humanistic perspective suggests that the change in adulthood comes about through the development of the self, specifically the development of self-actualization and self-awareness (Rogers, 1951).

Two additional perspectives on adult development are the views of Havighurst (1972) and Neugarten (1980). Havighurst described a number of developmental tasks that the individual must master in early adulthood (18-30 years):

1. Selecting a mate
2. Learning to live with a partner
3. Starting a family
4. Rearing children
5. Managing a home
6. Getting started in an occupation
7. Taking on civic responsibility
8. Finding a congenial social group
Havinghurst believed that men and women reach their peak in productivity and have their greatest impact on society in middle adulthood (30-55 years), and must master the following tasks:

1. Achieving adult civic and social responsibility
2. Establishing and maintaining an economic standard of living
3. Assisting teen-age children to become responsible adults
4. Developing adult leisure-time activities
5. Relating to one's spouse as a person
6. Accepting the physiological changes of middle age

Neugarten (1980) believed that too much attention has been given to the biological timetable of development and not enough to the social timetable. Choices and dilemmas do not sprout forth at 10-year intervals, and decisions are not made in a linear fashion. Rather, Neugarten argued, that the themes of adulthood appear and reappear in new forms over many periods of time. For example, issues of intimacy are not only a task of early adulthood, but surface for the rest of one's life.

Several authors have stressed, in addition to a stage approach, the importance of crises in adult development. Sheehy (1976) argued that all individuals go through developmental stages, roughly bound by chronological age. Each stage contains problems that must be solved before progression to the next stage is achieved. One strives to attain an "authentic identity" by facing and successfully handling one's life crises.
Gould (1980) also has linked stage and crisis with development. He emphasized that mid-life is just as turbulent as adolescence, except that during middle adulthood striving to handle crisis is likely to lead to a healthier, happier life. Gould's study of males and females led him to propose seven developmental stages of adult life. Four of those stages which pertain to early and middle adulthood are:

1. Learning the family; peer and group orientation
2. Developing independence: Commitment to a career and to children
3. Questioning self: Role confusion; marriage and career vulnerable to dissatisfaction
4. Period of urgency to attain life's goals: Awareness of time limitation and realignment of life's goals

Newman and Newman (1987) identify developmental tasks, psychosocial crises and process for resolution of psychosocial conflict which occurs in adulthood which are set forth in Table 1. Successful completion of tasks and resolution of conflicts enables healthier adult functioning.

Gilligan (1982) has strongly objected to the major models of human development as she feels that they are not representative of the actual female experience. Because most developmental models are based upon male experience, female deviations are generally seen as pathological. She sees major differences between male and female development. Males strive for
TABLE 1

DEVELOPMENTAL TASKS, PSYCHOSOCIAL CRISSES AND PROCESS FOR RESOLUTION OF CONFLICT FOR EARLY AND MIDDLE ADULTHOOD (NEWMAN & NEWMAN, 1987)

<table>
<thead>
<tr>
<th>Developmental Variable</th>
<th>Early Adulthood (23-34 yrs)</th>
<th>Middle Adulthood (35-60 yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developmental Tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Marriage</td>
<td></td>
<td>1. Management of a household</td>
</tr>
<tr>
<td>2. Childbearing</td>
<td></td>
<td>2. Child rearing</td>
</tr>
<tr>
<td>4. Lifestyle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychosocial Crisis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intimacy vs. Isolation</td>
<td></td>
<td>Generativity vs. Stagnation</td>
</tr>
<tr>
<td>Process for Resolution of Psychosocial Crises</td>
<td>Mutuality Among Peers</td>
<td>Environmental Fit and Creativity</td>
</tr>
</tbody>
</table>
separation, autonomy and individuation in the quest for adulthood. Females, in contrast, are more concerned with attachment, nurturance, caretaking, establishing and maintaining relationships.

Summary

Freud's development of a formal personality theory encouraged the creation of projective testing methods to assess personality. With these methods information about personality functioning is gathered by presenting neutral stimuli to the client who responds in terms of the meaning the particular stimulus has for him or her. It is thought that this method of assessment is unique in that projective techniques can enter a person's inner world and obtain information not otherwise revealed through traditional means.

The DAP, one of the more popular projective testing methods, is based upon the assumption that an individual's self drawing reflects self-image. However, like other projective methods, the reliability and validity of the DAP is questionable. In part, this may be due to methodological flaws, such as design and internal validity problems.

One of the major difficulties associated with DAP studies exploring the relationship between self-esteem, self-concept and drawing behavior is the complexity of the self construct. Definitional clarity for the terms self-esteem and self-concept is a recent occurrence as is self-concept theory. Recent
speculation in the self-concept theoretical literature seems to indicate that self-concept appears to be multifaceted and quite complex, but has yet to be confirmed by empirical test. Additionally, because of the subjective and complex nature of the construct, measurement problems are typical.

Adult development models indicate that in early adulthood, individuals begin to integrate previously acquired data about the adult role to themselves (Newman and Newman, 1987). Developmental tasks consist of marriage, childbearing, career. The psychosocial crises revolve around intimacy versus isolation, establishing mutuality among peers and competition among roles. Gilligan (1982) maintains that strong differences exist between adult male and female development. Female development does not parallel male development in that women, more so than men, strive for interdependence and attachment rather than autonomy and individuation.
CHAPTER III
METHODOLOGY

The purpose of this chapter is to describe the methods and procedures. A description of the research setting, subjects, judges, procedures, instrumentation, research design, pilot study, and statistical analysis are provided below.

Research Setting

Both the pilot study (N=41) and the research study (N=100) took place in an outpatient clinic of a community mental health center (CMHC) in a suburb south of Dayton, Ohio, a large metropolitan area in southwestern Ohio. Founded in 1974, the CMHC has five branches and is responsible for delivering mental health services to the southern portion of the county, a diverse population representing the entire socioeconomic spectrum of the county. The CMHC renders 12 types of services: Outpatient, Community Inpatient, Crisis, Hotline, Pre-Hospital Screening, Adult Day Treatment, Case Management, Rehabilitation, Youth Residential Placement, Youth Partial Hospitalization, Forensic Evaluation and Community Services (Consultation and Education). Approximately 185 clients are admitted monthly, with approximately 160 discharges a month. The clinical staff of the
CMHC consists of 45 therapists, representing various educational backgrounds, and includes counselors, mental health technicians, psychiatrists, psychologists, psychology interns, and social workers.

The actual site of the study was an outpatient clinic located in a satellite office of the CMHC. Outpatient services at this location include individual, family, marital, group psychotherapies, as well as psychodiagnostic assessment. This particular branch of the CMHC is located in a suburban municipality, an area which is essentially middle class residential, with a 1980 estimated population of 61,836. Median age is 34, with whites making up 98% of the population, Hispanics .07, and Blacks .04 of the population; median household income is $25,621 (1984) with the unemployment rate at about 3.9%. The three top employers of residents are a manufacturing plant, a government defense contractor and a hospital. Average home price is $70,000.

Subjects

Subjects in this study were drawn from clients in treatment at the outpatient branch of the CMHC described above. Subjects were referred to the study over a one-year period by 10 outpatient therapists at the CMHC. Table 2 contains a summary of both the pilot study and the research study demographic data. For a discussion of demographic subject data see Chapter IV. The modal self-reported marital status of pilot study subjects was
married; modal occupational status skilled trade; average (M) age 33.00 years; average (M) educational level 14.00 years and average (M) income $16,634. The modal marital status of general study subjects was married; modal occupational level skilled trade; the average (M) age of subjects was 35.12 years; average (M) education 14.4 years and average (M) income $18,511.

TABLE 2
PILOT STUDY AND RESEARCH STUDY DEMOGRAPHIC DATA

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pilot Study N=41</th>
<th>Research Study N=100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital Status</td>
<td>Married</td>
<td>Married</td>
</tr>
<tr>
<td>Occupation</td>
<td>Skilled Trade</td>
<td>Skilled Trade</td>
</tr>
<tr>
<td>Age (Years)</td>
<td>33.00</td>
<td>35.12</td>
</tr>
<tr>
<td>Education (Years)</td>
<td>14.00</td>
<td>14.14</td>
</tr>
<tr>
<td>Income</td>
<td>16,634</td>
<td>18,511</td>
</tr>
</tbody>
</table>

Judges

The four judges possessed similar educational and experience backgrounds. Three of the judges were doctoral level licensed psychologists, one judge a Ph.D. candidate. All possessed at least 10 years clinical experience, and training in the area of
testing and assessment. Individual descriptions of each judge follows.

Judge 1: Judge 1, the researcher, is a 36-year-old female, a Ph.D. candidate in Counseling, who possesses 10 years clinical experience in outpatient mental health settings.

Judge 2: Judge 2 is a 38 year old female, a licensed psychologist with a Ph.D. in clinical psychology who possesses 12 years clinical experience.

Judge 3: Judge 3 is a 42 year old male, a licensed psychologist, Ph.D. in clinical psychology, who possesses 14 years of clinical experience.

Judge 4: Judge 4 is a 56 year old male, licensed as a psychologist, with a Psy.D., who possesses 12 years of clinical experience.

Procedures

Subject Selection

A crucial aspect of this type of non-experimental research consists of controlling for the possible effects of extraneous variables. This was accomplished by holding potential confounds constant which means examining only one level of the potential extraneous variables. Because race, ethnic background, age, sex and pregnancy status were found to be possible confounds to drawing performance, only nonpregnant white females, ages 20-45, were selected for inclusion in the study to increase internal
rigor. Subjects were selected in the following manner: The 10 primary therapists in the outpatient clinic were requested by memo (See Appendix A) to assess their caseloads for suitable clients who met the following criteria:

(a) Female, age between 20 and 45, not pregnant.

(b) Native American born, white.

(c) Absence of chronic brain syndrome, mental retardation, psychosis (mental status must have been stable and subject possess adequate intellectual abilities to insure sufficient concentration and attention span to complete testing tasks).

(d) Absence of overt physical handicaps which may confound the drawing performance. The subjects must possess adequate motoric skill to complete drawing task.

Once the primary therapists assessed the potential client as sufficiently in reality contact and of adequate operating intelligence and motoric skill to understand and complete the tasks involved and as meeting the demographic requirements, the clients were asked by the primary therapists if they would be interested in participating in a study of women's self-esteem. Clients who agreed were then referred to the researcher for testing. At that time the client was given a brief explanation (See Appendix B for Script to Subjects) of the study, as required by The Ohio State University Human Subjects Review Committee, and
told that the study was entirely voluntary. If the client was agreeable, a written consent was obtained (see Appendix C).

The Human Subjects Review Committee of The Ohio State University reviewed this research to ensure proper protection of the rights and welfare of the individuals involved. The Committee unanimously approved of the methods used to obtain informed consent and they deemed that subjects were not at risk (Appendix D).

**Data Collection**

The subjects were administered the instrument battery individually and in one sitting taking approximately 30 minutes. The battery consisted of the following instruments administered in the order listed: 1) Draw-A-Person Test (DAP), Rosenberg Self-Esteem Inventory (SEI), 3) Miskimins Self-Goal-Other Discrepancy Scale (MSGO).

Presentation order of the instruments was deliberately not randomly mixed in order to preserve the integrity of the projective instrument. The DAP is seen as extremely sensitive to outside influences; to obtain an undistorted projection of self it is important that the subject not be cued in any way as to the nature of the instrument. To avoid drawing performance being inadvertently influenced by the unwanted influence of possible cues from the other instruments, the DAP was the first instrument presented to subjects. Appendices B and F provide the
instructions given to subjects and administration procedures for each of the three instruments.

Upon completion of the three instruments, the subject was asked to paper clip all three together and place them in a large manilla envelope. To ensure blind analysis by the researcher and avoid criterion contamination, crucial in this kind of validation (Anastasi, 1982), the envelopes with the protocols were then given to a psychology intern at the CMHC for compilation of data, and the scoring of criterion measures. This ensured that subsequent DAP scoring by the researcher and other raters proceeded in isolation.

The psychology intern was a female third year graduate student in a school of professional psychology. She was employed as an intern in the research site facility.

The intern did the following:

1. Assigned a random number code to each set of protocols from a table of random numbers (Kendall and Smith, 1950).

2. Placed this code on each of the criterion measures and the DAP.

3. Recorded pertinent demographic information on a criterion summary sheet for each subject (See Appendix G). The summary sheet contains the following information: age, income, marital status, occupation, scores on criterion measures.

4. Scored the criterion measures (SEI) and (MSCO).
5. Recorded the criterion scores on the Criterion Summary Sheet.

6. Retained the criterion protocols and the criterion summary sheet and communicating none of the results to the researcher, gave to the researcher only the DAP protocols which contained no identifying information.

The researcher and three other raters did the following:

1. In a blind analysis, independently scored each DAP protocol according the the scoring system developed in the pilot study (for instructions to raters see Appendix H). Each rater was given the scoring instructions with minimal additional verbal instructions.

2. All raters recorded the DAP scores on the DAP Summary Sheet (see Appendix I).

Instrumentation

Three instruments were utilized in both the pilot study and in the present research. The DAP was the predictor measure while the Rosenberg Self-Esteem Inventory (RSEI) and the Miskimins Self-Goal-Other Discrepancy Scale (MSGODS) were the two criterion measures (See Appendix E for copy of the RSEI). A copy of the MSGO is not provided due to copyright restrictions. The criterion measures were selected with care to ensure that they were an accurate reflection of the constructs as defined for the purpose of this study. Additionally, both instruments possess high reliability and validity. The RSEI was selected primarily
because it is one of the few unitary measures of self-esteem, and the MSGO because it is one of the few instruments which measure congruence between ideal and real self.

Draw-A-Person Test (DAP)

The DAP is a projective personality pencil and paper test, in which the client is requested to draw human figure drawings which theoretically represent various aspects of the self. The client was presented with three pieces of \(8\frac{1}{2} \times 11\) inch unlined paper (one at a time) and a pencil. The client was requested to "draw a person"; "draw a person of the opposite sex"; and "draw yourself". For a more extensive discussion of the DAP, see Chapter II and Appendix F.

Rosenberg Self-Esteem Inventory (SEI)

The SEI is considered to be one of the best unidimensional measures of self-esteem (Demo, 1985). The instrument is constructed as a Guttman scale and consists of 10 items based upon a four point Likert-scaling format, ranging from "strongly agree" to "strongly disagree".

While the original normative sample of the SEI consisted of over 5,000 high school students, it has been used extensively with adults since that time (Robinson, 1973). Crandall (1973) noted that although brief, the SEI is thorough in its measurement of self-esteem. Wylie (1961) in her classic and extensive review of self-esteem measures viewed the high reliability relationships supporting its construct validity as impressive for such a brief
scale. She recommended that the instrument be used more in research application. Gordon (1969) asserted that the SEI represents some of the best empirical work in the area of self-esteem research.

Rosenberg (1965) reported a coefficient of reproducibility of .92 with his original sample of high school students. Silber and Tippet (1965) found a test-retest coefficient of .85 for college students over a two-week period.

With regard to convergent validity, Silber and Tippet (1965), using a multitrait-multimethod matrix analysis, correlated the SEI with three other self-esteem measures. They reported correlations of .67, .83, and .56. Wylie (1974) commented that these convergent validities are among the highest observed in cross-instrument validation.

**Miskimins Self-Goal-Other Discrepancy Scale (MSGO)**

The MSGO is a psychometric device for measuring the self-concept. The three letters S, G, and O refer to the self-concept, the goal self-concept and the perceived responses of others, respectively. The three variables utilized in the construction of the scale have been drawn directly from a theoretical system proposed by Miskimins (1964). The theory upon which the MSGO is based is essentially behavioristic. The focus of this framework is the object self; it is maintained that a description of an individual's self may be obtained by observing the manner in which the person behaves toward this object. Given
the self as the center of a descriptive system, Miskimins differentiates this concept into three elements. The first, and most influential is the "Self-Concept". This refers to the cognitive scheme that an individual develops about the self. Operationally, the Self-Concept takes the form of the constellation of actional and verbal behaviors which persons direct toward themselves.

The second element of the self measured by the MSGO is the "Goal Self-Concept". This refers to the concept which persons hold regarding themselves as they would like to be. It is assumed by Miskimins that the meaningfulness of an ideal is an individual matter, and each person selects goals which affect his/her own behavior. Operationally, the Goal Self-Concept takes the form of the constellation of actual and verbal behaviors which persons direct toward themselves as they would like to be.

The third element of the self is the "Perceived Responses of Others" representing the concepts a person holds of the self as they feel others perceive them. Operationally, the Perceived Responses of Others takes the form of the constellation of actional and verbal behaviors which persons direct toward themselves as they feel others view them.

The MSGO was constructed to provide a relatively straightforward means of assessing the three dimensions of self in order to calculate the congruencies among the three. The discrepancy scale pertinent to the present study is the scale which measures
the discrepancy between self-concept and goal self-concept. The MSQO was selected because it is one of the few instruments which measures ideal and real self discrepancy and at the same time defines self-concept in a way identical to the present researcher's definition of the research construct.

Subjects are asked to rate: (1) themselves as they actually are, (2) themselves as they would like to be, (3) how they believe others see them. There are 20 items which assess the individual's view of real and ideal self in three content areas: general, social and emotional.

Reliability and validity of the MSQO appear to be fairly good (Robinson 1973). As a measurement instrument it is quite stable over time, as the test-retest reliabilities range from .68 to .86 (Miskimins, 1964). In terms of validity the MSQO correlations with other measures of personality and self-concept range from .42 to .67.

Research Design

For the purpose of this study construct validation paradigm was deemed the most appropriate research design. Construct validation consists of both logical and empirical processes. The procedure utilized in this study follows the construct validation design outlined by Cronbach (1982) and in the APA Standards for Test Validation (1985).
1. Construct Identification

This first step involved identification of constructs which might account for test performance, based upon observation or logical study of the test. It was assumed by the present researcher based upon the literature and projective theory that performance on the DAP is heavily influenced by an individual's self-esteem and self-concept.

2. Formulation of Hypotheses

In this step, the investigator developed tentative hypotheses which might explain differential test performance by examining the test characteristics of subjects with high scores on the criterion measures, in contrast to those with low scores on the criterion measures. In the present study, this was done via a preliminary pilot study, in which the test batteries of 41 subjects were scored, normed and then placed into four criterion groups (high and low self-esteem; poor and good self-concept). The researcher then examined the DAP characteristics of those subjects with high scores on the criterion measures in contrast to those subjects with low scores on the criterion measure (the pilot study procedures and results are described below). This examination resulted in the formation of hypotheses which build a theory about the nature of the construct the DAP is measuring and is based upon clinical judgment. Several authors (Sims et al, 1968) have strongly suggested that DAP hypotheses development and subsequent scoring systems be grounded in clinical judgment.
This type of research is referred to as naturalistic or qualitative research (Lincoln and Guba, 1985). In such research the researcher is using himself or herself as a primary data gathering instrument because it is not possible to invent a nonhuman instrument with the sufficient adaptability to encompass and adapt to the wide variety of stimuli present. Only a clinician is capable of understanding and evaluating the meaning of differential DAP performance.

3. Empirical Verification of the Hypotheses

The third and last step in the construct validation process consisted of the empirical verification of the hypotheses formulated in step two. In this study, the tentative hypotheses formulated to explain DAP performance also formed the basis of a DAP scoring system. In an attempt to empirically verify the hypotheses, additional test data was then gathered from subjects not used in the pilot study and independently scored by the researcher and three other raters in according to this scoring system. As part of the validation process, it was necessary to demonstrate how well the DAP scores correspond to measures of concurrent status. Towards that end the DAP scores were correlated with the criterion measures. Concurrent validation is employed when the tester would like to estimate an individual's present status on some variable external to the test (APA Technical Recommendations, 1978). As noted by Cronbach (1982)
construct validation involves two types of validation, both construct and concurrent.

**Pilot Study**

A pilot study (N=41) was conducted to aid in the hypotheses development upon which the scoring system was based. The pilot study research setting and identification of suitable subjects were similar to the subject selection procedures previously described. In terms of data collection for the pilot study, the researcher was referred 41 female clients who met the subject criteria. Chapter IV contains discussion of pilot study demographic data. Each subject was administered the instrument battery consisting of the three instruments (the DAP, and the two criterion measures, the SEI and MSOC).

The second step in the construct validation process consisted of developing hypotheses about how self-esteem and self-concept levels account for differential drawing performance in women. This logical process in which the DAP protocols of high and low scores on the criterion measures were examined to determine the drawing variables which differentiate the four criterion groups took place in the pilot study. A DAP scoring system was developed to measure two constructs separately, self-concept, self-esteem (Appendix J and Tables 39 and 40) based upon this examination as well as upon a review of the literature and the assumptions outlined in Chapter One. This scoring system
was then utilized to carry out step three of the construct validation procedure.

The criterion groups for the pilot study were determined by the compilation of normative data from the pilot study. Forty-one protocols were scored and from these raw scores the means and standard deviations for each criterion measure were computed (See Table 3). Only subjects with criterion measures scores one standard deviation above and below respective means were included in the pilot study.

**TABLE 3**

PILOT STUDY CRITERION SCORE MEANS, STANDARD DEVIATIONS, RANGES

<table>
<thead>
<tr>
<th>Criterion Measure</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>RANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miskimins</td>
<td>41</td>
<td>63.21</td>
<td>20.22</td>
<td>13-95</td>
</tr>
<tr>
<td>Rosenberg</td>
<td>41</td>
<td>27.35</td>
<td>7.01</td>
<td>12-40</td>
</tr>
</tbody>
</table>

**Hypotheses Generated from Pilot Study**

The purpose of the pilot study was to aid in hypothesis development which would then be used to formulate a scoring system. The hypotheses are discussed.

**Self-Concept**

A person with good self-concept will view her real and ideal self similarly and this will be reflected in her drawing performance. Individuals with good self-concept will draw the
first and third drawings in a congruent fashion. That is, the ego real and ego ideal will be represented in their drawings as similar. Conversely, those with poor-concept will exhibit a marked discrepancy between how they would like to be and how they actually see themselves. Because it was a basic assumption that the DAP #1 and DAP #3 represent the ego-ideal and ego-real respectively, a person with good self-concept will exhibit congruency in drawings #1 and #3. Thus self-concept will be diagnosed by identifying DAP #1 and #3 drawing congruency. The following drawing variables were considered to be strongly influenced by the individual's self-concept and were employed to determine congruency: (1) Size, (2) Placement, (3) Gender Identification, (4) Affect, (5) Age.

In general, an individual with good self-concept will draw the real and ideal selves congruently on these variables. DAP Drawings #1 and #3 will be approximately the same size, placed on the page in a similar location, be of the same sex, affect and age. Appendix J contains a description of these variables and scoring procedures.

Self-Esteem

Persons with high self-regard will view the real self in a positive way and this will be reflected in their portrayal of the real self with the highest overall quality when compared to the other two drawings. Specifically, the person with highest self-esteem will draw DAP #3, the real self, with the highest quality
and in the most positive way. Persons with low self-esteem will draw the ideal self and person of the opposite sex with better quality. In the DAP, this will be reflected by the subject drawing the real self with the highest quality in comparison to the ego ideal or others. Drawing variables associated with self-esteem are: (1) Size, (2) Overall Quality, (3) Sexual Differentiation, (4) Placement, (5) Omissions.

In general, persons with high self-esteem will draw the real self about 6-7 inches in height, with overall good quality, exhibiting secondary sexual characteristics, placed in the middle of the page, and with no major omissions of body parts. Appendix J and Tables 39 and 40 contain a more detailed account of these variables and the scoring procedure.

**Statistical Analysis**

The researcher and three other raters independently and in a blind analysis scored the DAP protocols according to the scoring system developed from the pilot study in order to empirically test the assumptions set forth in the pilot study. To determine the relationship between the DAP scores and the criterion measures, correlational statistics were used. The statistical analysis employed in this study are described below by the four research questions. Statistical analyses were performed by SAS and SPSS-X programs, The Ohio State University Computer Systems.
Research Question One:

The individual DAP drawing variables were correlated with the criterion measure scores by use of biserial correlation \( r_{bis} \) as is appropriate when examining the relationship between continuous scores and scores in the form of an artificial dichotomy (Borg and Gall, 1983). The relationship between the DAP composite scores and the criterion measures were examined with the Pearson product moment correlation \( r \) (Baggaley, 1964).

Research Question Two:

The ability of the DAP scoring system to differentiate between various levels of the criterion measures were demonstrated by Fisher's \( Z(r) \) transformation (Gravetter and Walnau, 1985). Criterion group scores were correlated with DAP drawing variable scores by use of biserial correlation \( r_{bis} \). The Pearson's product moment correlation was used to correlate DAP composite scores with criterion group scores.

Research Question Three:

Inter-rater reliability was examined, utilizing the tetrachoric correlation (Thorndike, 1978) for Drawing Variable scores and Pearson's product moment correlation for DAP composite scores with criterion measure scores. All judges pair-wise comparisons were examined for statistical significance at the .05 level.
Research Question Four:

To determine if inter-rater reliability is higher for various levels of the criterion measures Fisher's Z(r) transformation was utilized.

In addition, demographic data for both the pilot study and the research study were gathered. These data were analyzed by use of descriptive statistics.

It was also necessary to make some adjustments to criterion measure scores because for both instruments (SEI and MSGO), high numerical scores originally represented low values of self-esteem and self-concept. Therefore, as is typically the practice with these two instruments (Robinson, 1973), the criterion measure scores were subtracted from 100 to allow meaningful correlational analysis.
CHAPTER IV
RESULTS

This study was designed to examine the validity of the Draw-A-Person Test (DAP) as a measure of self-concept and self-esteem in outpatient females. Towards that end, using a construct validation design, a DAP scoring system was developed in a pilot study (N=41). This scoring system was then used by four independent judges to score DAP protocols from 100 subjects, adult female clients in active treatment at a community mental health center in a large midwestern city. Subjects also completed objective measures of self-esteem (Rosenberg SEI) and self-concept (Miskimin's SGODS) which were employed as criterion measures. The findings are reported below by demographic data and by research question, followed by a discussion and a summary of the major findings.

Findings by Demographic Variables

Subject demographic data are presented below by variable and by pilot study and research study (Tables 4 and 5). Demographic data by level of criterion group for both pilot and study subjects is also presented (Tables 6 and 7).
### TABLE 4

**DEMOGRAPHIC DISTRIBUTION OF PILOT STUDY SAMPLE**

**BY FREQUENCY AND PERCENT**

**N=41**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level</th>
<th>f</th>
<th>%</th>
<th>cum f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital Status</td>
<td>Single</td>
<td>13</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Married</td>
<td>14</td>
<td>35</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Divorced</td>
<td>14</td>
<td>33</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td><strong>Totals</strong></td>
<td><strong>41</strong></td>
<td><strong>100</strong></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td>Secretarial</td>
<td>6</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Skilled Trade</td>
<td>15</td>
<td>37</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>White Collar</td>
<td>6</td>
<td>15</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Professional</td>
<td>8</td>
<td>20</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Homemaker</td>
<td>6</td>
<td>15</td>
<td>102*</td>
</tr>
<tr>
<td></td>
<td><strong>Totals</strong></td>
<td><strong>41</strong></td>
<td><strong>102</strong>*</td>
<td></td>
</tr>
<tr>
<td>Education (Years)</td>
<td>11</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>11</td>
<td>27</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>2</td>
<td>5</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>9</td>
<td>22</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>15</td>
<td>36</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>2</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td><strong>Totals</strong></td>
<td><strong>41</strong></td>
<td><strong>100</strong></td>
<td></td>
</tr>
<tr>
<td>Age (Years)</td>
<td>20-24</td>
<td>6</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>25-29</td>
<td>9</td>
<td>23</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>30-34</td>
<td>13</td>
<td>32</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>35-40</td>
<td>4</td>
<td>9</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>41-45</td>
<td>9</td>
<td>22</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td><strong>Totals</strong></td>
<td><strong>41</strong></td>
<td><strong>100</strong></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>2,000-19,000</td>
<td>24</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>20,000-27,000</td>
<td>9</td>
<td>22</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>28,000-35,000</td>
<td>6</td>
<td>15</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>36,000-50,000</td>
<td>1</td>
<td>2</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>51,000-66,000</td>
<td>0</td>
<td>0</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>67,000-82,000</td>
<td>1</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td><strong>Totals</strong></td>
<td><strong>41</strong></td>
<td><strong>100</strong></td>
<td></td>
</tr>
</tbody>
</table>

* % > 100 due to rounding error
<table>
<thead>
<tr>
<th>Variable</th>
<th>Level</th>
<th>f</th>
<th>%</th>
<th>cum %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital Status</td>
<td>Single</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Married</td>
<td>43</td>
<td>43</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Divorced</td>
<td>24</td>
<td>24</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td>Secretarial</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Skilled Trade</td>
<td>33</td>
<td>33</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>White Collar</td>
<td>15</td>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Professional</td>
<td>20</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Homemaker</td>
<td>20</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Education (Years)</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>22</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>7</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>25</td>
<td>25</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>8</td>
<td>8</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>33</td>
<td>33</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Age (Years)</td>
<td>20-25</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>26-30</td>
<td>30</td>
<td>30</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>31-35</td>
<td>16</td>
<td>16</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>36-40</td>
<td>21</td>
<td>21</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>41-45</td>
<td>22</td>
<td>22</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>2,000-10,000</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>11,000-20,000</td>
<td>35</td>
<td>35</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>21,000-30,000</td>
<td>25</td>
<td>25</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>31,000-40,000</td>
<td>6</td>
<td>6</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>41,000-50,000</td>
<td>4</td>
<td>4</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>51,000-60,000</td>
<td>4</td>
<td>4</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>61,000-70,000</td>
<td>2</td>
<td>2</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>71,000-110,000</td>
<td>3</td>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 6

DEMOGRAPHIC DATA OF PILOT STUDY SAMPLE
BY CRITERION GROUP AND LEVEL

<table>
<thead>
<tr>
<th>Variable</th>
<th>Miskimin's Group N=41</th>
<th>Rosenberg's Group N=41</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High n=10, Medium n=23, Low n=8</td>
<td>High n=9, Medium n=22, Low n=10</td>
</tr>
<tr>
<td>Education M</td>
<td>15 yrs. 15 yrs. 13 yrs.</td>
<td>15 yrs. 14 yrs. 13 yrs.</td>
</tr>
<tr>
<td>Income M</td>
<td>$18,500 $13,750 $27,000</td>
<td>$40,143 $16,720 $13,000</td>
</tr>
<tr>
<td>Age M</td>
<td>33 33 38</td>
<td>36 31 36</td>
</tr>
<tr>
<td>Occupational Status by Mode</td>
<td>White Collar White Collar Skilled Trade</td>
<td>Profess. White Collar White Collar</td>
</tr>
<tr>
<td>Marital Status by Mode</td>
<td>Married Married Married</td>
<td>Divorced Married Married</td>
</tr>
</tbody>
</table>
### TABLE 7

DEMOGRAPHIC DATA OF RESEARCH STUDY SAMPLE
BY CRITERION GROUP AND LEVEL

<table>
<thead>
<tr>
<th>Variable</th>
<th>Miskimin's Group</th>
<th>Rosenberg's Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High n=18</td>
<td>Medium n=69</td>
</tr>
<tr>
<td>Education M</td>
<td>16 yrs.</td>
<td>15 yrs.</td>
</tr>
<tr>
<td>Income M</td>
<td>$27,500</td>
<td>$12,200</td>
</tr>
<tr>
<td>Age M</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Modal Occupation Status</td>
<td>White Collar</td>
<td>Skilled Trade</td>
</tr>
<tr>
<td>Modal Marital Status</td>
<td>Single</td>
<td>Married</td>
</tr>
</tbody>
</table>
Marital Status

**Pilot Study** (N=41)

The modal self-reported marital status of subjects in the pilot study was married. Thirty-five percent of the subjects reported they were married, 33% divorced and 32% single (never married) (Table 4).

**Research Study** (N=100)

The modal reported marital status of subjects was married. Forty-three percent of the research study subjects reported that they were married, 24% divorced and 33% single (never married) (Table 5).

Occupational Status

**Pilot Study** (N=41)

The modal self-reported occupational status of subjects in the pilot study was skilled trade. Thirty-seven percent reported occupations of skilled trade, 20% professional, 15% secretarial, 15% white collar and 15% homemaker (Table 4).

**Research Study** (N=100)

The modal self-reported occupational status of study subjects was skilled trade. Thirty-three percent reported occupation of skilled trade, 20% professional, 20% homemaker, 15% white collar and 12% secretarial (Table 5).
Educational Level

Pilot Study (N=41)

The average (M) years of education obtained by self-report of subjects was 14.00 years. Thirty-six percent reported 16 years of education, 27% reported 12 years, 22% reported 15 years, 5% reported 11 years, 5% reported 13 years and 0% reported 14 years (Table 4).

Research Study (N=100)

The average (M) reported educational level for the research study subjects was 14.14 years. Thirty-three percent reported 16 years of education; 25% reported 14 years; 22%, 12 years; 8%, 15 years; 7%, 13 years and 3%, 11 years (Table 5).

Age

Pilot Study (N=41)

The average (M) self-reported age of pilot study subjects was 33.00 years. Thirty-two percent reported 30-34 years; 23%, 25-29 years; 14%, 20-24 years and 9%, 35-40 years (Table 4).

Research Study (N=100)

The average (M) self-reported age of study subjects was 35.12 years. Thirty percent reported 26-30 years, 22% reported 41-45 years, 21% reported 36-40 years, 16% reported 31-35 years and 11% reported 20-25 years (Table 5).
Income

**Pilot Study** (N=41)

The average (M) income as reported by pilot study subjects was $16,634. Fifty-nine percent reported incomes in the range of $2,000-9,000, 22% reported 20,000-27,000 range, 15% reported 28,000-35,000 range, 2% reported 36,000-50,000 range, 2% reported 67,000-82,000 range and 0% reported 51,000-66,000 range (Table 4).

**Research Study** (N=100)

The average (M) income for study subjects was $18,511. Thirty-five percent reported incomes in the range of $11,000-20,000, 25% reported 21,000-30,000 range, 21% reported 2,000-10,000 range, 6% reported 31,000-40,000, 4% reported 41,000-5,000 range, 4% reported 51,000-60,000 range, 3% reported 71,000-110,000 range and 2% reported 61,000-70,000 range (Table 5).

**Demographic Variables by Criterion Group Level**

Subject demographic data by criterion group level for both pilot study and research study is presented below. Table 6 contains a summary of pilot study data; Table 7 contains a summary of research study data.

**Pilot Study** (N=41)

The Miskimin's high group (N=10) self- reported a modal marital status of married, a modal occupational level of white collar, average (M) age of 33 years, an average (M) income of
$18,500, average (M) educational level of 15 years. The Miskimin’s medium group (N=23) reported a modal status of married, a modal occupational status of white collar, an average (M) age of 33 years, an average (M) income of $13,750, an average (M) educational level of 15 years. The Miskimin’s low group (N=8) self-reported, consisted of a modal marital status of married, a modal occupational of status skilled trade, an average (M) age of 38 years, an average (M) income of $27,000, and an average (M) education level of 17 years (Table 6).

The Rosenberg high group (N=9) reported a modal marital status of married, a modal occupational level of professional, an average (M) age of 36 years, an average (M) income of $50,143 and an average (M) education of 15 years. The Rosenberg medium group subjects (N=22) reported a modal marital status of married, a modal occupational level of white collar, an average (M) age of 31 years, an average (M) income of $16,720 and an average (M) education of 14 years.

Rosenberg low group subjects (N=10) reported a modal marital status of married, a modal occupational status of white collar, an average (M) age of 36 years, an average (M) income of $13,000, and an average (M) education of 13 years (Table 6).

Research Study

The Miskimin’s high group (N=18) self reported the following: modal marital status of single, modal occupational status of white collar, average (M) age of 33 years, average (M)
income of $17,500, average (M) education of 16 years. Miskimin's medium group (N=69) self reported the following: modal marital status of single, modal occupational status of skilled trade, average (M) age of 33 years, average (M) income of $12,200, average (M) education of 12 years. The Miskimin's low group (N=13) consisted of the following: modal marital status of married, modal occupation status of white collar, average (M) age of 38 years, average (M) income of $25,300, average (M) education of 12 years (Table 7).

The Rosenberg high group (N=16) self-reported a modal marital status of married, a modal occupational status of status skilled trade, an average (M) age of 36 years, an average (M) income of $30,139, an average (M) education 16 years. The Rosenberg medium group (N=66) reported a modal marital status of married, a modal occupational status of skilled trade, an average (M) age of 31 years, an average (M) income of $13,200, an average (M) education 14 years. The low Rosenberg group (N=18) self reported a modal marital status of divorced, a modal occupation of skilled trade, an average (M) age of 36 years, an average (M) income of $12,000 and average (M) education 13 years (Table 7).

Findings by Research Question

The research study findings are presented below. The results are provided by research question.
Research Question One

Is there a relationship between the DAP scoring system measures of self-concept and self-esteem and respective criterion measures?

To answer this question DAP Self-Concept and DAP Self-Esteem scores obtained by four independent judges using the developed DAP scoring system were correlated with their respective criterion measures (Miskimin's SCODS and Rosenberg SEI). Because all DAP scores, except the DAP Composite scores, are in the form of an artificial dichotomy and were correlated with criterion measures in the form of continuous scores the biserial correlation \( r_{bis} \) was used. For DAP Composite scores, continuous scores, the Pearson's product moment correlation \( r \) was employed. The resulting correlation coefficients were then examined to determine statistical significance at the .05 level.

Table 8 presents correlations of the DAP Self-Concept scores and the criterion measure (Miskimin's SCODS) by judge. Inspection of the table shows that only two of the 24 correlations were statistically significant. The DAP Self-Concept variable of size is negatively correlated \( r_{bis} = -.22, p < .05 \) (Judge 4) with criterion measure scores. This indicates that, contrary to expectations, subjects who scored higher on the criterion measure (Miskimin's SCODS), reflecting high congruence between real and ideal self, received from one
judge lower DAP scores indicating low congruence between drawings of real and ideal self.

The DAP scoring variable gender identification is positively correlated ($r_{bis} = .24$, $p < .05$) (Judge 4) with the criterion measure score. This indicates that, as hypothesized, persons who scored higher on the Miskimin's SGODS also drew real and ideal self gender congruently.

**TABLE 8**

**BISERIAL CORRELATION OF DAP SELF-CONCEPT DRAWING VARIABLE SCORES AND PEARSON'S PRODUCT MOMENT CORRELATION OF DAP SELF-CONCEPT COMPOSITE SCORES WITH CRITERION MEASURE (MISKIMIN'S SGODS) BY JUDGE**

<table>
<thead>
<tr>
<th>DAP Variable</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>-.1706</td>
<td>-.1243</td>
<td>-.1842</td>
<td>-.2210*</td>
</tr>
<tr>
<td>Placement</td>
<td>.0516</td>
<td>-.0559</td>
<td>.0430</td>
<td>-.0716</td>
</tr>
<tr>
<td>Gender Identification</td>
<td>.0564</td>
<td>-.1300</td>
<td>.0841</td>
<td>.2366*</td>
</tr>
<tr>
<td>Affect</td>
<td>-.0291</td>
<td>.0168</td>
<td>.0548</td>
<td>-.0407</td>
</tr>
<tr>
<td>Age</td>
<td>.1678</td>
<td>.1792</td>
<td>.1217</td>
<td>-.0914</td>
</tr>
<tr>
<td>Composite Score</td>
<td>.0370</td>
<td>.0147</td>
<td>.0939</td>
<td>-.0871</td>
</tr>
</tbody>
</table>

* $p < .05$

Table 9 presents the correlations between DAP Self-Esteem scores with the criterion measure of self-esteem (the Rosenberg SEI) by judge. As indicated, Judge 1's scores on two DAP self-esteem variables achieved statistical significance: size
(r=.21, p<.05) and Composite Score (r=.23, p<.05) for Judge 1. The first significant correlation indicates that as subjects scored higher on the self-esteem criterion measure, as hypothesized, they also scored higher on the DAP size variable, drawing real self within 6-7 inches. The significant Composite Score correlation indicates that one judge's (Judge 1) overall DAP score was able to predict subjects' scores on the criterion measure. In summary, it can be concluded for Research Question One that, in general, there appears to be no significant relationship between the DAP scoring system used in this study and respective criterion measures.

**TABLE 9**

**BISERIAL CORRELATION OF DAP SELF-ESTEEM DRAWING VARIABLES AND PEARSON'S PRODUCT MOMENT CORRELATION OF DAP SELF-ESTEEM COMPOSITE SCORE WITH CRITERION MEASURE (ROSENBERG SEI) BY JUDGE**

<table>
<thead>
<tr>
<th>DAP Variable</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>.2130*</td>
<td>.1497</td>
<td>.1529</td>
<td>.1862</td>
</tr>
<tr>
<td>Quality</td>
<td>.0930</td>
<td>-.0249</td>
<td>-.1000</td>
<td>-.0634</td>
</tr>
<tr>
<td>Sexual Identification</td>
<td>.1976</td>
<td>.1744</td>
<td>.0291</td>
<td>.1847</td>
</tr>
<tr>
<td>Placement</td>
<td>.1162</td>
<td>-.0983</td>
<td>.0039</td>
<td>.1461</td>
</tr>
<tr>
<td>Omissions</td>
<td>.0483</td>
<td>.0062</td>
<td>.1634</td>
<td>.0163</td>
</tr>
<tr>
<td>Composite Score</td>
<td>.2277*</td>
<td>.0550</td>
<td>.0848</td>
<td>.1805</td>
</tr>
</tbody>
</table>

*p<.05*
Research Question Two

Can the DAP scoring system of self-concept and self-esteem discriminate between various levels of the criterion measure?

This question explores whether the DAP scoring system produces scores with higher validities for various levels of the criterion measures. Toward that end, for each of the criterion measures, subjects were placed into one of three groups (high, medium, low) based upon criterion measure scores. Score cutoffs for determination of group placement were determined by the criterion measure mean and standard deviation. Subjects achieving scores one standard deviation above mean were placed in the high groups \( (N=13) \), one standard deviation below the mean in the low groups \( (N=18) \) and scores within one standard deviation above and below the mean were placed in the medium groups \( (N=69) \). Table 10 presents the mean, standard deviation, and range for each of the criterion measures. Each subject's score was then correlated using the Pearson's product moment correlation for DAP Composite Scores and the biserial correlation for DAP Drawing Variables with respective criterion measure scores. Significant correlations were to be further analyzed via Fisher's \( z \) (r) transformation to examine whether the correlations differed significantly by group.

Tables 11, 12, 13 show the correlations between the DAP Self-Concept scores and the criterion measure (Miskimin's SGODS).
TABLE 10
CRITERION MEASURE SCORE RANGES, MEANS AND STANDARD DEVIATIONS

<table>
<thead>
<tr>
<th>Criterion Measure</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miskimin's SGODS</td>
<td>100</td>
<td>68.22</td>
<td>19.38</td>
<td>7-96</td>
</tr>
<tr>
<td>Rosenberg SEI</td>
<td>100</td>
<td>29.45</td>
<td>6.79</td>
<td>12-40</td>
</tr>
</tbody>
</table>

TABLE 11
BISERIAL CORRELATION OF DAP SELF-CONCEPT DRAWING VARIABLES AND PEARSON'S PRODUCT MOMENT CORRELATION OF DAP SELF-CONCEPT COMPOSITE SCORE WITH LOW CRITERION GROUP SCORES BY JUDGE N=13

<table>
<thead>
<tr>
<th>DAP Variable</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>.0014</td>
<td>.0895</td>
<td>.0014</td>
<td>.1144</td>
</tr>
<tr>
<td>Placement</td>
<td>.0373</td>
<td>-.1989</td>
<td>.0466</td>
<td>.1914</td>
</tr>
<tr>
<td>Gender</td>
<td>.3180</td>
<td>-.0319</td>
<td>.3338</td>
<td>-.0233</td>
</tr>
<tr>
<td>Affect</td>
<td>.1332</td>
<td>.0871</td>
<td>.0785</td>
<td>.1041</td>
</tr>
<tr>
<td>Age</td>
<td>-.0786</td>
<td>.0643</td>
<td>.025</td>
<td>-.1178</td>
</tr>
<tr>
<td>Composite Score</td>
<td>-.0915</td>
<td>-.0702</td>
<td>-.1663</td>
<td>.1267</td>
</tr>
</tbody>
</table>
TABLE 12
BISERIAL CORRELATION OF DAP SELF-CONCEPT DRAWING VARIABLE SCORES AND PEARSON'S PRODUCT MOMENT CORRELATION OF DAP SELF-CONCEPT COMPOSITE SCORES WITH MEDIUM CRITERION GROUP BY JUDGE
N=69

<table>
<thead>
<tr>
<th>DAP Variable</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>.0059</td>
<td>.0557</td>
<td>.0712</td>
<td>.0601</td>
</tr>
<tr>
<td>Placement</td>
<td>-.0344</td>
<td>-.2030</td>
<td>-.0756</td>
<td>-.0828</td>
</tr>
<tr>
<td>Gender</td>
<td>.0856</td>
<td>.1334</td>
<td>.0118</td>
<td>.1081</td>
</tr>
<tr>
<td>Affect</td>
<td>.0530</td>
<td>.0636</td>
<td>.2215</td>
<td>-.0667</td>
</tr>
<tr>
<td>Composite Score</td>
<td>.0942</td>
<td>.0867</td>
<td>.0632</td>
<td>-.0255</td>
</tr>
</tbody>
</table>

TABLE 13
BISERIAL CORRELATION OF DAP SELF-CONCEPT DRAWING VARIABLES AND PEARSON'S PRODUCT MOMENT CORRELATION OF DAP SELF-CONCEPT COMPOSITE SCORE WITH HIGH CRITERION GROUP SCORES BY JUDGE
N=18

<table>
<thead>
<tr>
<th>DAP Variable</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>-.3944</td>
<td>-.2692</td>
<td>-.2295</td>
<td>-.1443</td>
</tr>
<tr>
<td>Placement</td>
<td>.0165</td>
<td>.2349</td>
<td>.1764</td>
<td>.1017</td>
</tr>
<tr>
<td>Gender</td>
<td>.0610</td>
<td>-.0389</td>
<td>-.3394</td>
<td>.0610</td>
</tr>
<tr>
<td>Affect</td>
<td>.0603</td>
<td>-.1909</td>
<td>.4300</td>
<td>-.1614</td>
</tr>
<tr>
<td>Age</td>
<td>-.1112</td>
<td>-.1886</td>
<td>-.3250</td>
<td>.0147</td>
</tr>
<tr>
<td>Composite Score</td>
<td>-.1349</td>
<td>-.5455*</td>
<td>-.1032</td>
<td>-.0471</td>
</tr>
</tbody>
</table>

*p < .05
by group and by judge. Examination of the tables shows that only one correlation is statistically significant. One judge's DAP Self-Concept Composite score (Judge 2), as shown in Table 11, is negatively correlated with the criterion measure \( r = -0.55, p < .05 \). That is, contrary to expectations, as subjects scored higher on the Miskimin's SGODS, they received lower DAP Self-Concept scores from one judge.

Tables 14, 15, 16 summarize the correlations between the DAP Self-Esteem scores and the Rosenberg SEI by group and by judge. As indicated above, there are no statistically significant correlations between the DAP scoring system and its criterion measure.

The data gathered to address this research question failed to achieve statistical significance. Therefore, it was decided to forego further differential group analysis via Fisher's \( z (r) \) transformation. In summary, for Research Question Two, there appears to be no statistically significant relationship between the DAP scoring system and respective criterion measures when examining high, medium and low levels of self-concept and self-esteem.

Research Question Three

Will the four raters, independently utilizing the researcher's rating system, demonstrate agreement?
### TABLE 14

BISERIAL CORRELATION OF DAP SELF-ESTEEM DRAWING VARIABLE SCORES AND PEARSON'S PRODUCT MOMENT CORRELATION OF DAP SELF-ESTEEM COMPOSITE SCORES WITH LOW CRITERION GROUP SCORES BY JUDGE N=18

<table>
<thead>
<tr>
<th>DAP Variable</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>.2261</td>
<td>.1479</td>
<td>.1479</td>
<td>.1479</td>
</tr>
<tr>
<td>Placement</td>
<td>.0412</td>
<td>.0190</td>
<td>-.0369</td>
<td>.2236</td>
</tr>
<tr>
<td>Gender</td>
<td>.2467</td>
<td>.0966</td>
<td>-.1479</td>
<td>.2466</td>
</tr>
<tr>
<td>Affect</td>
<td>.1829</td>
<td>-.1539</td>
<td>-.0941</td>
<td>.1829</td>
</tr>
<tr>
<td>Age</td>
<td>.0966</td>
<td>.1109</td>
<td>.2084</td>
<td>-.0590</td>
</tr>
<tr>
<td>Composite Score</td>
<td>.1339</td>
<td>.0540</td>
<td>.0035</td>
<td>.2381</td>
</tr>
</tbody>
</table>

### TABLE 15

BISERIAL CORRELATION OF DAP SELF-ESTEEM DRAWING VARIABLE SCORES AND PEARSON'S PRODUCT MOMENT CORRELATION OF DAP SELF-ESTEEM COMPOSITE SCORES WITH MEDIUM CRITERION GROUP SCORES BY JUDGE N=66

<table>
<thead>
<tr>
<th>DAP Variable</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>-.0471</td>
<td>-.0462</td>
<td>.0471</td>
<td>.0228</td>
</tr>
<tr>
<td>Placement</td>
<td>.1486</td>
<td>-.0184</td>
<td>.0932</td>
<td>.1258</td>
</tr>
<tr>
<td>Gender</td>
<td>.1685</td>
<td>.1685</td>
<td>.0420</td>
<td>.1385</td>
</tr>
<tr>
<td>Affect</td>
<td>.0375</td>
<td>-.2504</td>
<td>.0282</td>
<td>.0810</td>
</tr>
<tr>
<td>Age</td>
<td>.0609</td>
<td>-.0061</td>
<td>.0493</td>
<td>.0279</td>
</tr>
<tr>
<td>Composite Score</td>
<td>.1276</td>
<td>-.0794</td>
<td>-.0320</td>
<td>-.0567</td>
</tr>
</tbody>
</table>
TABLE 16

BISERIAL CORRELATION OF DAP SELF-ESTEEM DRAWING VARIABLE SCORES AND DAP SELF-ESTEEM COMPOSITE SCORES WITH HIGH CRITERION GROUP SCORES BY JUDGE
N=16

<table>
<thead>
<tr>
<th>DAP Variables</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>.1270</td>
<td>.1709</td>
<td>.2304</td>
<td>.1270</td>
</tr>
<tr>
<td>Placement</td>
<td>.0893</td>
<td>.0689</td>
<td>-.3156</td>
<td>-.2259</td>
</tr>
<tr>
<td>Gender</td>
<td>.0000</td>
<td>-.1560</td>
<td>-.1561</td>
<td>.0000</td>
</tr>
<tr>
<td>Affect</td>
<td>.1632</td>
<td>.1709</td>
<td>.0272</td>
<td>.0499</td>
</tr>
<tr>
<td>Size</td>
<td>.2160</td>
<td>.2825</td>
<td>.1591</td>
<td>.0970</td>
</tr>
<tr>
<td>Composite Score</td>
<td>.2563</td>
<td>.1851</td>
<td>.0209</td>
<td>.0539</td>
</tr>
</tbody>
</table>
To determine inter-rater reliability, or the accuracy of scoring, correlation coefficients were calculated for all pair-wise comparisons. All DAP scores except for the Composite scores were analyzed with the tetrachoric correlation ($r_t$), used when both variables are in the form of artificial dichotomies. The Composite score reliabilities were determined with the Pearson's product moment correlation ($r$). The obtained correlations represent the strength of scorer agreement. Inter-rater reliability correlations of at least .21 were needed to achieve statistical significance at the .05 level. However, examination of the current DAP literature indicates that most DAP researchers are able to achieve fairly high reliabilities of greater than or equal to .60. It was expected, then, that inter-rater reliabilities for this study would fall within that range.

Tables 17-22 show the inter-rater reliability correlations for DAP Self-Concept scores by judge for each scoring variable. As indicated, the rater agreements range from .20 to .79. All of the 36 pair-wise rater comparisons were statistically significant at the .05 level, with the exception of one comparison (Table 18, Judge 1 and Judge 2, $r=.20$ for Placement variable). In addition, most (92%) of these reliability coefficients were significant at the .005 level. Thirty-eight percent of all pair-wise comparisons achieved correlations greater than or equal to .60. On several of the scoring
### TABLE 17

**DAP SELF-CONCEPT INTER-RATER RELIABILITY (TETRACHORIC CORRELATION) BY JUDGE**

**DAP SIZE VARIABLE**

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.6859**</td>
<td>.7356**</td>
<td>.6339**</td>
</tr>
<tr>
<td>Judge 2</td>
<td></td>
<td>1.00</td>
<td>.7940**</td>
<td>.7749**</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td></td>
<td>1.00</td>
<td>.7748**</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

**p < .005**

### TABLE 18

**DAP SELF-CONCEPT INTER-RATER RELIABILITY (TETRACHORIC CORRELATION) BY JUDGE**

**DAP PLACEMENT VARIABLE**

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.1958</td>
<td>.3940**</td>
<td>.2120*</td>
</tr>
<tr>
<td>Judge 2</td>
<td></td>
<td>1.00</td>
<td>.4477**</td>
<td>.5810**</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td></td>
<td>1.00</td>
<td>.5366**</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

* *p < .05
** *p < .005
**TABLE 19**

DAP SELF-CONCEPT INTER-RATER RELIABILITY 
(TETRACHORIC CORRELATION) BY JUDGE 
DAP GENDER IDENTIFICATION VARIABLE

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.7105**</td>
<td>.6422**</td>
<td>.6568**</td>
</tr>
<tr>
<td>Judge 2</td>
<td></td>
<td>1.00</td>
<td>.6510**</td>
<td>.5666**</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td></td>
<td>1.00</td>
<td>.6372**</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

**p < .005**

**TABLE 20**

DAP SELF-CONCEPT INTER-RATER RELIABILITY 
(TETRACHORIC CORRELATION) BY JUDGE 
DAP AFFECT VARIABLE

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.3623**</td>
<td>.5275**</td>
<td>.3895**</td>
</tr>
<tr>
<td>Judge 2</td>
<td></td>
<td>1.00</td>
<td>.2857**</td>
<td>.3201**</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td></td>
<td>1.00</td>
<td>.3035**</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

**p < .005**
### TABLE 21

**DAP SELF-CONCEPT INTER-RATER RELIABILITY**  
(TETRACHORIC CORRELATION) **BY JUDGE**  
**DAP AGE VARIABLE**

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>0.3967**</td>
<td>0.5029**</td>
<td>0.3104**</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td></td>
<td>0.4543**</td>
<td>0.2050</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td>1.00</td>
<td></td>
<td>0.3354**</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**p < .005**

### TABLE 22

**DAP SELF-CONCEPT INTER-RATER RELIABILITY**  
(PEARSON'S PRODUCT MOMENT CORRELATION) **BY JUDGE**  
**DAP COMPOSITE SCORE**

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>0.5717**</td>
<td>0.6446**</td>
<td>0.5206**</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td></td>
<td>0.6073**</td>
<td>0.5381**</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td>1.00</td>
<td></td>
<td>0.6485**</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**p < .005**
variables such as size and gender identification, 100% of the pair-wise comparisons fell in the more desirable range of greater than or equal to .60. Table 29 presents a summary of inter-rater reliabilities.

Tables 23-28 show the inter-rater reliability coefficients for the DAP Self-Esteem scores for all possible 36 pair-wise rater comparisons, representing the strength of scorer agreement. The rater agreements ranged from .23-.95. All inter-rater pair-wise correlations were statistically significant at the .005 level, with the exception of two rater comparisons which were significant at the .05 level. (Table 22, Judges 1 and 3, Quality variable; Judges 3 and 4, Quality Variable.) In addition, approximately 28% of the inter-rater comparisons fell in the more desirable range of correlations greater than or equal to .60.

Table 29 presents a summary of the DAP inter-rater reliability coefficients by scoring variable. All four judges, on all but two (Self-Concept Age and Placement) of the 12 scoring variables achieved statistically significant agreement. On the variables of Self-Concept Placement and Age, 83% of the pair-wise comparisons were significant at the .05 level. One hundred percent of the pair-wise comparisons for 3 of the 12 scoring variables (Self-Concept Size, and Self-Esteem Size and Composite Score) occurred greater than or equal to .60.

Of particular interest were the reliability coefficients for the Composite scores, as in the scoring system employed in the
### TABLE 23

**DAP SELF-ESTEEM INTER-RATER RELIABILITY (TETRACHORIC CORRELATION) BY JUDGE**

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.9230**</td>
<td>.9478**</td>
<td>.8956**</td>
</tr>
<tr>
<td>Judge 2</td>
<td></td>
<td>1.00</td>
<td>.9230**</td>
<td>.8750**</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td></td>
<td>1.00</td>
<td>.9478**</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

*<p><.05  
**<p><.005

### TABLE 24

**DAP SELF-ESTEEM INTER-RATER RELIABILITY (TETRACHORIC CORRELATION) BY JUDGE**

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.6370**</td>
<td>.2255*</td>
<td>.6147**</td>
</tr>
<tr>
<td>Judge 2</td>
<td></td>
<td>1.00</td>
<td>.4781**</td>
<td>.5864**</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td></td>
<td>1.00</td>
<td>.2433*</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

*<p><.05  
**<p><.005
### Table 25
DAP Self-Esteem Inter-rater Reliability (Tetrachoric Correlation) by Judge Gender Identification Variable

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>0.7103**</td>
<td>0.2308*</td>
<td>0.6422**</td>
</tr>
<tr>
<td>Judge 2</td>
<td></td>
<td>1.00</td>
<td>0.4901**</td>
<td>0.4876**</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.2269*</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05  
**p < .005

### Table 26
DAP Self-Esteem Inter-rater Reliability (Tetrachoric Correlation) by Judge DAP Placement Variable

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>0.5687**</td>
<td>0.6826**</td>
<td>0.7641**</td>
</tr>
<tr>
<td>Judge 2</td>
<td></td>
<td>1.00</td>
<td>0.5705**</td>
<td>0.6360**</td>
</tr>
<tr>
<td>Judge 3</td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.6989**</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

**p < .005
### Table 27
DAP Self-Esteem Inter-Rater Reliability (Tetrachoric Correlation) by Judge
DAP Omissions Variable

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.5622**</td>
<td>.4879**</td>
<td>.5351**</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>.6135**</td>
<td>.4646**</td>
<td></td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>.5666**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**p < .005

### Table 28
DAP Self-Esteem Inter-Rater Reliability (Pearson's Product Moment Correlations) by Judge
DAP Composite Score

<table>
<thead>
<tr>
<th></th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
<th>Judge 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.7527**</td>
<td>.7616**</td>
<td>.7439**</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>.7694**</td>
<td>.6928**</td>
<td></td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td></td>
<td>.6685**</td>
<td></td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**p < .005
<table>
<thead>
<tr>
<th>Scoring Variable</th>
<th>Percent r p &lt; .05</th>
<th>Percent r ≥ .60</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self Concept</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>100</td>
<td>100</td>
<td>.63 - .79</td>
</tr>
<tr>
<td>Placement</td>
<td>83</td>
<td>0</td>
<td>.20 - .54</td>
</tr>
<tr>
<td>Gender Ident.</td>
<td>100</td>
<td>83</td>
<td>.57 - .71</td>
</tr>
<tr>
<td>Affect</td>
<td>100</td>
<td>0</td>
<td>.30 - .53</td>
</tr>
<tr>
<td>Age</td>
<td>83</td>
<td>0</td>
<td>.20 - .50</td>
</tr>
<tr>
<td>Composite Score</td>
<td>100</td>
<td>50</td>
<td>.52 - .64</td>
</tr>
<tr>
<td><strong>Self Esteem</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>100</td>
<td>100</td>
<td>.88 - .95</td>
</tr>
<tr>
<td>Quality</td>
<td>100</td>
<td>33</td>
<td>.23 - .64</td>
</tr>
<tr>
<td>Gender Ident.</td>
<td>100</td>
<td>33</td>
<td>.23 - .64</td>
</tr>
<tr>
<td>Placement</td>
<td>100</td>
<td>66</td>
<td>.57 - .76</td>
</tr>
<tr>
<td>Omissions</td>
<td>100</td>
<td>16</td>
<td>.49 - .61</td>
</tr>
<tr>
<td>Composite Score</td>
<td>100</td>
<td>100</td>
<td>.67 - .77</td>
</tr>
</tbody>
</table>
present study overall DAP scores were considered to be more reflective of the constructs than individual scoring variables. Table 29 shows that 100% of the pair-wise comparisons for DAP Self-Concept Composite scores were statistically significant and at least 50% were .60 and above. One hundred percent of the DAP Self-Esteem Composite score inter-rater reliabilities were both statistically significant at the .05 level and .60 and above. It can be concluded that, in general, DAP scoring agreement between the four judges was fairly good. On several specific scoring variables, however, such as self-concept placement, affect, age and self-esteem omissions, judges were able to achieve statistically significant agreement, but did not achieve reliability greater than or equal to .60. Therefore, for Research Question Three it can be concluded that judges were able to reach scorer agreement beyond chance.

Research Question Four

Will judges using the DAP Scoring System exhibit differential scoring agreement based upon level of criterion measure group?

To answer this research question, inter-rater reliability was examined for high, medium, low levels of self-concept and self-esteem. The same group placement procedure was followed as described for Research Question Two. Subjects were placed into one of three groups (high, medium and low) for each of the criterion measures. Subject placement was determined by
criterion measure score (high = + 1 SD above mean, medium = within 1 SD above and below the mean, low = -1 SD below the mean). DAP inter-rater reliability correlations from each group were examined to determine whether level of criterion measure was related to inter-rater reliability.

Tables 30, 31, 32 present the DAP Self-Concept inter-rater reliability for low, medium, high levels of the criterion groups respectively. Table 30 shows the 36 inter-rater reliability correlations for the low level criterion group. Correlations ranged from .32 to .91. All but two correlations (Judges 2 and 3, age variable; Judges 1 and 3, gender identification variable) of the 36 correlations were significant at the .05 level and 64% of the pair-wise comparisons were significant at the .005 level. Sixty-nine percent of the inter-rater reliabilities fell into the more desirable range of .60 and above.

Table 31 shows the DAP inter-rater reliability coefficients for the medium level criterion group. All but three of the 36 pair-wise comparisons (Judges 2 and 4, placement variable; Judges 1 and 3, Judges 3 and 4, placement variable) were significant at the .05 level. Seventy-eight percent of the pair-wise comparisons were significant at the .005 level. At least 50% of the comparisons fell in the more desirable range of .60 and above.

Table 32 provides the DAP inter-rater reliability coefficients for the high level criterion group. Correlations
TABLE 30

DAP SELF-CONCEPT INTER-RATER RELIABILITY (TETRACHORIC CORRELATION)
FOR LOW SELF-CONCEPT CRITERION GROUP BY DAP VARIABLE AND JUDGE
N=13

<table>
<thead>
<tr>
<th></th>
<th>Size Variable</th>
<th></th>
<th>Affect Variable</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Judge</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.6860**</td>
<td>.6860**</td>
<td>.6860**</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Placement Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.5968**</td>
<td>.6364*</td>
<td>.4264</td>
<td>1.00</td>
<td>.5635*</td>
<td>.6614**</td>
<td>.8619**</td>
<td></td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>1.00</td>
<td>.7977**</td>
<td>.5345*</td>
<td>1.00</td>
<td>.1754</td>
<td>.4462</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>1.00</td>
<td>.6701**</td>
<td>1.00</td>
<td>.5701*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Gender Identification</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>.8367**</td>
<td>.3162*</td>
<td>1.00*</td>
<td>1.00</td>
<td>.8409**</td>
<td>.7849**</td>
<td>.9047**</td>
<td></td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>1.00</td>
<td>.6614*</td>
<td>.8367**</td>
<td>1.00</td>
<td>.6531**</td>
<td>.7201**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>1.00</td>
<td>.3162*</td>
<td>1.00</td>
<td>.7640**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
**p < .005
<table>
<thead>
<tr>
<th>Size Variable</th>
<th>Affect Variable</th>
<th>Placement Variable</th>
<th>Age Variable</th>
<th>Gender Identification</th>
<th>Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>0.9629**</td>
<td>0.9629**</td>
<td>0.9629**</td>
<td>0.9629**</td>
<td>0.9629**</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 4</td>
<td>0.9629**</td>
<td>0.9629**</td>
<td>0.9629**</td>
<td>0.9629**</td>
<td>0.9629**</td>
</tr>
</tbody>
</table>

*p < 0.05

**p < 0.005

TABLE 31

DAP SELF-CONCEPT INTER-RATER RELIABILITY (TETRACHORIC CORRELATION)
FOR MEDIUM SELF-CONCEPT CRITERION GROUP BY DAP VARIABLE AND JUDGE
N=69
<table>
<thead>
<tr>
<th>Size Variable</th>
<th>Affect Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge</td>
<td>1</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Placement Variable</th>
<th>Age Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge</td>
<td>1</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender Identification</th>
<th>Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge</td>
<td>1</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05  
**p < .005
range from .32 to 1.00. Seventy-two percent of the correlations were significant at the .05 level. Forty-eight percent were significant at the .005 level. Fifty-five percent of the correlations fell into the .60 and above range.

Tables 33, 34, 35 contain the DAP Self-Esteem reliabilities for low, medium, high levels of the criterion group respectively. Table 33 shows the inter-rater reliability correlations for the low level Self-Esteem criterion group. Correlations range from .11 to 1.00. Thirty-six percent of the correlations achieved statistical significance and 19% of the correlations fell in the desirable range of .60 and above. Table 34 shows the inter-rater reliabilities for the medium criterion group. Thirty of the 36 correlations (83%) were statistically significant at the .05 level. Thirty-six percent of the correlations were .60 and above. Table 35 shows the DAP inter-rater reliabilities for the high criterion group. Sixty-one percent were statistically significant, and 44% were .60 and above.

Tables 36 and 37 present a summary of inter-rater reliability coefficients by level of criterion group. Of particular interest in this study are Composite Score inter-rater reliabilities. As can be seen in Table 33, the DAP Self-Concept Composite Score inter-rater reliabilities for the low and medium groups were all statistically significant at the .05 and occur in the more desirable .60 and above range. DAP inter-rater reliabilities, however, for the high level criterion group were
<table>
<thead>
<tr>
<th>Size Variable</th>
<th>Placement Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00 **</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00 **</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00 **</td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00 **</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality Variable</th>
<th>Omissions Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00 **</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00 **</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00 **</td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00 **</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sexual Identification</th>
<th>Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00 **</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00 **</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00 **</td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00 **</td>
</tr>
</tbody>
</table>

*p < .05

**p < .005
<table>
<thead>
<tr>
<th></th>
<th>Size Variable</th>
<th>Placement Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>0.7328**</td>
<td>.3305**</td>
</tr>
<tr>
<td>Judge 3</td>
<td>0.6769**</td>
<td>.4229**</td>
</tr>
<tr>
<td>Judge 4</td>
<td>0.6253**</td>
<td>.4605**</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>.8720**</td>
<td>.2359*</td>
</tr>
<tr>
<td>Judge 3</td>
<td>.7597**</td>
<td>.3484**</td>
</tr>
<tr>
<td>Judge 4</td>
<td>.8118**</td>
<td>.3101*</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Quality Variable</th>
<th>Omissions Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>0.2445</td>
<td>0.2357</td>
</tr>
<tr>
<td>Judge 3</td>
<td>0.4179**</td>
<td>0.4669**</td>
</tr>
<tr>
<td>Judge 4</td>
<td>0.1480</td>
<td>0.2278</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>0.4484**</td>
<td>0.3754**</td>
</tr>
<tr>
<td>Judge 3</td>
<td>0.5942**</td>
<td>0.0425</td>
</tr>
<tr>
<td>Judge 4</td>
<td>0.4781**</td>
<td>0.1983</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 4</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sexual Identification</th>
<th>Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>0.7798**</td>
<td>0.6206**</td>
</tr>
<tr>
<td>Judge 3</td>
<td>0.5980**</td>
<td>0.6135**</td>
</tr>
<tr>
<td>Judge 4</td>
<td>0.5980**</td>
<td>0.4928**</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Judge 2</td>
<td>0.6086**</td>
<td>0.6430**</td>
</tr>
<tr>
<td>Judge 3</td>
<td>0.6086**</td>
<td>0.5741**</td>
</tr>
<tr>
<td>Judge 4</td>
<td>0.5779**</td>
<td>0.6304**</td>
</tr>
</tbody>
</table>

*P < .05  
**P < .005
### Table 35

**DAP Self-Esteem Inter-rater Reliability (Tetrachoric Correlation)**

For High Self-Esteem Criterion Group by DAP Variable and Judge

N=16

<table>
<thead>
<tr>
<th></th>
<th>Size Variable</th>
<th>Placement Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Judge 1</td>
<td>Judge 2</td>
</tr>
<tr>
<td>1</td>
<td>1.00</td>
<td>.3578</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>.3133</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>.3133</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Quality Variable</th>
<th>Omissions Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Judge 1</td>
<td>Judge 2</td>
</tr>
<tr>
<td>1</td>
<td>1.00</td>
<td>.2381</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>.6831**</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>.8704**</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sexual Identification</th>
<th>Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Judge 1</td>
<td>Judge 2</td>
</tr>
<tr>
<td>1</td>
<td>1.00</td>
<td>.4286</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>1.0000**</td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td>.4286</td>
</tr>
<tr>
<td>Judge 4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*P < .05

**P < .005
<table>
<thead>
<tr>
<th>Scoring Variable</th>
<th>N=18 Low Criterion Group</th>
<th>N=69 Medium Criterion Group</th>
<th>N=13 High Criterion Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%r</td>
<td>r</td>
<td>Range</td>
</tr>
<tr>
<td>Self Concept</td>
<td>p&lt;.05</td>
<td>≥.60</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>100</td>
<td>100</td>
<td>.69-1.00</td>
</tr>
<tr>
<td>Placement</td>
<td>83</td>
<td>66</td>
<td>.43-.80</td>
</tr>
<tr>
<td>Gender Ident.</td>
<td>66</td>
<td>66</td>
<td>.32-.84</td>
</tr>
<tr>
<td>Affect</td>
<td>100</td>
<td>66</td>
<td>.56-.78</td>
</tr>
<tr>
<td>Age</td>
<td>66</td>
<td>33</td>
<td>.18-.86</td>
</tr>
<tr>
<td>Composite Score</td>
<td>100</td>
<td>100</td>
<td>.78-.90</td>
</tr>
<tr>
<td>Scoring Variable</td>
<td>N=18 Low Criterion Group</td>
<td>N=66 Medium Criterion Group</td>
<td>N=16 High Criterion Group</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td></td>
<td>( r )  ( \geq .60 )</td>
<td>( r )  ( \geq .60 )</td>
<td>( r )  ( \geq .60 )</td>
</tr>
<tr>
<td>Self Esteem</td>
<td>( p &lt; .05 ) 100  .78-.1.00</td>
<td>( p &lt; .05 ) 100  .63-.87</td>
<td>33  .31-.76</td>
</tr>
<tr>
<td>Size</td>
<td>16  .03-.60</td>
<td>66  .14-.59</td>
<td>66  .24-.87</td>
</tr>
<tr>
<td>Quality</td>
<td>0  .08-.40</td>
<td>100  .60-.78</td>
<td>33  .43-1.00</td>
</tr>
<tr>
<td>Gender Ident.</td>
<td>0  .08-.40</td>
<td>100  .23-.46</td>
<td>50  .31-.86</td>
</tr>
<tr>
<td>Placement</td>
<td>33  .11-.62</td>
<td>33  .20-.47</td>
<td>100  .75-.88</td>
</tr>
<tr>
<td>Omissions</td>
<td>66  .43-.60</td>
<td>100  .49-.64</td>
<td>83  .42-.85</td>
</tr>
<tr>
<td>Composite Score</td>
<td>16  .43-.60</td>
<td>66  .49-.64</td>
<td>50  .42-.85</td>
</tr>
</tbody>
</table>
lower, with only 83% of the pair-wise comparisons achieving statistical significance and 66% of the correlations .60 and above.

Table 37 indicates that for the low criterion group DAP Self-Esteem inter-rater reliabilities Composite score, 66% were statistically significant, 16% of the correlations .60 and above. The medium group composite score reliabilities were 100% statistically significant, 66% of the correlations .60 and above. In the high group pair-wise comparisons 83% were significant, while 50% of the correlations .60 and above. Mean correlations for each level of criterion group were computed, then tested to determine statistical difference via Fisher's Zr transformation test. Table 38 presents the results of this analysis. As indicated, none of the inter-rater reliability means were statistically different at the .05 level or beyond. Thus, it can be concluded for Research Question Four there appears to be no difference between inter-rater agreement by level of criterion group.

Discussion

Research Question One:

Is there a relationship between the DAP scoring system measures of self-concept and self-esteem and respective criterion measures?

Individuals with good self-concept, as defined by the Miskimen's SGODS, are those who describe their real and ideal
### Table 38

**Fisher's Z r Transformation Test for Differences Among DAP Inter-Rater Reliability Correlation Means for DAP Composite Scores**

<table>
<thead>
<tr>
<th>Group</th>
<th>Composite Score Mean</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-Concept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low $M_{r1}$</td>
<td>77.5</td>
<td>.2310</td>
<td>.8165</td>
<td></td>
</tr>
<tr>
<td>Medium $M_{r2}$</td>
<td>72.5</td>
<td></td>
<td>.8163</td>
<td></td>
</tr>
<tr>
<td>High $M_{r3}$</td>
<td>66.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Self-Esteem</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low $M_{r1}$</td>
<td>51.5</td>
<td>.3775</td>
<td>.4963</td>
<td></td>
</tr>
<tr>
<td>Medium $M_{r2}$</td>
<td>59.5</td>
<td></td>
<td>.2536</td>
<td></td>
</tr>
<tr>
<td>High $M_{r3}$</td>
<td>64.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
self in a congruent way. It was hypothesized that individuals with good self-concept would also draw their real and ideal self in a congruent way, as reflected in the DAP in terms of size, placement, gender identification, affect and age. An individual with good self-concept was expected to draw the real and ideal self in a similar way on these variables.

The data from this study did not support these expectations; the link between self-concept and drawing behavior was not supported. There were no statistically significant correlations between the judges' DAP scores and respective criterion measures, with one exception. One judge's (Judge 4) DAP scores indicated a statistically significant relationship between DAP measured gender identification and criterion measure of self-concept ($r=.24, p<.05$). As expected, females with higher levels of congruency between real and ideal self as measured by the Miskimin's SCODS drew the real and ideal self the same sex. It is difficult to make direct comparisons of this finding to the research literature as no other study has examined this variable in a way similar to this study. However, when examining the question of which gender an individual draws when drawing the ideal self, the literature seems to suggest several trends. Men tend to draw the same sex when drawing the ideal self, women are less likely than men to do so (Culí and Hardy, 1971; Fisher, 1968).
Self-esteem, as defined by Rosenberg's SEI, is defined as overall self-regard, and an individual with high self-esteem was expected to draw the real self with the highest overall quality. It was expected, then, that persons with high self-esteem would draw the real self about 6-7 inches in height, with highest quality, exhibiting secondary sexual characteristics, drawing in the middle of the page, and with no major omission of body parts.

The data from this study did not support these expectations, with several exceptions. One judge's (Judge 1) DAP scores demonstrated a statistically significant relationship between self-esteem and size of drawing. Persons with good self-esteem did draw the self approximately 6-7 inches in height ($r=.21$, $p<.05$). While DAP theory holds that size of drawing is related to self-esteem level, the literature is mixed, with most studies failing to yield a relationship between size of drawing and self-esteem (Baugh & Prytula, 1974; Hjorth & Harway, 1981). However, most of the studies that have not yielded a relationship between size of drawing and level of self-esteem were based on child or adolescent populations. Drawing behavior may reflect self-esteem differently depending upon age of drawer.

Additionally, one judge's (Judge 1) DAP self-esteem Composite score was significantly positively correlated ($r=.23$, $p<.05$) with the Rosenberg SEI, indicating a statistically significant relationship between combined DAP self-esteem scores and criterion measure. This finding is similar to the bulk of
current DAP research, which seems to suggest that using a combination of multiple variables or global ratings are more successful in predicting criterion measures than individual signs (Kahill, 1984).

It should be noted that the above statistically significant correlations were only achieved by one of four judges. Additionally, it is important to note that these two significant correlations occurred in the .20 range. For purposes of validity, coefficient ranges of .20 to .30 show only a slight relationship between variables. Correlations in this range may have meaning in exploratory research (Borg & Gall, 1983) when relationships are being investigated using crude measures, but are of little value in practical prediction situations. In summary, the validity of the DAP as a measure of self-esteem and self-concept was not supported by this study.

Research Question Two:

Can the DAP scoring system discriminate between various levels of the criterion measures?

It was expected that judges may have more success in predicting criterion scores when examining high, medium or low self-concept and self-esteem groups. Validity correlations by criterion group level ranged from .01 to .43. The data gathered to address this research question failed to achieve statistical significance. The expected relationships between drawing
behavior and the constructs did not reveal itself even when DAP validity coefficients were examined by various levels of criterion measures. Comparisons to existing literature can not be made as no known study appears to have examined data in a way similar to this study.

**Research Question Three:**

Will the raters, independently utilizing the rating system, demonstrate agreement?

Of particular importance when examining projective measures, because scoring is subjective, is the amount of inter-rater agreement. An examination of the data concerning the strength of rater agreement revealed that for the DAP self-concept Composite scores, 100% of the 36 pair-wise comparisons were significant at the .05 level and 50% of the correlations were equal to or greater than .60. In terms of the DAP self-esteem composite scores, 100% of the pair-wise comparisons were statistically significant at the .05 level and 100% of the correlations were equal to or greater than .60. Thus while the DAP with this scoring method was subjectively scored, it appears that it was quantified in such a way as to achieve acceptable levels of inter-rater reliability. These findings coincide with the more recent body of DAP literature. In the past 15 years, there has been a concerted effort on the part of DAP researchers to objectify and standardize DAP rating procedures (Kahill, 1984).
Research Question Four:

Will judges using the DAP scoring system exhibit differential scoring agreement based upon level of criterion measure group?

Inter-rater reliabilities for high, medium and low levels of the criterion group were examined. For both DAP self-concept and self-esteem scores, the majority of inter-rater reliabilities across groups were statistically significant. Seventy-five percent of all pair-wise comparisons were significant at the .05 level. Statistically there was no significant difference between the various groups in terms of inter-rater reliability, meaning that rater agreement did not differ when examining high or low levels of self-concept and self-esteem. No other known study has examined inter-rater reliability by criterion group level, preventing comparisons to existing literature.

Summary

A summary of the findings are presented below. The summary is reported by demographic variables and research question.

Findings by Demographic Variable:

Forty percent of the research study subjects reported a married marital status and thirty-four percent of subjects reported an occupation of skilled trade. Sixty-three percent of subjects reported possessing at least 16 years of education and 60 percent of subjects indicated they were 20-35 years of age.
Last, 90 percent of subjects reported incomes of $50,000 or below.

Findings by Research Question

Research Question One:

Is there a relationship between the DAP scoring system measures of self-concept and self-esteem and respective criterion measures?

Examining 100 DAP protocols, DAP Self-Concept and DAP Self-Esteem scores were obtained by four independent judges using the DAP scoring system developed for this study, consisting of 10 variables and 2 composite scores. Each judge's 12 DAP Self-Concept and Self-Esteem scores for each of the 100 DAP protocols were correlated with respective criterion measure (Miskimin's SGODS and Rosenberg SEI) and examined to determine statistical significance at the .05 level, as well as to determine more practical significance in which correlations were greater than or equal to .60. Only 4 of the 48 validity coefficients (8%) achieved statistical significance at the .05 level: DAP Self-Concept variables size (r=-.22)(Judge 4) and gender identification (r=.24)(Judge 4); and DAP Self-Esteem variables size (r=.21)(Judge 1) and Composite score (r=.23)(Judge 1.)

These few significant correlations were achieved by two of the four judges, and occurred only in the .20 range, indicating only a weak relationship between DAP scores and respective
criterion measures. The validity coefficients for Composite scores were of particular interest, as it was hoped that overall DAP scores would exhibit a positive correlation with criterion measures. The data revealed that for the DAP Self-Concept Composite score, none of the four judges' scores were statistically significant. This indicates that there was no statistically significant relationship between the DAP Self-Concept scoring system and its criterion measure, the Miskimin's SGODS.

In terms of the DAP Self-Esteem Composite scores, only one judge's DAP scores correlated significantly at the .05 level with the criterion measure, the Rosenberg SEI. However, while statistically significant, the validity coefficient of \( r = .23 \) \( (p < .05) \) indicates only a very weak relationship between measures.

Thus, in terms of DAP scoring system validity, it can be concluded for Research Question One that there exists little, if any, relationship between the DAP scoring system and respective criterion measures. The data fails to support a significant relationship between the DAP scoring system and independent measures of self-concept and self-esteem.

Research Question Two:

Can the DAP scoring system discriminate between various levels of the criterion measures?
The researcher examined whether the DAP scoring system would produce higher validities for various levels of the criterion measures. Is the DAP scoring system more sensitive with high or low levels of self-esteem/self-concept as measured by the criterion instruments?

To this end, subjects were placed into one of three groups (high, medium, low) for each criterion measure. Placement was determined by subject's z scores on criterion measures. DAP scores were then correlated with criterion measures to determine differential validity. Correlations were obtained for each DAP scoring variable and composite score by judge. Only one of the 144 correlation coefficients was statistically significant. One judge's (Judge 2) DAP Self-Concept Composite score was negatively correlated with criterion measure ($r = -.55$, $p < .05$). Contrary to expectation, as subjects in high criterion group achieved higher scores on Miskimin's SGODS, they received low DAP Self-Concept scores from one judge.

The data analysis failed to support the DAP scoring system's differential validity. The scoring system was equally ineffective no matter whether examining subjects with high or low levels of self-esteem or self-concept. Therefore, it can be concluded for Research Question Two that level of criterion measure did not influence DAP scoring system validity.
Research Question Three:

Will the four raters, independently utilizing the author's scoring system, demonstrate agreement?

To determine accuracy of scoring, correlation coefficients were calculated between judges. Reliability coefficients were examined to determine statistical significance at the .05 level and to determine practical significance of correlations greater than or equal to .60. Seventy of the 72 pair-wise comparisons were significant at the .05 level, 65 of the 72 were significant at the .005 level, 35 of the 72 correlations were equal to or greater than .60. All four judges in all but two (Self-Concept Placement and Age) of the 12 scoring variables were able to achieve statistically significant agreement at the .05 level. On 5 of the 12 variables, at least 50% of the pair-wise correlations were greater than or equal to .60. For three variable categories (Self-Concept Size, Affect, Age) inter-rater agreement was statistically significant but did not achieve practical significance.

Examination of the data revealed that inter-rater reliability of the DAP scoring system was adequate. That is, in general, the judges achieved adequate levels of agreement in scoring judgment, with the exception of several categories as mentioned above.

Judges achieved adequate inter-rater reliability for Composite scores, with 100% of judges reaching agreement on the
DAP Self-Concept Composite score, 50% achieving correlations greater than or equal to .60. In terms of Self-Esteem Composite score, 100% of judges achieved inter-rater reliability coefficients which were significant at the .05 level and equal to or greater than .60. It can be concluded for Research Question Three that inter-rater reliability for the DAP scoring system developed for this study was adequate.

**Research Question Four:**

Will judges using the DAP scoring system exhibit differential scoring agreement based upon level of criterion measure group?

DAP inter-rater reliability coefficients were examined for three levels for each of the two criterion measures. Subjects were placed into one of three groups for each criterion measure (high, medium and low based upon criterion measure scores; high = + 1 SD above mean, medium = 1 SD above/below mean, low = 1 SD below mean). Inter-rater reliabilities were computed for each of the resultant six groups to determine whether level of self-esteem and self-concept was related to inter-rater reliability. Of particular interest in this study were composite score reliabilities.

For DAP Self-Concept inter-rater reliability, 100% of the pair-wise comparisons for both low and medium groups were statistically significant at the .05 level and were equal to or greater than .60. Eighty-three percent of the high criterion
group pair-wise comparisons were statistically significant, while 66% of the comparisons fell in the \( r=.60+ \) range. Sixty-six percent of the DAP Self-Esteem Composite scores for low criterion group were significant at the \( .05 \) level, 16% of the correlations were equal to or greater than \( .60 \). One hundred percent of the medium group were significant at the \( .05 \) level, 66% at the \( r=.60+ \) level. For the high group, 83% were significant at the \( .05 \) level, while 50% of the comparisons achieved an \( r \) of \( .60 \) or higher. Mean composite score correlations for each group were computed and tested to determine statistical difference via Fisher's \( z \) \( r \) transformation test for independent samples. None of pair-wise comparisons for composite scores were significant at the \( .05 \) level. Thus, for Research Question Four it can be concluded that no statistical differences exist between the judges' inter-score agreement when judging high, medium or low levels of self-esteem and self-concept. Table 36 contains a summary of inter-rater reliability differences by level of criterion measure.
CHAPTER V
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This chapter contains a summary of the study. In addition, conclusions and recommendations for future research are provided.

Summary

The Draw-A-Person Test (DAP) is one of the most widely used personality instruments in the United States. Despite its popularity in clinical usage, there appears to be little certainty as to what the test actually measures. The DAP is not a frequently researched instrument, and the research can be characterized as inconsistent and not supportive of the validity of the DAP.

The purpose of this study was to validate the DAP as a measure of self-concept and self-esteem in outpatient females at a community mental health center. Using a construct validation design, a pilot study (N=41) was conducted in which hypotheses were developed to account for differential drawing performance. These hypotheses and certain basic assumptions about drawing behavior were then used to develop a scoring system for the DAP to separately measure self-esteem and self-concept.

127
The scoring system, consisting of 12 variables, was then used by four judges making independent scoring judgments to score 300 drawings from 100 subjects. Three of the judges were Ph.D. clinical psychologists with at least 10 years clinical experience and training in assessment. The researcher, a doctoral level candidate in Counseling at The Ohio State University, served as the fourth rater and possessed 11 years clinical experience.

The DAP scores obtained by the four judges were correlated with external objective measures of self-concept and self-esteem to assess the validity of the DAP scoring system. Inter-rater reliability coefficients were also determined to assess the strength of judge agreement.

One hundred adult, female volunteer subjects were recruited by 10 therapists from patient caseloads at a community mental health center in a suburb of a large midwestern city. Females were chosen as subjects because the literature suggests that gender influences drawing performance and that women, more so than men, express self concept via drawing behavior. Subjects included only individuals who were estimated by referring therapists to possess adequate operating intelligence, mental status and motoric skill to complete the drawing tasks. Subjects were administered three instruments in the following order: the DAP, the Rosenberg Self-Esteem Inventory (SEI) and the Miskimin's Self-Goal-Other Discrepancy Scale (SGODS). The latter two instruments were the criterion measures.
The DAP is a projective personality measure in which the client is requested to draw three separate human figure drawings which theoretically represent various aspects of the self. The Rosenberg SEI is considered to be one of the best unidimensional measures of self-esteem (Demo, 1985). The Miskimin's SCODS is an instrument which measures the self-concept, specifically congruency between real and ideal self.

This study was designed to improve upon previous research by the following:

1. Utilization of a more appropriate research design which focuses upon developing a set of related hypotheses underlying the test rather than simply correlating long lists of drawing variables with criterion measures.

2. This study was one of the few known studies to examine the constructs of self-concept and self-esteem. To date, most DAP researchers confuse the two terms, making no clear distinction between the two constructs and have not linked the DAP studies theoretically with the current body of self-esteem/self-concept research.

3. Development of a clearly detailed scoring system which would also be of practical use in a clinical setting.

4. Taking adequate steps to ensure internal validity by controlling for possible confounds. This was accomplished by holding potential confounds consistent by examining only one level of extraneous variables.
5. This is believed to have been the first study which was designed to examine the inter-relationships among an individual's drawings rather than examine each drawing in isolation. In summary, this study, then, more so than previous research, was designed to reflect the way in which the DAP is actually used in the clinical practice.

Findings

Four research questions were the focus of this study. The findings are summarized below by demographic variables and research question.

Demographic Variables

By self-report, two-fifths of the adult female subjects indicated they were married; one-third employed in the skilled trades. One-third of the subjects indicated they were 25-34 years of age and two-thirds of the subjects possessed at least 15 years of education. Five-sixths of the subjects reported incomes of $2,000-$35,000.

Research Questions:
Research Question One: Is there a relationship between the DAP scoring system measure of self-concept and self-esteem and respective criterion measures?

The DAP Self-Concept and DAP Self-Esteem scores obtained by four independent judges using the developed DAP scoring system were correlated with their respective criterion measures (Miskimin's SGODS and Rosenberg SEI). DAP Drawing Variables were
analyzed with the biserial correlation and the DAP Composite Scores with the Pearson's product moment correlation to determine statistical significance at the .05 level.

The 48 resulting correlations ranged from .03 to .23, and four were statistically significant at the .05 level. The DAP self-concept drawing variable of size was negatively correlated with criterion measure score \( r_{bis} = -.22, \ p < .05 \) (Judge 4). The DAP Self-Concept drawing variable gender identification was positively correlated \( r_{bis} = .24, \ p < .05 \) (Judge 4) with criterion measure scores. The DAP self-esteem drawing variable of size \( r_{bis} = .21, \ p < .05 \) (Judge 1) was positively correlated with criterion measure scores. The DAP Self-Esteem composite score \( r = .23, \ p < .05 \) (Judge 1) was positively correlated with criterion measure scores. There appears to be no statistically significant relationship between the DAP scoring system used in this study and respective criterion measures.

**Research Question Two:** Can the DAP scoring system discriminate between various levels of the criterion measures?

The DAP self-concept and DAP self-esteem scores obtained by four independent judges using the developed DAP scoring system were correlated with three levels (high, medium and low) of respective criterion measures. DAP drawing variables were analyzed with the biserial correlation and the DAP composite scores with the Pearson's product moment correlation to determine
statistical significance at the .05 level. The 72 pair-wise comparisons ranged from .01 to .55.

Only one out of 72 correlations were statistically significant at the .05 level. One judge's DAP self-concept composite score ($r=-.55$, $p<.05$) was negatively correlated with criterion measure scores. Therefore, there appears to be no statistically significant relationship between the DAP scoring system and high, medium and low levels of self-concept and self-esteem.

Research Question Three: Will the judges, independently utilizing the rating system, demonstrate agreement?

To determine inter-rater reliability, or strength of judges' scoring agreement, correlation coefficients were calculated for all pair-wise comparisons. All DAP scores were analyzed with the tetrachoric correlation ($r_t$), except for DAP composite scores which were analyzed with the Pearson's product moment correlation ($r$). The obtained correlations were examined to determine statistical significance at the .05 level, and practical significance which was set at greater than or equal to .60.

The 72 pair-wise rater comparison correlations ranged from .20 to .95. Seventy-one of the 72 pair-wise comparisons (99%) were statistically significant at the .05 level. Seventy of the 72 comparisons (97%) were significant at the .005 level.

Thirty-eight percent of the DAP self-concept inter-rater reliability correlations were greater than or equal to .60.
Twenty-eight percent of the DAP self-esteem reliability correlations were greater than or equal to .60. One hundred percent of the pair-wise comparisons for DAP self-concept and self-esteem composite reliabilities were statistically significant at .05 level with 50% of the DAP self-concept composite reliabilities and 100% of the DAP self-esteem composite reliabilities equal to or greater than .60. It can be concluded that DAP scoring agreement between judges was adequate.

Research Question Four: Will judges using the DAP Scoring System exhibit differential scoring agreement based upon level of criterion measure group?

DAP Scoring System inter-rater reliability was examined for three levels (high, medium, low) of each criterion measure. Tetrachoric correlational analysis was utilized for DAP Drawing Variables and Pearson's Product Moment Correlation for DAP composite scores.

The resultant pair-wise comparisons ranged from correlations of .32 to 1.00. Seventy-five percent of all pair-wise comparisons were significant at the .05 level. Fifty-five percent of the correlations were greater than or equal to .60. Mean Composite Score correlations for each level of criterion group were computed and analyzed via Fisher's ZR transformation test for independent samples. None of the pair-wise comparisons were significant at the .05 level, indicating no statistical difference between the judges' scoring agreement by level of
criterion measure. Therefore, it can be concluded that there is no significant difference between inter-rater agreement by level of criterion group.

Conclusions

The following section offers conclusions based upon the findings in light of the theoretical framework. Conclusions will be related to the four hypotheses addressed in the study.

Research Question One

Is there a relationship between the DAP scoring system and the criterion measures of self-esteem and self-concept?

The results of the study failed to yield any significant relationship between the DAP scoring system and objective outside measures of self-concept and self-esteem in adult female subjects recruited from an outpatient mental health clinic. The DAP as reflected by the scoring system utilized in the present study did not support DAP validity as a measure of self-concept and self-esteem in outpatient females. The low, non-significant correlation coefficients resulted in little certainty as to what the DAP scores actually mean. A number of possible explanations for the inconclusive findings are enumerated below:

1. Human figure drawings may not in actuality be a valid measure of self-concept. The DAP as a personality measure may be quite meaningless; drawing behavior may have no relationship with personality variables.
2. The DAP may have some connection to self-concept, but because the idea of self-concept projection is so complex and varies so much from individual to individual, it is impossible to quantify and predict. Figure drawings as personality instruments may not be researchable. The DAP may not be amenable to simplification and operationalization.

3. The difficulty may not be with the DAP itself but a reflection of the fact that the understanding of what is meant by the term self and how it is evaluated is only at a very formative stage. Indications at this point from the self literature (Wylie, 1974; Shavelson et al, 1982) suggest that the construct is complex and multifaceted. Thus, in terms of the present research, it is possible that the DAP does measure the self in some way but differently than the criterion measures utilized in this study. It may well be that until the self body of literature is more substantial, the role, if any, of the projective drawings in its measure will be unknown.

4. The inconclusive findings in this study may be due to problems with its research design. It may be that the DAP did not measure the constructs because of a volunteer, female, outpatient population which biased the results of the study with some unknown confound. For example, level and specific type of psychopathology may impact upon drawing behavior in such a way as to confound the measurement of self-concept or self-esteem. For example, a subject with high self-esteem who is depressed may
draw differently than a non-depressed high self-esteem subject. Utilizing a more homogenous or normal population may have resulted in different findings. Future studies may benefit from building type of psychopathology into the research design.

Additionally, because the sample consisted of volunteers, as well as clients at a community mental health center, the results may be different from results obtained if a non-volunteer or random sample had been utilized. The subjects' client and volunteer status may have impacted upon how they responded to the objective measures of self-esteem/concept. For example, clients may have agreed to participate in the study to please their therapists and then responded to research instruments by under or over reporting levels of self-esteem.

Also, it was suspected that at least some of the subjects may have been responding to cues in the environment to come to some conclusions about what the researcher was hoping to achieve. In this study, subjects were told that they would be participating in a study dealing with "women and self-esteem". A number of subjects appeared to volunteer because of a desire to prove that women in general possess certain amounts of self-esteem. Thus, subjects may have either under or over reported when taking objective measures in order to meet their perceived notion of what the study was attempting to achieve.

5. Another possible source for the nonsignificant results may center around the scoring system developed to reflect the
stated assumptions. It is possible that the DAP is a valid measure of self but that the scoring system did not accurately reflect those assumptions. The variables selected may not reflect the assumptions or the mechanics of scoring may have distorted underlying construct measurement. The scoring system utilized a dichotomous forced choice rating scale in part to aid in ease of scoring and reliability which may have sacrificed the system's ability to adequately represent the underlying assumptions.

6. It is possible that the assumptions underlying the scoring system gathered from the pilot study were erroneous and/or the basic assumptions of the research as outlined in Chapter I are not viable. For example, DAP #2 may not represent the real self; DAP #3 may not represent the ideal self.

**Research Question Two**

Can the DAP scoring system discriminate between various levels of the criterion measures?

It was expected that judges would have more success in predicting criterion measure scores with the DAP when examining protocols from high, low or medium levels of self-concept and self-esteem. The study failed to support that expectation.

In addition to the issues raised in the discussion related to Question One, a possible cause of the nonsignificant results for this research question deals with the way in which subjects were placed into the high, medium and low groups. Because
subjects were placed into groups based upon similarity of criterion scores, they constituted a homogenous sample. Whenever correlation coefficients are computed from scores that are restricted in range, as was the case for Research Questions Two and Four, interpretations of the correlations should be done in a cautious manner. Most typically, such correlations will be lower than if a more variable group was utilized, and the possibility of committing a Type II error exists. That is statistically significant results may not be found when they do indeed exist.

In addition, the high and low group N's were small, approximately N=20; the small sample size of the subgroup increased the probability of finding no significant results. This also increased the possibility of a Type II error.

**Research Question Three**

Will the four raters, independently utilizing the author's rating system, demonstrate agreement?

The data indicated that the judges using the DAP Scoring System demonstrated strong agreement. The researcher made a strong effort to develop a scoring system that would not only be valid and reliable, but be of use in actual practice. Toward that end, clear scoring procedures and directions were developed along with a rating scale that was expected to be efficient and require minimal training. Thus, it appears that projective measures can be quantified in such a way as to allow ease and inter-rater reliability (Kahill, 1984).
Research Question Four

Will judges using the DAP scoring system exhibit differential scoring agreement based upon level of criterion measured?

The possibility of differential inter-rater reliabilities for high or low levels of the criterion measures was explored. Inter-rater reliabilities by criterion group level ranged from .35 to 1.00, with no statistical difference at the .05 level between groups. It appears that inter-rater agreement did not depend upon level of criterion group. Perhaps because of the clarity of the scoring system, judges were able to reach strong rater agreement no matter what the drawer's level of self-esteem or self-concept.

Recommendations

This section contains a discussion of recommendations based upon findings from the present study. Suggestions pertinent to future research as well as for practice are presented.

Future Research

1. Future research efforts involving the present study should center on re-evaluating and/or refining the scoring system. The system may benefit from modification of scoring procedures for certain variables such as the replacement of the dichotomous rating scale with a continuous one in hopes of producing a scoring system which may more accurately reflect underlying assumptions. Additionally, it may be profitable to
examine whether more extensive rater training prior to the actual scoring of protocols would increase DAP validity.

2. Additional focus on the impact of the judge characteristics upon scoring results would be beneficial. This research could include examining validity and reliability by sex, training of judges.

3. Future research should also be designed to explore the use of this DAP scoring system with different populations, such as nonclinical or male populations. Research with a clinical population may benefit from a design in which level and type of pathology are controlled.

4. Future research with the DAP may benefit from a shift away from the use of the DAP instrument alone, to the DAP interpreter and the interpretive process. This approach would examine the clinician as part of the DAP test. It is believed that one of the major reasons clinicians refuse to relinquish their use of the DAP is that many have come into contact with a DAP expert who appears to gather accurate and invaluable data from the instrument. No study to date has been conducted to examine the expert DAP diagnostician.

5. Future research to develop alternative methods to measure the self may be valuable. Emphasis upon objective measures rather than projective measures of the self may be a productive avenue to follow because of the higher reliability and
6. It is also necessary to raise the issue of the utility of any future DAP research. Because of the lack of a substantial theoretical base for the self-concept, it may be more appropriate to call a moratorium on all DAP research until a more sound theoretical base is established.

**Practice**

The validity of a test is not an attribute that can be described by a single study or isolated correlation coefficients. Thus, the specific meaning that can be drawn from this study is limited. However, the many years of DAP research has failed to support the validity of the instrument as a measure of self. In spite of this, clinicians are fond of the use of projective drawings in personality assessment, and have continued to use the instrument despite very mixed and inconclusive research findings concerning its validity. Perhaps rather than discourage its use, clinicians should be asked to use it more as a therapy tool, that is, a method by which to encourage further discussion with the client, rather than as an assessment tool.
APPENDIX A

MEMO TO CLINICIANS
MEMO

To: All Outpatient Clinicians
From: Julie Ganley
Re: Research Project

As most of you are aware, I am in the process of gathering data for my research project and would very much appreciate your assistance. The title of my study is the Draw-A-Person As A Measure of Self-Esteem and Self-Concept in Outpatient Females and is conducted under the supervision of The Ohio State University Professor Dr. James Wigtil. You can help me by going through your caseloads and referring to me clients appropriate for participation in my study. I am looking for clients who meet the following criteria:

(a) Female, age between 20 and 40, not pregnant
(b) Native American born, white
(c) Absence of chronic brain syndrome, mental retardation, psychosis (mental status must be stable and subject possess adequate intellectual abilities to insure sufficient concentration and attention span to complete testing tasks)
(d) Absence of overt physical handicaps, which may cause motor skill problems

If you have any such clients on your caseload, I would appreciate your querying them as to the possibility of participating in my study. In your discussion with the clients, please say that the study is dealing with "women and self-esteem", and that the researcher will be giving some testing tasks that are relatively easy and take approximately 30 minutes. In addition, please inform the clients that their participation in the study is entirely voluntary, they will not be charged for the testing time, and all data is to be gathered anonymously.

Thank you for your help and assistance. Please feel free to talk with me should you have any questions.
APPENDIX B

SCRIPT TO SUBJECTS
SCRIPT TO SOLICIT VOLUNTEER SUBJECTS*

Study: The Draw-A-Person As A Measure of Self-Esteem and Self-Concept in Outpatient Females.

A. Clients were solicited on a strictly volunteer basis. (See "memo to therapists"). Therapists at the counseling center were requested to query their female clients about the possibility of participating in a study which looks at self-esteem in women. Clients were asked in a low-key manner, with an emphasis upon the voluntary nature of the study. If a client expressed no interest, the matter was dropped. Clients who expressed an interest were referred to the researcher who covered the following points with the subjects:

1. The purpose of the research is to examine self-esteem and self-concept in women, and is being conducted under the supervision of The Ohio State University Professor Dr. James Wigtill. Any information gathered is to be kept strictly confidential.

2. The subject is requested to complete three brief pencil and paper tests. The first test consists of the drawing of human figures and the other two tests are inventories of self-esteem and self-concept.

3. Subjects are given an opportunity to decline participation in the study, are again assured that participation is entirely voluntary, and that they may withdraw at any time.

B. In addition the therapists were clearly instructed by the researcher to avoid any coercion, and that the clients should not in any way feel obligated to participate in the study.

C. Subjects are individuals of normal intelligence and adequate mental competence who are outpatient clients at a community mental health center.

* This script was created by request of The Ohio State University Human Subjects Review Committee and reflects their specific requirements.
APPENDIX C

SUBJECT CONSENT FORM
CONSENT FOR PARTICIPATION IN
SOCIAL AND BEHAVIORAL RESEARCH

I consent to participating in research entitled:

THE DRAW-A-PERSON TEST AS A MEASURE OF SELF-ESTEEM AND
SELF-CONCEPT IN OUTPATIENT FEMALES

Signed: __________________________ (Principal Investigator)

or his authorized representative has explained the purpose of the study, the procedures to be followed, and the expected duration of my participation. Possible benefits of the study have been described as have alternate procedures, if such procedures are applicable and available.

I acknowledge that I have had the opportunity to obtain additional information regarding the study and that any questions I have raised have been answered to my full satisfaction. Further, I understand that I am free to withdraw consent at any time and to discontinue participation in the study without prejudice to me.

Finally, I acknowledge that I have read and fully understand the consent form. I sign it freely and voluntarily. A copy has been given to me.

Date: _____________________________ Signed: _____________________________ (Participant)

Signed: __________________________________________________________________ (Principal Investigator or Authorized Representative)

Witness: __________________________

HS-027 (Rev. 3/87)-- (To be used only in connection with social and behavioral research.)
APPENDIX D

HUMAN SUBJECTS REVIEW APPROVAL
BWEBERIAL AND SOCIAL SCIENCES
HUMAN SUBJECTS REVIEW COMMITTEE (HSRC)
THE OHIO STATE UNIVERSITY

RESEARCH PROTOCOL:

8700136 THE DRAW-A-PERSON TEST AS A MEASURE OF SELF-ESTEEM AND
SELF-CONCEPT IN OUTPATIENT FEMALES, James V. Wigtii,
Richard C. Kelsey, Julie A. Canley, Human Services Education

Presented for review by the Behavioral and Social Sciences Review Committee
to ensure proper protection of the rights and welfare of the individuals
involved with consideration of the methods used to obtain informed consent
and the justification of risks in terms of potential benefits to be gained,
the committee action was:

_____ APPROVED
_____ DEFERRED*
X APPROVED WITH CONDITIONS*
_____ DISAPPROVED
_____ NO REVIEW NECESSARY

*CONDITIONS/COMMENTS:

Subjects were deemed NOT AT RISK and the protocol was
unanimously APPROVED WITH THE FOLLOWING CONDITIONS:

1. The HS-027 consent form should be used. Provide a completed
copy including Dr. Wigtii's signature.

2. State in the memo to the therapists that the study is being
done under the supervision of Dr. Wigtii at Ohio State University.

3. Revise the script to potential subjects to be more specific
and provide a copy to the Committee.

If you agree to the above conditions, PLEASE SIGN THIS FORM IN THE SPACE
PROVIDED BELOW AND RETURN WITH ANY ADDITIONAL INFORMATION REQUESTED TO ROOM
205, THE OHIO STATE UNIVERSITY RESEARCH CENTER, 1314 KINNEAR ROAD, COLUMBUS,
OHIO 43212, WITHIN ONE WEEK. Upon such compliance, the approval form will be
mailed to you. (In case of a deferred protocol, please submit the requested
information at your earliest convenience. The next meeting of the Committee
will be two weeks from the meeting date indicated above.)

DATE Sept. 8, 1987
Signature(s) /Wigtii
(Principal Investigator)

HS-025A (Rev. 3/85)
(Conditions/Comments)
APPENDIX E

ROSENBERG SELF-ESTEEM SCALE
ROSENBERG SELF-ESTEEM INVENTORY

DATE:

For each of the following statements, please mark an X in the response that best describes how you are feeling about yourself lately.

1. I feel that I'm a person of worth, at least on an equal basis with others.
   ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree

2. I feel that I have a number of good qualities.
   ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree

3. All in all, I am inclined to feel that I am a failure.
   ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree

4. I am able to do things as well as most other people.
   ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree

5. I feel I do not have much to be proud of.
   ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree

6. I take a positive attitude toward myself.
   ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree

7. In the whole, I am pretty satisfied with myself.
   ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree

8. I wish I had more respect for myself.
   ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree

9. I certainly feel useless at times.
   ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree

10. At times I think I am no good at all.
    ______ Strongly Agree ______ Agree ______ Disagree ______ Strongly Disagree
APPENDIX F

INSTRUMENT ADMINISTRATION PROCEDURES

BY INSTRUMENT
DAP ADMINISTRATION:

The subject is presented with a blank sheet of 8 1/2 x 11 inch unlined paper and a well-sharpened #2 pencil with eraser. The paper is white and contains no watermark of any sort. The subject is provided with a drawing surface that is flat and smooth with adequate illumination. The client has comfortable seating arrangement and can rest her arms comfortably on the drawing surface. One sheet of paper is placed in front of the subject in a vertical position, with one pencil. The instructions are as follows: "Please draw a picture of a person." Little other guidance is given, despite the frequent questions asked in an attempt to have the examiner structure the task more clearly. Because it is a projective task, the clinician must avoid the adding of any additional instructions which will prevent free performance. When the subject asks, "What kind of a person?" or "Do you want a whole person or just the head?" or "Is a stick figure O.K.?" the examiner responds with, "It's your drawing; do it any way you like."

Additionally, it is not uncommon for the subject to respond with comments denigrating her artistic talent. In such cases, the examiner makes such comments as, "This is not a test of artistic ability; I am not concerned with how good an artist you are...just do the best you can and don't worry." If the client draws a stick figure, she is given another piece of paper and told: "This time I'd like you to draw a real person, not a stick person."

When the first drawing is finished, the examiner presents the subject with another sheet of paper in the identical manner as was the first, with the instructions: "Now I'd like you to draw a person of the sex opposite of your first drawing." If the first drawing is not clearly one sex or the other, the client is asked at this time to identify the gender of the first drawing to facilitate the drawing of the second. If the S is not sure, she is asked to guess the sex of the first. When the second drawing is completed, the examiner presents the client with a third piece of paper with the instructions, "Now draw yourself." If the subject responds that she has already drawn herself in one of the other drawings, she is told: "Please draw yourself again."

As each drawing is completed, the examiner indicates the chronological order by placing in the upper right hand corner of each drawing DAP#1, DAP#2, DAP#3 as appropriate. The three drawings are then attached together with a paper clip.
Rosenberg Self-Esteem Scale

"For each of the following statements, please mark an X on the response that best describes how you are feeling lately."

Miskimins Self-Goal-Other Discrepancy Scale

"The purpose of this questionnaire is to measure your ideas about important areas of living. You will be asked to rate yourself, according to your own experience and feelings, on a total of 20 items. Each of these items is simply a pair of opposite words, such as 'good-bad', on which you will be required to give your standing by placing yourself nearer to 'good', nearer to 'bad', or somewhere in between."
APPENDIX G

CRITERION SUMMARY SHEET
CRITERION SUMMARY SHEET

Age: __________

Income: __________

Marital Status: __________

Occupation: ____________________

Education: __________

Scores on the Criterion Measures:

Self-Esteem ____________________ (Rosenberg)

Self-Concept ____________________ (Miskimins)
APPENDIX H

INSTRUCTIONS TO Raters
Instructions to Raters

Enclosed are 100 sets of figure drawings and corresponding scoring sheets. Each set consists of three separate drawings by the same subject. You will find a summary of scoring procedures attached. Carefully following these procedures, please score each DAP protocol.

Each rater is asked to independently judge each set of drawings in terms of ten variables. The rater will score each variable according to instructions below.

Please place your responses in the proper space on the DAP scoring sheet.

Please do not make any marks on the drawings themselves.
SCORING SYSTEM

Self-Concept Variables

1. Size
   1) Overall size is obtained (in inches) by taking the most distant lower and upper points of the drawing and projecting imaginary parallel lines at right angles to the axis of the body. The height of the figure is then measured on the line between these extremes.
   2) Measure drawing #1 and drawing #3 (by following the instructions above using the ruler provided).
      * Measure to the nearest \( \frac{1}{2} \) inch
      * Do not include non-clothing items (e.g., umbrella, fishing pole, but do include clothing (e.g., hat, etc.).

2. Placement
   1) Placement is scored by using the provided transparent sheet which has been divided into quarters, producing four quadrants equal in size. Quadrants are designed a) upper-left, b) upper right, c) lower left, d) lower right.
   2) Quadrant placement is determined by the quadrant(s) in which over 50% of a drawing is placed. Determine quadrant placement for drawings #1 and #3.

3. Gender Identification
   The rater is asked to determine the sex of drawings #1 and #3. If gender is drawing can not be determined, circle N.

4. Affect
   1) Emotional Tone will be categorized as
      Positive: happy, excited, content, enthusiastic,
      Neutral: indifferent, business-like, indeterminate
      Negative: sad, angry, bored
   2) Determine the affect of drawings #1 and #3.

5. Age
   1) The rater is asked to determine which of the following categories both DAP #1 and DAP #3 falls.
      Child: 0-16 years
      Adult: 17-55 years
      Elderly: Over 55 years
      * On this variable, as in all categories, rely upon your judgment, not given age stated on drawing page.
      * Cannot determine score CD
Self-Esteem Variables

1. **Size**
   Using the same instructions for size measurement as previously described, determine the size in inches of DAP #3.

2. **Overall Quality**
   1) Overall quality is defined as "the quality of the whole drawing; that is, its goodness or artistic quality.
   2) Using this definition, compare overall quality of DAP #3 to DAP's #1 and #2.
      * Highest quality also indicated by drawing with:
      - most detail
      - most complete

3. **Sexual Differentiation**
   Refers to the clarity of the DAP #3 sexual identity.
   * Can you determine sexual identity of drawing? Yes or No.

4. **Placement**
   Placement of the drawing is generally in the middle of the page. Placement is determined by drawing a verticle line which divides the page exactly in half as well as a horizontal line which does the same. Middle of the page placement is scored when the drawing is approximately in the center of the intersection of the two lines drawn as described above.

5. **Omissions**
   1) The typical drawing of a person consists of a head, facial features (eyes, nose, mouth, ears, hair), legs, feet, arms, hands, fingers, neck, shoulders and trunk. The typical drawing also consists of clothing such as dress, skirt, pants and blouse.
   2) Omission is scored when any of the above body parts omitted or a part has been cut off by the edge of the paper (paper chopping).
      * For drawings of female, omission scored when breasts, or bust outline missing.
      * Head only is scored omission.
APPENDIX I

DAP SUMMARY SHEET
Random Number ______________________
Rater ID ______________________

DAP SUMMARY SHEET

Self-Concept Scoring

+ _____ - _____
   _____ SIZE    DAP#1= _____ in.    DAP#3= _____ in.
   _____ PLACEMENT    DAP#1= a b c d    DAP#3= a b c d
   _____ GENDER ID    DAP#1= M F N    DAP#3= M F N
   _____ AFFECT    DAP#1= P Neg Neu    DAP#3= P Neg Neu
   _____ AGE    DAP#1= C A E C D    DAP#3= C A E C D

Composite Score ________

Self-Esteem Scoring

+ _____ - _____
   _____ SIZE    DAP#3= 6-7 in.    Under/over 6-7 in.
   _____ OVER QUAL    DAP#3= highest quality    Not highest quality
                      Same quality
   _____ SEX DIFF    DAP#3= ident sex    Not identif
   _____ PLACEMENT    DAP#3= center    Not center
   _____ OMISSIONS    DAP#3= no omiss    Omiss

Composite Score ________
APPENDIX J

SCORING SYSTEM
SCORING SYSTEM

Utilizing Handler's suggestions (1966) in the comparison of drawing scoring (+) indicating present more frequently, (-) indicating present less frequently.

DRAWING VARIABLES ASSOCIATED WITH SELF-CONCEPT

Good Self-Concept is demonstrated by congruent Drawings #1 and #3
Poor Self-Concept is demonstrated by incongruent Drawings #1 and #3.

1. Size. Persons with good self concept will see the real and ideal self as similar in size. Size congruency is established by taking measurements of each drawing and comparing the real and ideal selves. Overall size is obtained (in inches) by taking the most distant lower and upper points of the drawing and projecting parallel lines at right angles to the axis of the body. The height of the figure is then measured on the line between these extremes.

Scoring Procedure
+ DAP drawings #1 and #3 are within one inch of each other
- DAP drawings #1 and #3 are not within one inch of each other

2. Placement. Persons with good self concept will place the ego ideal and the ego real in similar positions on the page. Placement is scored by dividing a transparent sheet of plastic into quarters, producing four quadrants equal in size. Quadrants are designated a 1) upper left, 2) upper right, 3) lower left, 4) lower right.

Scoring Procedure
+ DAP #1 and #3 same placement
- DAP #1 and #3 not same placement

3. Gender Identification. Persons with good self concept will identify the real and ideal selves as the same sex.

Scoring Procedure
+ DAP #1 and #3 same sex
- DAP #1 and #3 not the same sex
4. Affect. Persons with congruency of ideal and real self will draw DAP #1 and #3 with similar affect. The categories used in scoring Affect are straightforward and taken from Wanderer (1969) and require few subjective judgments. Emotional Tone will be categorized as Positive (happy, excited, content, enthusiastic), Neutral (indifferent, business-like, indeterminate) and Negative (sad, angry, bored).

   Scoring Procedure
   + DAP #1 and #3 same emotional tone
   - DAP #1 and #3 not same emotional tone

5. Age. The congruent person will see the ideal and real selves as similar in age, indicating satisfaction with where one is in the lifespan.

   Scoring Procedure
   + Judge rates DAP #1 and #3 as falling in same age category
   - Judge rates DAP #1 and #3 as not falling in same age category
DRAWING VARIABLES ASSOCIATED WITH SELF-ESTEEM

High Self-Esteem is demonstrated by Drawing #3 highest quality and in most positive manner.
Low Self-Esteem is demonstrated by Drawing #3 lowest quality.

1. Size. The person with high self-esteem will draw a typical drawing about six to seven inches in size (Handler, 1985). The low self-esteem person will draw a figure either smaller or larger than norm. Overall size is obtained (in inches) by taking the most distant lower and upper portions of the drawing and projecting parallel lines at right angles to the axis of the body. The height of the figure is then measured on the line between these extremes. Measurements are made with a compass to enhance precision.

Scoring Procedure
+ drawing 6-7 inches in size
- drawing under or over 6-7 inches in size

2. Sexual Differentiation. Refers to the clarity of the figure's sexual identity.

Scoring Procedure
+ presence of secondary sexual characteristics and identifiable sex
- inability of judge to identify sex of drawing and/or no secondary sexual characteristics

3. Overall Quality. Persons with high self-esteem are thought to draw the self-drawing (DAP #3) as highest quality.

Scoring Procedure
+ DAP #3 highest quality, most accurate, detailed, complete drawing
- DAP #3 lowest quality

4. Placement. Placement of the drawing is generally in the middle of the page. Placement is determined by drawing a vertical line which divides the page exactly in half as well as a horizontal line which does the same. Middle of the page placement is scored when the drawing is approximately in the center of the intersection of the two lines drawn as described above.

Scoring Procedure
+ placement in the center of two lines which bisect the page
- placement not in the center as described above
5. Omissions. The typical drawing of a person consists of a head, facial features (eyes, nose, mouth, ears, hair), legs, feet, arms, hands, fingers, neck, shoulders, trunk. The typical drawing also consists of clothing such as dress, skirt, pants and blouse (Handler, 1985). Omission is scored when an essential body part is omitted or a part has been cut off by the edge of the paper (paper chopping).

**Scoring Procedure**
+ to receive a positive score on omissions, the drawing must have present: head, trunk, hair, nose, chin, neck, shoulders, arms, legs, feet, hands, eyes, mouth
- any of the above body parts missing
<table>
<thead>
<tr>
<th>Variables</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Size</td>
<td>DAP #3 6-7 in. in height</td>
</tr>
<tr>
<td>2. Overall Quality</td>
<td>DAP #3 highest quality compared to DAP #1, 2 most detailed, complete. &quot;artistic goodness&quot;</td>
</tr>
<tr>
<td>3. Sexual Differentiation</td>
<td>DAP #3 presence of secondary sexual characteristics and/or identifiable sex</td>
</tr>
<tr>
<td>4. Placement</td>
<td>DAP #3 placement in center of two lines which bisect the page</td>
</tr>
<tr>
<td>5. Omissions</td>
<td>DAP #3 must have present: head, trunk, hair, nose, chin, neck, shoulders, arms, legs, feet, hands, eyes, mouth</td>
</tr>
<tr>
<td>Variables</td>
<td>Positive (1)</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>1. Size</td>
<td>DAP #1 and #3 within 1 inch of each other</td>
</tr>
<tr>
<td>2. Placement</td>
<td>DAP #1 and #3 same placement</td>
</tr>
<tr>
<td>3. Gender Identification</td>
<td>DAP #1 and #3 same sex</td>
</tr>
<tr>
<td>4. Affect</td>
<td>DAP #1 and #3 same emotional tone</td>
</tr>
<tr>
<td>5. Age</td>
<td>DAP #1 and #3 fall into same age category (Child/Adult/Elderly or Cannot Determine)</td>
</tr>
</tbody>
</table>

**TABLE 40**

SELF-CONCEPT SCORING PROCEDURE BY VARIABLE
Sample DAP Protocol Scoring

The following scored DAP protocols are presented to illustrate the DAP scoring procedures utilized in this study. Four typical DAP protocols were chosen to serve as examples.

Please note that a DAP protocol consists of three separate drawings. These drawings were scored according to the scoring system used in this study (Appendix J and Tables 39 and 40).
Random Number ____________
Rater ID ____________

DAP SUMMARY SHEET

Self-Concept Scoring

+ ______
  + ______ SIZE
  - ______ PLACEMENT
  + ______ GENDER ID
  + ______ AFFECT
  + ______ AGE

Composite Score ______

Self-Esteem Scoring

+ ______
  - ______ SIZE
  + ______ OVER QUAL
  - ______ SEX DIFF
  + ______ PLACEMENT
  - ______ OMISSIONS

Composite Score ______
Random Number 2
Rater ID JAB

DAP SUMMARY SHEET

Self-Concept Scoring

+ ______
+ ______ SIZE  DAP#1= 2.5 in.  DAP#2= 3.5 in.  DAP#3= 3.5 in.
+ ______ PLACEMENT  DAP#1= a b c d  DAP#2= a b c d
- ______ GENDER ID  DAP#1= M F N  DAP#2= M F N
+ ______ AFFECT  DAP#1= P Neg Neu  DAP#2= P Neg Neu
- ______ AGE  DAP#1= C A E CD  DAP#2= C A E CD

Composite Score 3

Self-Esteem Scoring

+ ______
- ______ SIZE  DAP#3= 6-7 in.  Over/over 6-7 in.
+ ______ OVER QUAL  DAP#3= highest quality  Not highest quality  Same quality
+ ______ SEX DIFF  DAP#3= ident sex  Not identif
- ______ PLACEMENT  DAP#3= center  Not center
- ______ OMISSIONS  DAP#3= no omiss  Omiss

Composite Score 2
DAP #2
Random Number \[3\]
Rater ID \[JA6-\]

DAP SUMMARY SHEET

Self-Concept Scoring

\[+---\]
\[+____ SIZE \quad \text{DAP}\#1= 6 \text{ in.} \quad \text{DAP}\#3= 7 \text{ in.}\]
\[+____ PLACEMENT \quad \text{DAP}\#1= a b c d \quad \text{DAP}\#3= a b c d\]
\[+____ GENDER ID \quad \text{DAP}\#1= M \, F \, N \quad \text{DAP}\#3= M \, F \, N\]
\[--- AFFECT \quad \text{DAP}\#1= P \text{ Neg (Neu)} \quad \text{DAP}\#3= P \text{ Neg Neu}\]
\[+____ AGE \quad \text{DAP}\#1= C A E CD \quad \text{DAP}\#3= C A E CD\]

Composite Score \[4\]

Self-Esteem Scoring

\[+---\]
\[+____ SIZE \quad \text{DAP}\#3= 6-7 \text{ in.} \quad \text{Under/over 6-7 in.}\]
\[--- OVER QUAL \quad \text{DAP}\#3= \text{highest quality} \quad \text{Not highest quality} \quad \text{Same quality}\]
\[+____ SEX DIFF \quad \text{DAP}\#3= \text{ident sex} \quad \text{Not identif}\]
\[+____ PLACEMENT \quad \text{DAP}\#3= \text{center} \quad \text{Not center}\]
\[+____ OMISSIONS \quad \text{DAP}\#3= \text{no omiss} \quad \text{Omiss}\]

Composite Score \[4\]
Random Number 4
Rater ID J#6

DAP SUMMARY SHEET

Self-Concept Scoring

+ ___ -
+ ___ SIZE  DAP#1= 4 in.  DAP#3= 4 in.
+ ___ PLACEMENT  DAP#1= a b c d  DAP#3= a b c d
+ ___ GENDER ID  DAP#1= M F N  DAP#3= M F N
+ ___ AFFECT  DAP#1= Neg Neu  DAP#3= Neg Neu
+ ___ AGE  DAP#1= C A E CD  DAP#3= C A E CD

Composite Score 5

Self-Esteem Scoring

+ ___ -
+ ___ SIZE  DAP#3= 6-7 in.  Under/over 6-7 in.
+ ___ OVER QUAL  DAP#3= highest quality  Not highest quality  Same quality
+ ___ SEX DIFF  DAP#3= ident sex  Not identif
- ___ PLACEMENT  DAP#3= center  Not center
+ ___ OMISSIONS  DAP#3= no omiss  Omiss

Composite Score 3
LIST OF REFERENCES
REFERENCES


