11-1983

The Relationship Between Psychologists' Judgments of Artistic Quality and Judgments of Emotional Adjustment from Children's Human Figure Drawings

Karen Collier

THE RELATIONSHIP BETWEEN PSYCHOLOGISTS' JUDGMENTS OF ARTISTIC QUALITY AND JUDGMENTS OF EMOTIONAL ADJUSTMENT FROM CHILDREN'S HUMAN FIGURE DRAWINGS

A Thesis

Presented to

the Faculty of the Department of Psychology

Western Kentucky University

Bowling Green, Kentucky

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Karen Tucker Collier

November 1983

AUTHORIZATION FOR USE OF THESIS

Permission is hereby

☑ granted to the Western Kentucky University Library to make, or allow to be made photocopies, microfilm or other copies of this thesis for appropriate research or scholarly purposes.

☐ reserved to the author for the making of any copies of this thesis except for brief sections for research or scholarly purposes.

Signed Karen Collier

Date December 5, 1983

Please place an "X" in the appropriate box.

This form will be filed with the original of the thesis and will control future use of the thesis.

THE RELATIONSHIP BETWEEN PSYCHOLOGISTS'
JUDGMENTS OF ARTISTIC QUALITY AND
JUDGMENTS OF EMOTIONAL ADJUSTMENT FROM
CHILDREN'S HUMAN FIGURE DRAWINGS

Recommended November 21, 1983
(Date)

William Blohl
Director of Thesis

Doris L. Redfield

Robert E. Simpson

Approved December 5, 1983
(Date)

Elmer Gray
Dean of the Graduate College

Acknowledgments

This thesis was completed with the assistance of many individuals to whom I owe a debt of gratitude. I would like to express my apprecation to the thesis committee members--Dr. William Pfohl, Dr. Doris Redfield, and Dr. Robert Simpson-- for their guidance, encouragement, and enthusiasm throughout the various stages of the thesis. In particular, I would like to thank Dr. Pfohl, thesis director, for his constant assistance and support and Dr. Redfield for her help with the methodology and advice on other aspects of the thesis. In addition, I would like to acknowledge the Faculty Research Committee for supplying a grant for the thesis, Dr. Leroy Metze for his help with word processing, the educators and psychologists who assisted in collecting the drawings for the study, and the twelve psychologists who participated in the study.

Table of Contents

List of Appendixes

THE RELATIONSHIP BETWEEN PSYCHOLOGISTS' JUDGMENTS OF ARTISTIC QUALITY AND JUDGMENTS OF EMOTIONAL ADJUSTMENT FROM CHILDREN'S HUMAN FIGURE DRAWINGS

Karen T. Collier November 1983 108 pages
Directed by: W.F. Pfohl, D.L. Redfield, and R.E. Simpson
Department of Psychology Western Kentucky University

Children's human figure drawings (HFDs) have frequently been used as a projective technique to indicate emotional problems. Despite the popularity of this technique, research has shown contradictory findings on its validity as a measure of emotional adjustment. As a reason for the inconsistent findings, researchers have suggested that the artistic quality of HFDs may interfere with successful interpretation of adjustment from the drawings. However, the issue of the possible influence of artistic quality has not been adequately researched.

The major purpose of this study was to determine if a relationship existed between psychologists' judgments of artistic quality and judgments of emotional adjustment from children's HFDs. Children diagnosed as emotionally disturbed and normal children were randomly selected to produce HFDs. These children were matched according to age, sex, and IQ. Twelve psychologists were randomly selected to rate the drawings for emotional adjustment and artistic quality without knowledge of the children's adjustment status. The psychologists were allowed to employ methods

of interpretation they use in their practice. In addition, each psychologist was asked to list or describe the methods/criteria used in rating each drawing. Each psychologist was also asked to re-rate a random sample of the HFDs after one month in order to determine intrarater reliability.

Comparisons were made between (a) the artistic quality and emotional adjustment ratings, (b) the methods/criteria used in both ratings, and (c) the level of identification of actual adjustment from each set of ratings. The interrater agreement and intrarater stability of the ratings were also determined.

A positive, but nonsignificant, correlation was found between the artistic quality and emotional adjustment ratings, indicating that the two ratings may be measuring different dimensions of children's HFDs. The psychologists' perceptions of artistic quality of the HFDs evidently did not influence their ratings of emotional adjustment to a significant degree. An analysis of the criteria used in classifying drawings indicated that the same types of criteria were frequently cited for both types of ratings. Since the ratings were not highly correlated, the criteria were presumably interpreted differently in the two types of ratings.

A relatively high degree of interrater agreement was found for the artistic quality ratings and emotional adjustment ratings. The intrarater stability for both

types of ratings was also relatively high. However, the emotional adjustment ratings were not significantly related to the actual emotional adjustment status of the children making the drawings. These findings indicated that the methods used by the psychologists in interpreting the HFDs for emotional adjustment were of questionable validity. The artistic quality ratings were also not significantly related to the children's actual adjustment, indicating that emotionally disturbed children's HFDs were not necessarily perceived as having low artistic quality.

The results did not support the contention that artistic quality of HFDs is a confounding influence on HFD interpretation for emotional adjustment. The findings from this study also contribute to the body of research suggesting that children's HFDs are not valid for indicating level of emotional adjustment.

CHAPTER I

Introduction

Children's drawings have long been used as a means for the understanding of children. Since the late nineteenth century, the interest in children's drawings has been well-documented in America and in Europe. According to Goodenough (1926), as early as 1885 a study describing developmental stages in children's drawings was reported in England. Since that time, numerous studies of children's drawings have been conducted by psychologists and educators. The focus of interest in the majority of these studies has been on children's renditions of the human figure.

According to Hulse (1951), the rationale for employment of human figure drawings with children has been that the drawings allow children to more accurately express themselves, since they have limited ability to express themselves verbally. Children are able to convey thoughts, attitudes, feelings, and maturity through drawings that they cannot possibly express in words or in writing. As indicated by Klepsch & Logie (1982), "drawing speaks louder than words in the early stages of a child's development. It is, therefore, ideally suitable as a technique for uncovering information" about children (p. 8).

Communication through drawing is also basic and universal. Since all cultures are familiar with the human figure, the use of human figure drawings easily crosses language and cultural barriers. The brevity and nonverbal nature of human figure drawings allow their use "with those whose language production or attention span is problematic. Special populations such as the very young, the mentally retarded, and the learning handicapped can perform this task adequately without the frustration encountered with language-oriented measures" (Scott, 1981, p. 483). In addition, most children enjoy drawing, which makes the technique of using human figure drawings unobtrusive and nonthreatening to children (Scott, 1981).

These unique characteristics have led to the development of various assessment techniques using children's human figure drawings. The techniques can be divided into two main types of interpretations: use of the drawings as a projective technique and use as an objective test.

According to Lindzey (1961),

> A projective technique is an instrument that is considered especially sensitive to covert or unconscious aspects of behavior; it permits or encourages a wide variety of subject responses, is highly multidimensional, and it evokes unusually rich or profuse data with a minimum of subject awareness concerning the purpose of the test (p. 45).

Projective techniques used with children include measures such as word association tests, interpretation of play, sentence completion tests, interpretation of pictures, arranging pictures, and drawing techniques (Klepsch & Logie, 1982). Children's human figure drawings are used as a projective technique through the interpretation of drawings for indications of personality traits and unconscious needs, and to diagnose emotional problems (Koppitz, 1968; Machover, 1949).

In contrast to a projective technique, an objective test usually requires a limited and structured response. Answers are used to "arrive at measurements of a dimension or trait that relates to a criterion. The responses or scores obtained are usually treated as correlates of something else" (Sundberg, 1977, p. 174). Children's human figure drawings are often used as an objective test to determine mental maturity or IQ (Goodenough, 1926; Harris, 1963).

In general, both approaches (projective and objective) involve asking children, either individually or in groups, to draw a picture of a whole person, using letter-size (8½ x 11 inch) white paper and a number two pencil with an eraser. A drawing of an opposite sex figure and a self-portrait figure are sometimes requested (Harris, 1963; Machover, 1949). The examiner or person collecting drawings must avoid any kind of sugggestion, comment or criticism (Harris, 1963).

These techniques of using figure drawings are known as Draw-A-Persons (DAPs) or Human Figure Drawings (HFDs). Although the term DAP specifically refers to Machover's projective drawing technique (1949), it is often used interchangeably with the more general term HFD. Thus, DAPs and HFDs refer to the use of figure drawings, both projectively and objectively. This study will use the term HFDs for clarity.

Most psychologists seem to adhere to one of the two approaches to interpretation (i.e., projective or objective) exclusively (Koppitz, 1968). Since some of the same HFD items are considered by Harris (1963) as indicators of mental maturity and by Machover (1949) as indicators of emotional conflict, a clear differentiation is needed between the two approaches in order for meaningful interpretation of HFDs to occur (Koppitz, 1968). Therefore, research is usually conducted separately for projective and objective uses of HFDs. This study will focus on a projective use of HFDs to indicate children's emotional adjustment.

Current Status of HFDs as a Projective Technique

Since Machover developed the Draw-A-Person Test as a projective technique in 1949, it has continued to be one of the most frequently used psychological tests in clinical settings. A 1961 survey (Sundberg) of 185 clinics, hospitals, Veterans Administration facilities and other clinical settings revealed that the DAP was second in frequency

of psychological test usage to the Rorschach Inkblot Test. Results of a similar survey in 1971 (Lubin, Wallis, & Paine) revealed that the DAP was the fourth most popular psychological test. A more recent survey of the practice of psychological assessment among 274 school psychologists showed that the DAP was one of the most frequently used instruments for personality assessment (Goh & Fuller, 1983; Goh, Teslow & Fuller, 1981). In addition to the DAP, Goh and Fuller (1983) and Goh et al. (1981) found that two other frequently used personality instruments involved interpretation of HFDs: the House-Tree-Person and Kinetic Family Drawings.

In spite of its general popularity as a diagnostic instrument, the DAP has been the subject of contradictory research. The literature contains numerous studies which concluded that the DAP is valid for differentiating between those who are maladjusted (diagnostic categories such as emotional disturbance or schizophrenia) and those who are adjusted (Albee & Hamlin, 1950; Burton & Sjoberg, 1964; Goldman & Velasco, 1980; Goldman & Warren, 1976; Hall & Ladriere, 1970; Hiler & Nesvig, 1965; Holzberg & Wechsler, 1950; Koppitz, 1966a, 1966b, 1966c, 1968; Vane & Eisen, 1962).

In contrast, other studies have reported the opposite results--the DAP did not discriminate between various diagnostic categories (Adler, 1970; Cauthen, Sandman, Kilpatrick, & Deabler, 1970; Diffenbach, 1978; Fisher &

Fisher, 1950; Pihl & Nimrod, 1976; Reiznikoff & Nichols, 1958; Ribler, 1957; Ries, Johnson, Armstrong & Holmes, 1966; Royal, 1949; Schaeffer, 1964; Sherman, 1958; Strumpfer & Nichols, 1962; Wanderer, 1969; Watson, 1967b). Furthermore, the literature indicates that psychologists experienced with the DAP were no more successful in interpretation than persons not experienced with the DAP (Albee & Hamlin, 1949; Cressen, 1975; Fisher & Fisher, 1950; Hiler & Nesvig, 1965; Jue, 1976; Schmidt & McGowan, 1959; Wanderer, 1969; Watson, 1967a).

Researchers have attempted to explain the inconsistent research findings by hypothesizing that the artistic quality of drawings is erroneously influencing psychologists' interpretations of drawings (Feldman & Hunt, 1958; Roback, 1968; Schaeffer, 1964). For example, drawings of low artistic quality might erroneously be perceived as having been drawn by maladjusted children. Some research (Cressen, 1975; Lewinsohn, 1965; Nichols & Strumpfer, 1962; Strumpfer & Nichols, 1962) has indicated that artistic quality reflected in drawings may not be related to actual level of adjustment. Therefore, the influence of artistic quality might interfere with psychologists' accurate interpretation of drawings for adjustment.

However, the issue of the possible influence of artistic quality has not been adequately researched. Previous studies of the possible influence of artistic quality have been conducted using adults' drawings, which

have limited generalizability to children's drawings. One exception was a study by McIntosh (1981) using children's drawings, but this study used different groups of judges for ratings of artistic quality and ratings of adjustment and did not account for individual differences between judges. No studies have been conducted which compared individual psychologists' ratings of artistic quality and ratings of adjustment from children's HPDs.

The purpose of this study is to determine if a correlational relationship exists between psychologists' judgments of artistic quality and judgments of emotional adjustment from children's human figure drawings. Hypotheses were as follows: (a) there will be a significant positive relationship between psychologists' judgments of artistic quality and judgments of emotional adjustment from children's HFDs, (b) the criteria or methods used for ratings of artistic quality will be similar to those used for ratings of adjustment, and (c) ratings of artistic quality and adjustment will have a low level of identification of actual level of adjustment.

CHAPTER II

Review of the Literature

The research on the use of children's human figure drawings, both as an objective test and a projective technique, are examined in this literature review. Since the focus of this study is on the projective use of HFDs as a measure of adjustment, projective uses are emphasized, with objective uses covered briefly. The development of various scoring systems and techniques for interpretation are reviewed, along with studies of the reliability and validity of the scoring systems and interpretations. The research on the possible influence of artistic quality of drawings on psychologists' HFD interpretations is also reviewed.

Objective Uses

Objective uses of childrens HFDs generally involve scoring systems which estimate intellectual maturity or IQ (Buck, 1948; Goodenough, 1926; Harris, 1963; Koppitz, 1966). Each of these scoring systems is based on the assumption that as children increase in age their drawings reflect developmental changes. For instance, a typical three-year-old draws a person as a head, four-year-olds make "tadpole-like" drawings, five-year-olds draw a body and a head. At each successive age, detail in the drawings

increases and additional features, such as ears and fingers, are included. These details and features, along with position and proportion, are given points in most of the scoring systems, with the total score indicating a child's status relative to other children (Klepsch & Logie, 1982).

The first scoring system for estimating general intellectual maturity or IQ, the Draw-A-Man Test, was developed by Goodenough in 1926. This test was the first for systematically evaluating children's drawings on a point scale method (Klepsch & Logie, 1982). Harris restandardized and revised the scoring system of this test in 1963, resulting in the Goodenough-Harris Drawing Test, which is frequently part of a psychologist's battery of assessment techniques today (Sattler, 1982). Harris (1963) viewed this test as a measure of intellectual maturity, which he described as the ability to perceive (to discriminate likenesses and differences), to abstract (to classify objects according to likenesses and differences), and to generalize (to assign a discriminated object to a correct class).

Other scoring systems similar to the Goodenough-Harris Drawing Test have been developed. Buck (1948) established a scoring system for estimating IQ from figure drawings with the House-Tree-Person Test. Koppitz (1968) developed a scoring system for children ages 5 to 11 which assesses general level of mental maturity through the number of expected and exceptional items (e.g., pupils, hair, fin-

gers, etc.) on HFDs. HFDs are also part of the McCarthy Scales of Children's Abilities, which provides a general level of intellectual functioning for children ages 2½ to 8½ (McCarthy, 1972), and the Denver Developmental Screening Test, which is used to detect developmental disorders in children from birth through 5 years of age (Frankenburg & Dodds, 1975).

Reliability and Validity of Objective Uses

In a review of approximately 100 studies on the Good-enough-Harris Drawing Test (GH), Scott (1981) examined the reliability and validity of objective uses of HFDs to estimate IQ. Scott reported that GH intrascorer and interscorer reliability coefficients are uniformly high and significant, generally in the low .90s. Test-retest scores, measuring the consistency of the GH drawings themselves, were correlated in the low .70s. However, Goodenough (1926) reported a correlation of .94 between scores on two successive days, and Harris (1963) reported correlations of up to .91 over retest intervals of one week to three months.

Scott's major conclusions concerning the validity of the GH included the following: GH scores effectively discriminated performance between age levels from 5 through 12, an upward bias of approximately 10 standard score points was found in Harris's norms, the GH was found to have little utility as a predictor of academic achievement, the relation between GH performance and learning disabil-

ities is unclear, the HFDs of the mentally retarded are comparable to those of normal children of the same mental age, and socioeconomic status was the cultural variable which had the most effect on GH performance.

Overall, Scott concluded that the GH is a reliable test which effectively discriminates the performance of children at different age levels, but it is a poor predictor of individual and group performance on the major criterion intelligence tests. "These discrepancies are large enough to render the GH useless for predictive purposes in the average and upper ranges of intelligence. The GH holds some promise as a gross screening device for those of below average intelligence" (Scott, 1981, p. 503).

The reliability and validity of Buck's system for estimating IQ is questionable. Bolander (1977) noted that Buck's system has been criticized because his normative experiment used only twenty subjects in each of seven groups.

Koppitz (1968) reported no reliability data on her developmental scoring system for estimating IQ. She did indicate that it was valid for a majority of 347 subjects in a study in which HFD interpretation for general intelligence categories significantly correlated with intelligence test scores.

Projective Uses

In contrast to the use of HFDs as a measure of mental maturity, representatives of different schools of thought

use HFDs as a projective technique which involves analyzing drawings for signs (ways in which parts of the figure are drawn) of personality traits, unconscious needs, and conflicts (Koppitz, 1968). HFDs have been investigated to determine if certain signs occur more often in drawings of children with certain conditions (e.g., organic problems, learning disabilities, deafness or hearing impairment, obesity, congenital heart disease, encephalitis, and mental retardation) than in drawings of children without these conditions (Klepsch & Logie, 1982). However, most of the research using HFDs projectively is concerned with indications of emotional disturbance or maladjustment from drawings. This research includes scoring systems and techniques such as those developed by Machover (1949), and Koppitz (1966a).

Machover's Draw-A-Person Test

In 1949, Machover standardized the administration and formalized the interpretation of figure drawings as a projective technique, known as the Draw-A-Person (DAP) Test. She offered numerous hypotheses based on psychoanalytic theory regarding interpretations of DAPs, such as the detection of paranoid pathology, schizophrenia, or homosexuality through certain signs on drawings (Machover, 1949).

Machover's analyses were based on the body-image hypothesis--the assumption that certain emotions, perceptions, and sensations are located in various body parts

(Machover, 1949). Particular aspects of drawings were considered important to Machover, such as pencil pressure, variability and solidarity of lines used, rapidity of graphic movement, size of figure, succession of parts drawn, placement on the page, rigidity or spontaneity, and the use of background.

Specific body parts were associated with certain meanings. For example, the head was considered to be "the center of intellectual power, social dominance and control of body impulses" (Machover, 1949, p. 36). Those who place significance on intellectual achievement or those who suffer organic brain damage might draw disproportionately large heads (Machover, 1949). Although Machover's hypotheses were considered significant and influential, she offered no scoring system and presented no controlled research to support her claims (Koppitz, 1968).

Koppitz's Human Figure Drawing Test

Koppitz (1966a, 1968) presented the first refined scoring system for interpreting children's HFDs, which was based on the Interpersonal Relationship Theory of Harry Stack Sullivan. HFDs were considered to reflect a child's level of development and his or her interpersonal relationships. Koppitz did not adhere to Machover's body image hypothesis and felt that HFDs represent a child's current developmental stage and attitudes, which may change over time with experience and maturation.

In developing her scoring system, Koppitz determined that 30 items, called emotional indicators, occurred significantly more often in drawings of children with emotional problems than in drawings of well-adjusted children. The presence of two or more emotional indicators in a drawing was considered to be indicative of emotional problems and unsatisfactory interpersonal relationships.

Koppitz (1968) classified three different types of emotional indicators: (a) items related to the quality of the HFD, such as broken or sketchy lines, shading, gross asymmetry of limbs, tiny or big figures, and transparencies; (b) special features not usually found on HFDs, such as tiny or large head, vacant or crossed eyes, teeth, genitals, monster or grotesque figure, and sun, moon or clouds; (c) omissions of items expected in HFDs at children at a given age level, such as eyes, nose, mouth, arms, or legs.

Variations of the DAP

The projective technique of the drawing of a single human figure has been extended to include drawings of other people and objects. Two examples of this extention are the House-Tree-Person (HTP) Test and Kinetic Family Drawings (KFDs).

Buck (1948) developed the House-Tree-Person Test, which consists of a drawing of a house and a tree in addition to a human figure drawing. Buck considered the HFD as a projection of the drawer's self image, the drawing

of a tree as the projection of adjustment to the natural world, and the drawing of a house as adjustment to the human or social world. Emphasis was also placed on post-drawing interrogation and use of color in drawings.

Buck's method has been criticized as being vague (Bolander, 1977) and "not clear as to procedure of evaluation, or wholly satisfactory as a guide to interpretation" (Harris, 1963, p. 49). Two other scoring methods for the HTP (Hammer, 1954; Jolles, 1952) have been developed which differ from Buck's in a number of respects and, according to Harris (1963), offer no firm basis for qualitative study.

As developed by Burns and Kaufman (1970), the Kinetic Family Drawing (KFD) Technique involves asking children to draw a picture of everyone in their family doing something, including themselves. Analysis of KFDs is focused on action (movements of energy depicted between people), symbols (interpretations from a analytical frame of reference), and style (drawing characteristics suggestive of defensiveness). Burns and Kaufman based their scoring systems for KFDs on their clinical experience, providing no formal evidence of reliability or validity (Klepsch & Logie, 1982). According to Falk (1981), a small amount of research on KFDs has yielded positive results in using family drawings to understand children, but much more research is needed with this technique.

Reliability of Projective Uses

Swensen (1968) reviewed the literature on reliability studies on the projective uses of DAPs. He reported that interscorer reliability of drawings was adequate if judges or scorers were provided with training or explicit instructions. In studies assessing the reliability of the drawings themselves, specific signs involving structural and content variables (such as line quality and presence or absence of certain body parts) were found to have reliabilities "probably too low for making reasonably reliable clinical judgments" (Swensen, 1968, p. 40). However, global ratings, or ratings based upon the drawings as a whole, generally have satisfactory reliability, leading Swensen to conclude that "the reliability of a particular sign is a direct, linear function of the amount of drawing behavior included to assess that sign" (1968, p. 40).

Hammer and Kaplan (1966) asked 1300 fourth, fifth, and sixth grade children to draw a person and then draw a person of the opposite sex. They had the children repeat this task one week later in order to determine if the children drew the figures the same both times. They found the following indicators to be reliable: heads without bodies; type of mouth; missing fingers; erasures; shading; frontal view drawings; and upper, lower, and left placement on a page. Indicators found to be different on each administration or unreliable included the following: the drawing of teeth; right profile drawings; the omission of hands,

feet, and nose; and placement on the right side of the page. Gittelman-Klein (1978) criticized this study: "in such a large group, a significant correlation may account for very little variance (e.g., a correlation of 0.06 is significant beyond the 0.05 level of chance in a sample size of 1,000)" (p. 158).

Validity of Projective Uses

According to Falk (1981), studies evaluating the validity of projective uses can be divided into two types: (a) those testing the validity of individual hypotheses developed by Machover, and (b) those assessing the validity of the DAP as a whole, using diagnostic sorting tasks.

Validity of Machover's Hypotheses

To date, studies designed to test Machover's hypotheses have been at best inconclusive (Koppitz, 1968). In reviews of the literature on figure drawings, Klopfer and Taulbee (1976), Roback (1968), and Swensen (1957, 1968) concluded that Machover's hypotheses concerning the DAP generally had not been supported in the reviewed studies. According to Swensen (1957),

> more of the evidence directly contradicts her hypotheses than supports them. And, even in the studies where some support for her hypotheses can be found, many of the cases did not render the human figure drawings in the way that would be expected according to Machover (p. 460).

Swensen (1957) reviewed studies prior to 1956 which researched 30 of Machover's hypotheses concerning body parts and structural and formal aspects of the DAP. He found that only 1 hypothesis was supported, 13 others presented conflicting evidence, and the remaining 16 had not been supported. Similarly, Roback (1968) reported that 3 Machover hypotheses generally had been supported, 7 presented conflicting evidence, and 10 generally had not been supported in research from 1956 to 1967. Based on these questionable validity findings, both Swensen (1957) and Roback (1968) concluded that the utility of the DAP may be in the determination of gross level of adjustment.

Klopfer and Taulbee (1976) reviewed the literature on Machover's hypotheses from 1971 through 1976 and concluded that

> drawings can only be regarded as a suggestive kind of graphic behavior that will take on meaning as it is discussed with the subject and viewed in the context of other information. Many of the hypotheses formed by authors like Machover are at a level not clearly related to either conscious self-concept or behavior (p. 561).

Validity of the DAP as a Whole in Diagnosis

Nonsupportive studies. Numerous studies have concluded that diagnoses based on results of the DAP are invalid. These studies often used figure drawings by neurotic and/or schizophrenic adults and normal adults and

found that ratings of the drawings cannot discriminate between the different groups (Adler, 1970; Cauthen, Sandman, Kilpatrick & Deabler, 1970; Fisher & Fisher, 1950; Reiznikoff & Nichols, 1958; Ribler, 1957; Ries, Johnson, Armstrong & Holmes, 1966; Royal, 1949; Schaeffer, 1964; Sherman, 1958; Strumpfer & Nichols, 1962; Wanderer, 1969; Watson, 1967b).

Two of the most comprehensive of these studies were conducted by Strumpfer and Nichols (1962) and Ries et al. (1966). Strumpfer and Nichols found that of 16 DAP measures, none were able to differentiate the drawings of normal, neurotic and schizophrenic adults at a level exceeding chance. Ries et al. determined that only 3 out of 80 signs derived from the literature were able to distinguish normal subjects from schizophrenics.

Additional nonsupportive studies have assessed the validity of Koppitz's emotional indicators. Studies by Diffenbach (1978), Eno, Elliot, and Woehlke (1981), Pihl and Nimrod (1976), Snyder and Gaston (1970), and Szasz, Baade, and Paskewicz (1980) have reported questionable validity of the Koppitz scoring system and advised caution in its use as an indicator of emotional problems. A cross validation study by Fuller, Preuss, and Hawkins (1970) supported Koppitz's contention that emotional indicators occur more often in the HFDs of children with emotional problems than in those of normal children; however, it did not support the view that two or more indicators suggest

emotional problems. Many children diagnosed as emotionally disturbed had either one or no indicators in their drawings in this cross validation study.

Selfe (1983) also questioned the validity of Koppitz's scoring system. She stated:

> The difficulty with [Koppitz's] work is that it lacks a central congruence and fails to represent a satisfactory integrated model of emotional adjustment. An aggregate of "indicators" does not gel into a theory and, in any case, it is highly questionable that individual drawing features or habits rise from the emotional state of the subject alone. These features could have been formed in many ways: through instruction, or imitation of other children, cartoons, etc. (p. 22-23).

Other nonsupportive research using children's drawings was conducted by Springer (1941) and Stolz and Coltherp (1961). Springer found no differences between maladjusted and adjusted groups of children (defined by incidence of behavior problems) using their HFDs. Stolz and Coltherp showed that three clinical psychologists were able to predict intelligence to a signficant degree from the drawings, but were unable to predict either emotional adjustment or sociability.

Methodological problems were evident in some of the nonsupportive studies reviewed here. For example, Wanderer

(1969) asked psychologists to sort five pairs of drawings into five diagnostic groups. Only one pair of drawings could be placed in each category, so when a judge classified one set incorrectly, he or she was automatically wrong in another category. This procedure resulted in 40 percent error for one mistake rather than 20 percent if judges had been able to classify different pairs of drawings in the same category. Studies by Schaeffer (1964) and Watson (1967a) also forced psychologists to evaluate an equal number of drawings into three or four categories, inflating the error involved.

Many of the nonsupportive studies did not control for intelligence of subjects from which drawings were obtained (Adler, 1970; Diffenbach, 1978; Fisher & Fisher, 1950; Pihl & Nimrod, 1976; Schaeffer, 1964; Watson, 1967b). Failure to control for intelligence may represent an important source of error, because HFDs have been shown to reflect intellectual maturity or IQ (Goodenough, 1926; Harris, 1963). Developmental or cognitive factors reflected in drawings may have differed for the various classification groups and may have confounded the results of these studies.

The majority of the nonsupportive studies have limited generalizability to the use of the DAP for diagnostic classification with children's drawings. With the exception of studies of Koppitz's emotional indicators and studies by Springer (1941) and Stolz and Coltherp (1961),

the nonsupportive research used drawings by adults, usually Veterans Administration (VA) patients. While the selection of subjects from VA hospitals probably provided easy access to drawings and other records, it has resulted in a lack of studies concerning the validity of children's drawings in diagnosis.

In addition, the nonsupportive studies as a general rule did not adequately define the diagnostic categories. Instructions given to judges on how to sort the drawings were often vague. For instance, Sherman (1958) asked ten psychologists to discriminate between drawings made by "patients" and "nonpatients," leaving each psychologist to determine what was meant by "patient" and "nonpatient" status. Other studies asked judges to sort drawings into diagnostic categories such as neurosis and schizophrenia without giving any further description (Rieznikoff & Nichols, 1950, Royal, 1949; Schaeffer, 1964; Wanderer, 1967; Watson, 1967b).

Supportive studies. In contrast to the body of findings suggesting that the DAP is invalid for discriminating between diagnostic categories, other similarly designed studies reported the opposite results. Holzberg and Wechsler (1950) reported statistical data which significantly differentiated drawings of normal and schizophrenic subjects. Albee and Hamlin (1950) developed a criterion scale of drawings representing, according to the judgment of a number of psychologists, a continuum of

adjustment for individuals making the drawings. They then asked psychologists to rate drawings from schizophrenics, neurotics, and normals on level of adjustment using the criterion scale. This method proved effective in differentiating the normal group from the two groups of psychiatric patients, although it did not differentiate between the neurotic and schizophrenic groups.

Hiler and Nesvig (1965) determined that six criteria of drawings, "bizarre," "distorted," "incomplete," "transparent," "happy expression," and "nothing pathological" were valid in differentiating between normal and psychiatric adolescents. Murray and Deabler (1958) demonstrated that diagnostic judgments were accurate at a level greater than chance when clinicians were given ongoing corrective feedback. Burton and Sjoberg (1964) and Kay (1978) also presented data supporting the validity of figure drawings for discriminating between schizophrenics and normals.

In other studies showing positive results, Koppitz (1966a) found that four emotional indicators "grotesque figure," "no mouth," "no body," and "no arms" seemed to occur exclusively in drawings of clinic patients, which she felt might enhance their clinical validity. Vane and Eisen (1962) found that these same items were able to predict the emotional adjustment of kindergarten children.

In further research, Koppitz (1966b, 1966c) claimed that certain emotional indicators were of value for predicting school achievement among children at the kinder-

garten level and in the first two grades, and also that emotional indicators were found more often in the HFDs of aggressive children than in the HFDs of shy children. However, a study by Lingren (1971) found no significant differences between drawings of shy and aggressive children. These inconsistent results may have been due to sampling differences; Koppitz used a sample of children in a mental health clinic, whereas Lingren's sample was drawn from a sample of non-clinic children.

Other studies showing positive results with Koppitz's emotional indicators were conducted by Goldman and Warren (1976) and Goldman and Velasco (1980), who developed a scale showing that body-part omissions are the most predictive items of emotional high risk in kindergarten children. Contrary to Koppitz's contentions of specific body-part omissions, Goldman and Velasco's results suggested that an interchangeable number of omissions predict emotional risk, rather than specific items.

Hall and Ladriere (1970) compared the relative screening potential of HFD scales for children's drawings and found that the Koppitz emotional indicators significantly discriminated between problem and nonproblem children. Problem children were those diagnosed as emotionally disturbed (ED) and brain damaged or perceptually handicapped (BD). The scales did not discriminate between ED and BD children, which may have been due to difficulties in defining and discriminating between the diagnostic cate-

gories of ED and BD (Hall & Ladriere, 1970).

The supportive studies reviewed have some of the same problems as the nonsupportive research. Intelligence was again not controlled for in certain studies (Albee & Hamlin, 1950; Burton & Sjoberg, 1964; Holzberg & Wechsler, 1950; and Murray & Deabler, 1958). Interestingly, Burton and Sjoberg (1964) compared their control group and a schizophrenic group on home ownership, number of offspring, number of cars owned, education, marital status, age, religion, height, and even shoe size and dress size, but they did not attempt to determine if IQ differences existed between the two groups. Also, with the exception of studies of Koppitz's scoring system, the supportive research was conducted using adult's drawings.

Studies Involving Experienced vs. Naive Judges

Other tests of DAP validity are often comparisons of diagnostic success of experienced judges vs. naive judges (those having no experience with DAPs) in interpreting drawings (Falk, 1980). Schaeffer (1964) asked psychologists and nonpsychologists to discriminate between normal, neurotic, and psychotic subjects. Despite wide differences in clinical training, the judges did not differ in their ability to correctly identify the DAPs; none of the judges' total correct identification exceeded chance. Other studies have also shown that diagnostic accuracy does not vary significantly with expertise (Albee & Hamlin, 1949; Cressen, 1975; Fisher & Fisher, 1950; Hiler & Nesvig, 1965;

Jue, 1976; Schmidt & McGowan, 1959; Wanderer, 1969; Watson, 1967a).

In a study by Hiler and Nesvig (1965), the elimination of invalid criteria led to sharpening of clinical judgment and more accuracy in the evaluation of figure drawings. Stricker (1967) made Hiler and Nesvig's (1965) findings concerning valid criteria available to clinical students and experienced clinicians for purposes of judging drawings made by psychiatric patients and normals. The students, accurate in 73 percent of their judgments, were superior to the clinicians.

Arkell (1976a) found no significant differences among five categories of judges--elementary school administrators, elementary school secretaries, elementary school teachers, seventh grade students, and trained personnel in figure drawing interpretation--in inferring emotional maladjustment in HFDs. However, the groups ranged in accuracy of judgments from 79 percent to 83 percent, suggesting that HFDs may aid in the identification of emotional maladjustment in children ranging in age from seven to nine.

Overall, no studies have concluded that experienced judges show diagnostic superiority over naive or inexperienced judges. Various explanations have been suggested to account for these results. Arkell (1976b) hypothesized that adults untrained in HFD interpretation may have knowledge of how drawings made by children at different ages should look. He asked adults to make drawings which would

be representative of drawings made by children from five to ten years old, and significant correlations were found between these drawing-estimates and scores using Harris (1963) and Koppitz (1968) scoring systems. However, in a similar study which involved asking adults to simulate the drawings of children ages three to five, Leichtman (1979) found that untrained adults could not accurately simulate HFDs. These studies suggest that adults may be better able to estimate older children's drawings than those of younger children, which may be due to the older children's increased motor development. Further research is needed to determine if possible knowledge of how children's drawings should look influences adults' interpretations of drawings.

As another explanation for the difficulties in interpreting HFDs, Chapman and Chapman (1967) have suggested that interpretation is impaired by systematic errors on the part of judges. When judges observed human figure drawings paired with statements of the symptoms of the alleged patients who made the drawings, they tended to agree with one another by reporting that they observed in the drawings the same "illusory correlates" of the symptom statement. These "illusory correlates" are erroneously reported correlates which correspond to associative connections between symptoms and drawing characteristics, which according to Chapman and Chapman (1967), illustrate the relative ease with which one can interpret relationships

which do not exist.

The Influence of Artistic Quality of Drawings

Many researchers attempt to explain the contradictory DAP validity studies and the apparent ineffectiveness of training in DAP interpretation by claiming that judges are influenced by the artistic quality of the drawings. For example, according to Schaeffer (1964), "the factors responsible for this interjudge consistency need further investigation, but it is the author's subjective impression that this consistency may be related to artistic quality of the drawings" (p. 383). In his conclusion of a literature review on the DAP, Roback (1968) also expressed this viewpoint--"the 'clinical' cues which the psychologist believes are influencing his interpretations may actually be a reflection of the artistic qua..ty of the drawing" (p. 17). Similarly, Feldman and Hunt (1958) noted that

> a considerable portion of the variance in figure drawing interpretation may be attributed to structural aspects of the drawings as distinct from symbolic aspects presumed to reflect personality traits or dimensions . . . Clinicians are evidently influenced by the 'artistic' dimension, both in their overall evaluations of the drawing and with regard to specific areas of the drawing upon which they base judgment (p. 219).

These statements challenge assumptions made by Good-

enough (1926), Harris (1963), Koppitz (1968), and Machover (1949) that differences in drawing ability among those tested would not interfere with successful interpretation of drawings. Goodenough (1926) searched unsuccessfully for children "whose drawings appeared to possess artistic merit. . . comparable to the musical genius occasionally shown by children" (p. 53). Subsequently, she concluded that artistic ability was not a potent factor in producing high scores.

As evidence that the GH method of scoring drawings is independent of artistic quality of drawings, Harris (1963) cited a study by Phatak in which artistic drawings received more points on clothing and action items and nonartistic drawings exceeded on the proper location of body parts, which did not significantly influence total scores. In this study, artistic quality was defined as the characteristics of "pleasing, appealing, and interesting."

Koppitz (1968) found that HFDs were not markedly influenced by a child's performance ability (measured by performance scales on intelligence tests), which she assumed was necessary for artistic ability. Her assumption is questionable because the performance scales on intelligence tests in the study did not require the children to draw. As a consequence, the study cannot be interpreted as evidence that artistic quality does not influence interpretation of children's drawings.

None of the major researchers in the development of

HFD techniques have specifically addressed the question of whether artistic quality of drawings influences the interpretation of HFDs. Only a few researchers (Sherman, 1958; Whitmyre, 1953) have conducted studies which attempt to answer this question. Whitmyre (1953) found that the overall artistic value of drawings was highly related to clinical ratings of adjustment. He found ratings of personal adjustment from drawings correlated .88 and .86 with independent ratings of artistic quality for the same drawings. Whitmyre (1953) concluded

> as judged by the 'average' clinical psychologist today, human figure drawings executed by persons of average or above-average intelligence seem to indicate art achievement but do not seem to indicate any consistent relationship to level of personal adjustment (p. 424).

Similarly, Sherman (1958) found that psychologists' evaluations of drawings by psychiatric patients and normal adults for adjustment were significantly related to artists' evaluations of the same drawings for artistic ability. The art or the adjustment ratings in both of these studies (Sherman, 1958; Whitmyre, 1953) did not consistently show significant relationship to the dichotomy of psychiatric vs. nonpsychiatric status. However, it must be noted that one group of judges rated adjustment status while another group of judges rated artistic quality, and comparisons were made between the two. There may have been

differences between the groups of judges in their ability to deal with HFDs; therefore, the results of these studies must be interpreted with caution.

Additional studies lend empirical support to the position that artistic quality of drawings influences judges' evaluations and represents an important source of error. According to Strumpfer and Nichols (1962), an Artistic Quality Scale was developed by Wagner and Shubert (1955) in order to quantify global judgments about artistic quality of DAPs by late adolescents and young adults. Wagner and Shubert's scale was made up of four series of seven illustrative drawings each--front and profile series for both male and female figures. Verbal descriptions were added in order to aid in rating drawings from the series, along with instructions on rating unusual drawings.

Strumpfer and Nichols (1962) found that the Artistic Quality Scale and measures of adjustment, sexual differentiation, maturity, aggression, and body image disturbance did not discriminate between the drawings of normal, neurotic, and schizophrenic adults. Strumpfer and Nichols also found significantly high correlations between the Artistic Quality Scale and the other drawing scales, leading them to conclude that psychologists' judgment of personality factors are influenced by artistic merit of drawings. However, the results of Strumpfer and Nichols's study have questionable impact, because the validity of the Artistic Quality Scale was not mentioned.

Feldman and Hunt (1958) found considerable overlap between those parts of figure drawings rated as most difficult to draw by artists and those areas of drawings most frequently selected by clinicians as revealing emotional disturbance. These results led Feldman and Hunt to conclude that "a proficiency in drawing" dimension strongly influences clinical evaluation of HFDs, because there was not an adequate reason for believing that maladjustment would lead to irregularities in just those body parts which are most difficult to draw.

Lewinsohn (1965) conducted a study of overall quality of HFDs, which was defined as "the quality of the whole drawing as a drawing, that is, its goodness or artistic quality" (p. 504). This overall quality was found to be unrelated to specific aspects of psychopathology, improvement in clinical condition, and a wide variety of personality trait ratings, leading Lewinsohn to conclude that "a lack of relationship between overall quality and symptomatic manifestations of emotional disorder" was suggested (p. 310).

Two factor analyses of HFDs (Adler, 1970; Nichols & Strumpfer, 1962) yielded a single factor accounting for most of the common variance among drawing scores. This major factor was interpreted as overall quality or artistic quality in a technical rather than aesthetic sense. According to Nichols and Strumpfer (1972), the major factor "seems mainly to reflect the technical skill of the subject

in executing a drawing and has little to do with aesthetic appeal. Picasso would score very low on [this] factor" (p. 160). Furthermore, Nichols and Strumpfer found that the overall quality factor was unrelated to adjustment of VA patients, and concluded that overall quality of drawings has little relationship to psychological adjustment.

Two other studies lend empirical support to the position that artistic quality of drawings influences judges' evaluations of DAPs. Cressen (1975) reported that trained and naive judges erroneously tended to see drawings of low art quality as being drawn by schizophrenic patients and drawings of high overall quality as being drawn by nonpatients. McIntosh (1981) asked psychologists to sort drawings into categories of adjusted and maladjusted, and asked artists to sort the same drawings into categories of more artistic and less artistic. It was found that both groups of judges essentially used the same basic set of criteria in making their decisions.

However, studies by Lewinsohn (1965) and Maloney and Glasser (1982) questioned the lack of relationship between artistic quality and adjustment. Lewinsohn found low but statistically significant relationships between overall quality and three ratings of adjustment (ratings of patients' adjustment made by relatives, ratings by nurses, and ratings of cooperativeness while taking psychological tests). Maloney and Glasser found that ratings of overall quality discriminated between the drawings of psychiatric

and normal adults.

Thus, it appears from the research that (a) artistic or technical quality of drawings may not be related to level of adjustment, and (b) when judging drawings for adjustment, psychologists may be influenced by the artistic quality of the drawings, which may partially explain the contradictory research on DAP validity. The artistic quality may be a source of error in psychologists' judgments of drawings, and controlling for it may be necessary for more valid use of the DAP.

However, the research is lacking in several important areas. With the exception of one study (McIntosh, 1981), all of the research related to the question of whether artistic quality influences HFD interpretation has been done using drawings from adults. Also, the methodology of the studies does not account for individual differences between judges, since different judges were used for ratings of artistic quality and ratings of adjustment.

Evaluation of the Literature

Overall, the body of research presented here can be criticized in several areas. These criticisms include the use of adults' drawings instead of children's drawings, the lack of adequate definitions of categories, and other methodological problems.

A major problem is that few of the studies used children to produce drawings, which is inconsistent with previous research on DAP techniques. Goodenough, Harris, and

Koppitz all focused on children in their DAP research, and although Machover's DAP test was largely based on her experience with adolescent and adult patients, she extended her findings to children (1953). Since all of these major figures in the development of HFDs as a diagnostic tool dealt with children, it does not follow logically that most of the subsequent research was done with adults' drawings. According to Falk (1981),

> clinically there are numerous possible explanations for using diagnostic drawing techniques primarily with children. One, in everyday life, children have a greater tendency to communicate by giving "clues" about things they feel and think. . . Two, drawing is generally considered something children do; many adults feel foolish when given a drawing task. . . Three, young children are more likely to become absorbed in doing the drawing, whereas adults may concentrate more on the interpretative aspect of the task (i.e., what the psychologist is going to read into it). In other words, an adult's psychological defense structure is much more developed and resistant to projection in a drawing task (p. 468).

The primary emphasis on research with adults' drawings has resulted in a limited amount of accumulated knowledge regarding the usefulness of children's drawings. As stated

by Gittelman-Klein (1978),

> it is reasonable to assume that disturbed children differ from adults and adolescents in important respects that preclude the assumption that various age groups represent a homogeneous psychopathological population. It therefore seems unjustified to generalize to children from studies performed with adults (p. 141).

Another significant problem in the studies cited in this review is the lack of adequate definition of diagnostic categories. It is not clear that inconsistent findings in diagnostic sorting tasks are due to DAP invalidity; they may partially result from differential interpretation of the labeled categories into which psychologists are asked to sort drawings. Very few studies have operationally defined the categories used in the research. As a typical example of this problem, Whitmyre (1953) stated "each psychologist used his own concept of what constitutes 'adjustment' as it is commonly used by clinicians" (p. 422). Thus, almost all of the research reviewed here was based on the assumption that psychologists agree upon or have a common understanding of the various labels or categories.

However, research shows that this assumption of common understanding of categories is faulty. As Hobbs (1975) noted, there is little agreement as to what constitutes emotional disturbance; despite the prevalence of children

labeled emotionally disturbed, it is exceedingly difficult to agree on a definition for the term. Hobbs described the range of meanings given to emotional disturbance:

> emotional disturbance is a generic term referring to conitions ranging from mild and temporary reactions to profound and prolonged disabilities. There are numerous related terms: adjustment problems of childhood, behavior disorders, mental illness, neurosis, childhood schizophrenia, and infantile autism (1975, p. 55).

Categories such as neurosis and schizophrenia are also difficult to define. Falk (1981) stated:

> Psychologists' ideas of what constitutes schizophrenia and other 'mental disorders' vary considerably. . . Few psychologists would disagree that the diagnosis of schizophrenia is not as clear-cut as the diagnosis of measles. However, there is the unquestioned assumption that diagnoses such as schizophrenia can be used in the same matter-of-fact way that diagnoses of measles is used (p. 468).

Other methodological problems were evident in the studies reviewed. IQ was not controlled for, forced sorting procedures led to increased error in accuracy of judgments, and individual differences between judges were often not accounted for.

Summary

The DAP research in this literature review was characterized by inconsistent and contradictory findings. While many studies concluded that the DAP is invalid for discriminating between diagnostic categories, various other studies reported just the opposite results. It also appears from the research that experienced judges are no more successful in interpreting DAPs than naive or inexperienced judges.

Methodological problems (e.g., failure to control for factors such as IQ and individual differences between judges, and failure to adequately define categories) probably contributed to the inconsistent research findings and lack of comparability across studies. Another explanation for the difficulties in DAP research often cited is that artistic quality of drawings may interfere with successful judgment of adjustment from the drawings. It has been demonstrated in several studies that artistic quality appears to be unrelated to adjustment. Therefore, if psychologists' ratings of adjustment are highly related to ratings of artistic quality, then the validity of judgments made from drawings is questionable without control for artistic quality.

However, the question of the influence of artistic quality has not been adequately researched. The studies reviewed have used adults' drawings, which limits generalizability to children's drawings. They have also not

assessed how psychologists view artistic quality of drawings, because other nonpsychologist judges have been used for the artistic quality ratings. Ratings of artistic quality by artists or other judges are not directly comparable to psychologists' ratings of adjustment because individual differences may influence the comparisons between the two types of ratings. Thus, there is a need for further, more methodologically sound research which seeks to determine if a relationship exists between psychologists' judgments of artistic quality and judgments of emotional adjustment.

CHAPTER III

Methodology

Subjects

Ten children were randomly selected from a pool of 30 children diagnosed as emotionally disturbed (ED) in Colorado, North Carolina, and Kentucky. These children were diagnosed as ED by their school systems or in a clinical setting and were currently placed in ED classes. The children were all male and ranged in age from seven years-ten months to ten years-eleven months, with a mean age of nine years-seven months. They were of "average" intelligence with a mean IQ of 101.5 from qroup and/or individual intelligence tests. Average is defined as scoring within one standard deviation of the mean on a standardized intelligence test.

Ten other children, also males, were randomly selected from a pool of 30 children who had not been identified as having emotional problems. The non-ED children ranged in age from eight years-one month to ten years-ten months, with a mean age of nine years-eight months. They were also in the average range of intelligence, with a mean IQ of 101.0.

Judges

Twelve psychologists were randomly selected to rate the human figure drawings in this study. These judges were certified/licensed clinical or school psychologists (at both the master's and doctoral level) in Kentucky. The psychologists had at least two years of experience in regular use of children's human figure drawings as a projective technique. An equal number of male (n=6) and female (n=6) psychologists was randomly selected from a pool of 40 psychologists.

Procedures

Each child was given an 8½ x 11 inch sheet of unlined white paper and a number two pencil. Instructions were given to "draw a picture of a whole person." Immediately afterwards, each child was given another piece of paper and instructed to draw another picture of a whole person. This procedure resulted in the collection of 40 drawings. After the drawings were obtained, any identifying material, with the exception of the child's age, was removed.

One drawing from each of the 20 pairs of HFDs was presented to each judge, counterbalancing for the order in which the children produced the drawings. Three male and three female judges were asked to rate the drawings for artistic quality; the other three male and the other three female judges were asked to rate the drawings for level of emotional adjustment. A rating sheet was attached to each drawing for both artistic quality and emotional adjustment

ratings (See Appendixes E and F). The judges were also informed that the drawings were made by male children in the average range of intelligence.

Those judges rating artistic quality were asked to indicate if the individual drawings were of high, medium, or low artistic quality. Artistic quality was defined as the goodness or the technical accuracy of the drawing (i.e., how well it represents a person), rather than its aesthetic appeal. These judges were then asked to further differientiate among the artistic quality of each drawing by rating the drawing on a 7 point scale. Ratings of 1 and 2 corresponded to low artistic quality, ratings 3, 4, and 5 corresponded to medium quality, and ratings 6 and 7 to high quality.

The judges were also instructed to list or describe the methods or criteria they used in determining each categorization (e.g., global impressions, content, details, or other elements in the drawing). They were asked to be as specific as possible in listing or describing the criteria from each drawing which resulted in its classification in a particular category.

Those judges rating the first drawing of each pair for level of emotional adjustment were asked to use the definition of emotional disturbance from Public Law 94-142 as a guideline. The judges were asked to indicate if the drawing best reflected emotional adjustment or maladjustment (emotional disturbance). Then, each judge was asked to

list, for each drawing, the criteria or methods used in determining the emotional adjustment ratings (e.g., a scoring system; content, detail or other elements in the drawing; global impressions; or a unique system of interpreting drawings).

Finally, the rating procedure was repeated with the remaining drawing from each pair of HFDs. The counterbalancing procedure for order of student drawing remained in effect. Those judges who rated the first drawing of each pair for artistic quality rated the second drawing for level of adjustment; those judges who rated the first drawing of each pair for level of adjustment rated the second drawing for artistic quality.

The instructions for making both types of ratings emphasized rating the drawings individually. The psychologists were instructed to rate each drawing and list the criteria used without referring to the other drawings. They were asked not to make comparisons between the drawings. The complete instruction sheets for both artistic quality and emotional adjustment ratings are presented in Appendixes B and C.

The psychologists participating in this study were also asked to complete an information form in order to determine (a) their training in the projective use of children's HFDs, (b) how they use children's HFDs projectively in their assessments, and (c) what emphasis they place on children's HFDs in determination of level of emo-

tional adjustment. This information form is presented in Appendix D.

One month after originally rating the drawings, the psychologists were asked to re-rate a random sample (n=20) of the drawings for emotional adjustment and artistic quality. This procedure was employed in order to determine the intrarater reliability or degree of stability of the ratings over time.

Analyses

The analyses addressed six questions:

1. Did the level of artistic quality reflected in human figure drawings influence psychologists' judgments of the drawings with regard to emotional adjustment? In other words, what was the degree of relationship between psychologists' ratings of artistic quality and ratings of emotional disturbance from children's human figure drawings?

2. Were the criteria the judges used to categorize the children's drawings according to artistic quality and level of adjustment similar?

3. Were children previously diagnosed as maladjusted or emotionally disturbed presently judged to be maladjusted on the basis of their human figure drawings alone?

4. What was the relationship between the ratings of artistic quality of human figure drawings and actual level of adjustment, i.e., ED vs. non-ED?

5. How closely did the judges agree in their ratings of artistic quality and adjustment?

6. How stable were each judges' ratings of artistic quality and adjustment over time?

To address the first question on the relationship between ratings of artistic quality and emotional adjustment, a point biserial correlation coefficient between the two ratings for all subjects was computed for each judge. The average correlation coefficient (McNemar, 1969) among the judges was calculated.

Question 2 regarding the criteria used for ratings was addressed descriptively. These descriptive data are evaluated in the Results Section.

The third question regarding the validity of the ratings of emotional adjustment required calculation of the percent agreement between ratings of adjustment and actual adjustment across subjects for each judge. The average percent agreement across judges was computed. (Hendricks, Balzer, and Sheehy (1980) recommended the use of percent agreement when estimating the reliability of nominal data.)

Question 4, concerning the relationship between ratings of artistic quality and actual level of adjustment, required the calculation of a Pearson product moment correlation coefficient for each judge between artistic quality ratings and actual adjustment. An average correlation coefficient was also computed.

Question 5 on the interrater reliability or degree of consistency among judges' ratings was addressed by the computation of intraclass correlations (Winer, 1971).

Intraclass correlation coefficients were computed for both the artistic quality ratings and emotional adjustment ratings.

Question 6, concerning the intrarater reliability or the degree of consistency between each judge's ratings, required that each judge re-rate a random sample (n=10) of the drawings originally rated for artistic quality and a random sample (n=10) of the drawings originally rated for emotional adjustment. A Pearson product moment correlation was computed for the artistic quality variable, and the percent agreement was determined for the adjustment variable. Both of these analyses were averaged across judges.

CHAPTER IV

Results

This chapter presents the results of the analyses of (a) the relationship between artistic quality and emotional adjustment ratings, (b) the methods/criteria used in both ratings, (c) the level of identification of actual adjustment status from each set of ratings, (d) the inter- and intrarater reliability, and (e) information on the projective use of HFDs obtained from the judges. These results indicate whether the three hypotheses were supported or rejected. First, it was hypothesized that there will be a significant positive relationship between psychologists' judgments of artistic quality and judgments of emotional adjustment. The second hypothesis was that the criteria or methods used for ratings of artistic quality will be similar to those used for ratings of adjustment. Third, it was hypothesized that ratings of artistic quality and emotional adjustment will have a low level of identification of actual level of adjustment.

The Relationship Between the Two Types of Ratings

The point biserial correlation coefficients for the relationship between the artistic quality ratings and the emotional adjustment ratings ranged from -.22 to +.71; the average correlation coefficient across the twelve judges

was +.37. This correlation coefficient of .37 is not significant at the .05 level. Therefore, the hypothesis of a significant positive relationship between the two types of ratings was not supported. Each psychologist's correlation between the two ratings is presented in Appendix G.

Methods/Criteria Used for Ratings

The methods or criteria used by the psychologists in making artistic quality and emotional adjustment ratings were often of the same type or category, lending support to the second hypothesis that the criteria would be similar. A cumulative frequency analysis of the methods/criteria used by the judges in classifying the 40 drawings indicated that 68 percent of the criteria categories mentioned in making emotional adjustment ratings were also mentioned in making artistic quality ratings. Similarly, 69 percent of the criteria categories listed in rating drawings for artistic quality were also mentioned in ratings for emotional adjustment. In other words, only 32 percent of the total emotional adjustment criteria categories were unique to emotional adjustment ratings, and only 31 percent of the total artistic quality criteria categories were unique to artistic quality ratings.

The five most frequently listed criteria for emotional adjustment ratings, in decreasing order of frequency, were global impressions, detail, size, placement, and content. For artistic quality ratings, the five most frequently mentioned criteria, in decreasing order of frequency, were

detail, proportion, global impressions, form, and line quality. Two of the criteria, detail and global impressions, accounted for 23 percent of the total number of criteria (rather than categories) used for emotional adjustment ratings and 29 percent of the total criteria used for artistic quality ratings.

The majority of the critera common to both ratings could be classified into three types: (a) vague general criteria (e.g., content, form, immaturity, quality, general mood); (b) criteria related to the execution of the drawings (e.g., proportion, size, placement, line quality, shading); and (c) criteria concerning specific content or body parts (e.g., arms, head, hands, trunk). Other criteria were listed only by single judges (e.g., primitive, partial profile), but were listed in making both types of ratings. Appendix H presents a cumulative frequency analysis of the criteria which were common to both artistic quality and emotional adjustment ratings.

The most frequently listed criteria unique to emotional adjustment ratings were distortion and midline emphasis. The majority of the other criteria unique to emotional adjustment ratings were mentioned only once or twice in the twelve psychologists' ratings of 20 drawings. These criteria were generally related to specific content in the drawing (e.g., opposite sex drawing, eyes, violent scene). The most frequently listed criteria unique to artistic quality ratings were structure, basic features,

and composition. Like the emotional adjustment criteria, the majority of the other criteria unqiue to artistic quality ratings were listed only once or twice in the total ratings. These criteria could be classified in general as related to the execution of the drawings (e.g., sureness of stroke, texturing, balance, sketching). A cumulative frequency analysis of the criteria unique to artistic quality and emotional adjustment ratings is presented in Appendix I.

Only four of the twelve judges indicated a specific scoring system used for either of the ratings. One judge used the Koppitz (1966a) scoring system for the emotional adjustment ratings. The Goodenough-Harris (1963) scoring system was used by three judges--two used it as one of the criteria for emotional adjustment ratings, while one judge used it as one of the criteria for artistic quality ratings.

Although they were not instructed to do so, five of the judges made diagnoses on the basis of the drawings. Examples of these diagnoses include schizoid personality, poor self concept, sexual disturbance, psychosis, guilt, psychosomatic tendencies, and learning disabilities. Four of these judges made these diagnoses only occasionally in the emotional adjustment ratings; the other judge consistently made diagnoses on each drawing of the emotional adjustment section.

The Relationship Between Ratings and Actual Adjustment

The point-biserial correlation coefficients for the relationship between artistic quality ratings and actual emotional adjustment status of the children making the drawings ranged from -.06 to +.22; the average across the twelve judges was +.09. The average correlation coefficient of .09 is not significant at the .05 level. Each judge's correlation coefficient for this relationship is listed in Appendix J.

The average percent agreement between the emotional adjustment ratings and actual adjustment was 57.92 percent, which is not significantly greater than chance. The individual judges' percent agreement between the emotional adjustment ratings and actual adjustment ranged from 35 percent to 65 percent. Each judge's percent agreement is listed in Appendix K. These results support the third hypothesis that ratings of artistic quality and emotional adjustment would have a low level of identification of actual level of adjustment.

Interrater Reliability

As determined by an intraclass correlation, the interrater reliability for the artistic quality ratings was .86. Intraclass correlation procedures also indicated that the interrater reliability for the emotional adjustment ratings was .75. These results indicate that the twelve judges substantially agreed with each other on both types of ratings. Additional indicators of interrater agreement

include (a) on 18 of the 20 drawings, the twelve judges agreed on the emotional adjustment ratings 75 percent or higher, and (b) on 14 of the 20 drawings, the standard deviation of the artistic quality ratings was less than 1 on a 7 point scale. However, it should be noted that this high degree of consistency was not indicative of the children's actual adjustment status.

Intrarater Reliability

The average correlation coefficient for the intrarater reliability of the artistic quality ratings was .90. The the individual judges' correlations of the artistic quality ratings ranged from .62 to .98. Appendix L presents each judge's correlation between the artistic quality ratings.

The average percent agreement for the intrarater reliability of the emotional adjustment ratings was 87.78 percent. The individual judges' percent agreement for the emotional adjustment ratings over time ranged from 70 percent to 100 percent. Each judge's percent agreement is presented in Appendix M. These results indicate that the ratings of both artistic quality and emotional adjustment in this study were relatively stable over time.

Judge Characteristics

The psychologists who rated the drawings were trained in the projective use of children's human figure drawings through a variety of methods which included personality assessment/projective techniques courses, workshops, and/or practicum experiences. They reported how they generally

use children's HFDs projectively in their assessments. These types of projective uses included (a) to establish rapport, (b) as a screening device to identify or rule out emotional disturbance, (c) to form hypotheses and gain supportive information, and/or (d) to get a "notion" for the child's attitudes and outlooks.

The average length of use of children's HFDs as a projective technique was 7 years, with a range of 2 to 13 years. The judges reported using children's HFDs projectively in an average of 63 percent of their assessments, with a range of 10 percent to 100 percent. (Three judges reported 10 percent and one reported 25 percent usage whereas the other eight judges reported 75 percent or higher.) When using a 7 point scale to indicate the degree of emphasis placed on HFDs in making decisions about emotional adjustment, the judges averaged 3.8. On this scale, 1 represented no emphasis and 7 represented great emphasis. Thus, all of the judges placed at least some emphasis on HFDs in making decisions about emotional adjustment; the scores on this scale ranged from 2 to 7.

CHAPTER V

Discussion and Summary

This chapter presents a discussion of the results of this study, comparing the findings to previous research. The issue of validity of HFDs as a measure of emotional adjustment is evaluated, including positions taken by various researchers. Finally, limitations of this study and sugggestions for future research are presented. A summary of the study concludes the chapter.

Interpretation of the results of the various correlations yielded by this study would not be meaningful without agreement of ratings among judges (interrater reliability) and consistency of ratings within judges (intrarater reliability). The inter- and intrarater reliability for both types of ratings in this study were relatively high, considering that the psychologists were allowed to use the techniques they usually employ, instead of training in one particular method of interpretation. The relatively high agreement among the judges may be partially attributed to the types of criteria used by the majority of the judges. As noted in the literature review, the reliability of projective uses of HFDs is higher when global ratings or larger amounts of the drawing (e.g., the whole figure rather than the face) are used in interpre-

tation (Swensen, 1968). The judges in this study most frequently used global impressions and detail (which seemed to refer to overall amount of detail rather than specific details) in rating the drawings, which may have contributed greatly to the high interrater agreement.

Since the judges significantly agreed with each other in rating the drawings for artistic quality and emotional adjustment, the correlation between the two ratings can be interpreted more readily. The nonsignificant results of the average correlation between artistic quality and emotional adjustment ratings indicate that the two ratings may be measuring different dimensions of children's human figure drawings. The perceived artistic quality of the drawings in this study evidently did not influence the psychologists' ratings of emotional adjustment, at least not to a significant degree. This finding supports contentions by Goodenough (1926), Harris (1963), Koppitz (1968), and Machover (1949) that interpretations of drawings are not influenced by differences in children's drawing ability.

However, the finding of no significant relationship between artistic quality and emotional adjustment ratings contradicts the findings of studies by Cressen (1975), Sherman (1958), and Whitmyre (1953), who found that artistic quality and emotional adjustment ratings were related. One possible explanation for the discrepancy between the results of this study and past research is that Cressen

(1975), Sherman (1958), and Whitmyre (1953) used adults' drawings, while the present study used children's drawings. It has been noted that findings from adults' drawings cannot be generalized to children's drawings (Falk, 1981; Gittelman-Klein, 1978). Thus, the children's drawings in this study may have been qualitatively different from the adults' drawings in previous studies (e.g., developmental factors may have played an important role in ratings of artistic quality).

Another difference between this study and past research which may partially explain the contradictory results was the choice of raters making the artistic quality ratings. The research by Cressen (1975), Sherman (1958), and Whitmyre (1953) used artists to rate drawings for artistic quality and compared them to psychologists' ratings of emotional adjustment. In contrast, this study used the psychologists themselves for the artistic quality ratings. Psychologists may have a different idea of what constitutes artistic quality than do the artists; this difference may contribute to the contrasting results between this study and past research.

Since there was not a significant correlation between the artistic quality and emotional adjustment ratings, it is difficult to explain the similarities between the criteria used for the two ratings. Most of the criteria categories mentioned in making one type of rating were also listed in making the other rating. This finding supports a

study by McIntosh (1981), who found that the same basic set of criteria was used in the two ratings. In the present study, the psychologists often used the same types of criteria for making both artistic quality and emotional adjustment classifications (see Appendix H). Therefore, it would appear that the correlation between the ratings would be higher than the correlation found in this study.

Although they frequently used the same criteria, the judges in the present study may have interpreted the criteria differently for the two ratings. For example, "detail," the criterion most frequently mentioned in artistic quality ratings, may have been given more weight in artistic quality classifications than in rating drawings for emotional adjustment. Even though "detail" was the second most frequently used criterion in emotional adjustment ratings, it was cited more than twice as often in artistic quality ratings. Other criteria were also listed more often in one type of rating than the other and thus, appeared to be given more weight. For instance, "size" was listed 33 times in emotional adjustment ratings compared to 7 times in artistic quality ratings; "form" was mentioned 25 times in artistic quality ratings and only once in emotional adjustment ratings.

Although two major scoring systems (Koppitz, 1966a; Machover, 1949) have had considerable impact on the use of children's HFDs as a projective technique, only one rater cited the use of Koppitz's scoring system, and none men-

tioned the use of Machover's interpretations. However, several of the judges actually made diagnoses on the basis of the drawings which seemed to be consistent with Machover's interpretations, e.g., a large head signifying intellectualism. All of the judges reported some formal training on the projective use of HFDs through courses and/or workshops. The judges' formal training has not resulted in their adherence to a structured system of interpretation as a whole; rather, it seems that they use parts of systems or their own method of interpretation.

This finding raises the issue of the need for consistency with a particular theoretical basis of personality when measuring emotional adjustment through HFDs. For example, the use of Machover's DAP interpretations would be consistent with adherence to a psychoanalytic theory of personality, whereas the use of Koppitz's scoring system would be consistent with support of Sullivan's Interperonal Relationship Theory. Therefore, it does not appear logical for psychologists to use parts of Machover's system of interpretation, for instance, without adherence to a psychoanalytic theory of personality.

A surprising finding from this study was that four of the twelve judges used the Goodenough-Harris scoring system for ratings of emotional adjustment. This finding clearly represents an inappropriate use of the Goodenough-Harris scoring system, because it was intended to be an objective measure of intellectual maturity (Harris, 1963) rather than

a projective technique. If the Goodenough-Harris system were used in either of the two ratings, the more logical use would seem to be for artistic quality ratings (one rater did, in fact, use the Goodenough-Harris system for artistic quality ratings). The Goodenough-Harris system might be more appropriate for artistic quality classifications than emotional adjustment ratings because it involves the scoring of details which represent the likeness of a person. However, it does not measure the dimension of artistic quality singularly (e.g., drawings from children in the same general range of intelligence received varying ratings of artistic quality in this study).

Another surprising finding, which resulted from asking the psychologists to list methods/criteria used, was that diagnoses were occasionally made on the basis of the drawings. Although two of the judges specifically mentioned that drawings should not be used in isolation to diagnose emotional disturbance, they proceeded to make detailed diagnoses such as obsessive/compulsive tendencies, psychosomatic tendencies, and problems with sexual identity on the basis of the drawings they were asked to rate in this study. These two judges (and the three others who made diagnoses) may have responded in this manner because they misinterpreted the instructions and thought that they were expected to make diagnoses, or they may usually interpret drawings in this way in their practice.

The Issue of Validity

As was hypothesized, neither the artistic quality ratings nor the emotional adjustment ratings were significantly related to the actual emotional adjustment of the children making the drawings. The finding that artistic quality does not discriminate between adjusted and maladjusted individuals confirms studies by Lewinsohn (1965), Nichols and Strumpfer (1962), and Strumpfer and Nichols (1962), who found that ratings of artistic quality were unrelated to measures of adjustment and emotional disorder. Similarly, the results showing that ratings of emotional adjustment did not accurately identify actual adjustment are comparable to research by Diffenbach (1978), Pihl and Nimrod (1976), Snyder and Gaston (1970), Szasz, Baade, and Paskewicz (1980), Springer (1941), and Stolz and Coltherp (1961). These studies all found that HFD interpretations for emotional adjustment did not discriminate between adjusted and maladjusted children.

Thus, this study provides further evidence which questions the validity of the use of children's human figure drawings as a measure of emotional adjustment. When attempting to account for difficulties in the interpretation of HFDs for emotional adjustment, researchers such as Feldman and Hunt (1958), Roback (1968), Schaeffer (1964), and Strumpfer and Nichols (1962) have suggested that artistic quality of HFDs is a confounding influence. However, the results of this study indicated that psycho-

logists' ratings of children's HFDs for emotional adjustment were not significantly related to ratings of artistic quality from HFDs of the same children. In other words, the perceived artistic quality was not found to be a possible confounding influence on emotional adjustment interpetations of HFDs. Therefore, the validity of emotional adjustment interpretations appears even more questionable, since the findings of this study indicated that (a) emotional adjustment ratings were not related to actual emotional adjustment status, and (b) a proposed confounding influence (artistic quality of HFDs) was not supported.

The impact of this finding of questionable validity is even greater considering the frequency of use and emphasis placed on usage of HFDs projectively by the psychologists participating in this study. The judges reported using HFDs projectively in an average of 63 percent of their assessments, and they all indicated that they placed at least some emphasis on HFDs in making decisions about emotional adjustment. The frequent projective use of HFDs by the judges participating in this study may be representative of psychologists in general who employ projective techniques. As previously noted, Goh and Fuller (1983) and Goh, Teslow, and Fuller (1981) found that HFDs (through the Draw-A-Person, House-Tree-Person, and Kinetic Family Drawings) were among the most frequently used projective instruments for personality assessment. This relatively high frequency of usage makes the finding of questionable

validity of HFDs from this study significant.

In addition, if the emotional adjustment ratings are indicative of how psychologists rate drawings in their practice, the validity of HFDs is further decreased by the use of inappropriate scoring techniques (e.g., the Good-enough-Harris scoring system). It is seriously doubtful that the psychologists participating in this study were trained to use the Goodenough-Harris scoring system as a measure of emotional adjustment. It seems more likely that they were trained in the objective use of HFDs through the Goodenough-Harris scoring system and then generalized this training to projective uses of HFDs. This finding of the projective use of the Goodnough-Harris scoring system supports statements by Ysseldyke and Algozzine (1982):

> Diagnosticians regularly administer and use the results of tests for purposes other than those for which they were designed. At its very simplest level is the use of the results of a pupil's performance on a screening measure to make a classification or placement decision. More often, tests are used for more than one purpose, and in the process, they are used for purposes for which they were not designed (p. 137-138).

This issue of questionable validity of projective uses of HFDs has been debated by researchers such as Hammer (1969), who supported and defended HFDs as a projective

technique. Hammer contended that HFDs have been evaluated inappropriately because projective techniques cannot be appraised on the basis of whether or not they invariably focus on the same dimension in each subject. Furthermore, he criticized validity studies which involved "blind" interpretation (interpreting drawings without other information about the subject), because integration of findings from other sources of data is not possible with this approach. In defending his position, Hammer (1969) stated "No one uses the DAP alone. It was never intended by Machover, or anyone else, as anything more than a supplement, a graphic adjunct to the verbal technique" (p. 154).

In further defense of projective drawings, Hammer (1981) noted that correlations between ratings of traits reflected in drawings and personality characteristics of the subject are low mainly because two HFDs are a small sample of an individual's expressiveness. Hammer recommended the use of a battery of drawings including (a) the drawing of a house, tree, and person of each sex; (b) four crayon drawings--a house, tree, male and female person; (c) the Draw-A-Family procedure; (d) a drawing of an animal; (e) the drawing of the most unpleasant concept a subject can think of; and (f) other miscellaneous drawings. According to Hammer, this battery of drawings "would only then actually provide a pool of data sufficient to more validly 'test' projective drawings" (p. 179). However, this recommendation appears to be impractical, considering

the amount of time it would take a subject to complete the various drawings in an actual assessment.

Falk (1981) also criticized the use of only one drawing or a few drawings from each subject in research concerning the validity of HFDs. Like Hammer, Falk expressed the concern that HFD interpretation is used only as part of a whole diagnostic process, which generally includes other psychodiagnostic devices, a personal history, and the subject's behavior and associations to tests. Falk further noted that researchers have often concluded that the DAP, despite its limitations, is a useful tool when used in conjunction with other tests.

The issues concerning number of drawings and "blind" interpretation raised by researchers such as Hammer (1969, 1981) and Falk (1981) have also been addressed by researchers such as Martin (1983) and Wanderer (1969). While Hammer and Falk defended the projective use of HFDs, Martin (1983) sharply criticized this technique. In accordance with Hammer and Falk, Martin noted that a one-item test (i.e., one figure drawing) is inappropriate. The difference between Martin's position and that of the others is that he recognized that only one or a few drawings are used in actual practice, rather than other procedures such as the large battery of drawings suggested by Hammer. Hammer defends the validity of HFDs based on the use of many drawings, but research such as the present study and studies by Goh and Fuller (1983) and Goh, Teslow, and

Fuller (1981) has shown that in actual practice, psychologists use only one or a few drawings from each individual. Martin stated:

> Most psychologists would recognize that a research project on one subject could not be generalized to a population, or that a jury should not be constituted of only one juror. Yet these same psychologists may be tempted to interpret the small size of drawings as a sign of depression, lack of self-confidence, or "shrunken ego". In this case, generalizations are being made from one response. It is a situation exactly analagous to that of a one item test and therefore the response and its interpretation is inherently unreliable. It is therefore inherently invalid (p. 6).

Interestingly, Martin's statement appears to be of particular importance to certain findings of this study. As noted earlier, five of the twelve psychologists participating in this study volunteered specific interpretations (very similar to those listed by Martin) on the basis of one drawing.

The issue of "blind" interpretation of drawings in research criticized by Hammer (1969) and Falk (1981) has also been debated. Wanderer (1967) concluded that clinicians may attibute knowledge obtained from interviews and observation to the drawings themselves, "seeing" in them

what they already knew. According to Wanderer, "blind" studies eliminate this hazard. As noted in the literature review of this study, Chapman and Chapman (1967) have shown that this phenomenon--observing in drawings "illusory correlates" of known symptoms--results in systematic errors on the part of judges. Therefore, it appears that "blind" studies are necessary in order to objectively measure the validity of projective uses of HFDs.

Through the position of researchers such as Falk (1981), it can still be argued that the validity of HFDs should not be evaluated in this strict manner because they are used only as a source of hypotheses that will be supported or rejected by the outcome of other procedures in the assessment. (The majority of the judges in this study reported using HFDs in this manner). However, Martin (1983) challenged this contention by stating several reasons why the use of the DAP to support other data is inappropriate. First, he noted that adding one unreliable measure (the DAP) to other reliable measures (such as standardized rating scales and test scores) serves only to decrease the reliability and validity of the entire set of measures.

Second, according to Martin (1983), it is inappropriate to use HFDs because "interpretations of any one index on the Draw-A-Person are so ambiguous and manifold that virtually any hypothesis could be supported from such data" (p. 6). This statement also seems applicable to the

present study, due to the finding that judges frequently used global impressions and overall detail as criteria for rating the HFDs. These two criteria seem to be vague, and as a result, they presumably could be used to support differing hypotheses.

Third, Martin (1983) concluded that HFDs should not be used to support other data because they can be the basis for the formation of a strongly held hypothesis, or they can reinforce a bias or stereotype held by the psychologist. Either of these occurrences might lead the psychologist to search for supportive data. Martin is in effect describing the "illusory correlation" found in research by Chapman and Chapman (1967). Wanderer (1969) further addressed this phenomenon, suggesting that the DAP is popular despite of questionable validity because it reinforces clinicians' beliefs.

One final issue regarding the questionable validity of projective uses of HFDs is the difficulty involved in accounting for variability in children's drawings. Hammer (1981) stated that emotionally disturbed individuals may have distorted views of the world which are sampled by projective drawings. However, according to Falk (1981), children labeled as emotionally disturbed may be facing the same conflicts as normal children. It is the manner in which emotionally disturbed children deal with these conflicts which results in their being labeled; their responses to conflicts may differ only in degree, not in

type of response. While drawings from children at the extremes of the normal-abnormal continuum may definitely reflect differences in how they cope with conflicts, those near the middle of the continuum may not reflect differences in adjustment. It is not safe to assume that drawings from normal children will not show any signs of stress or conflicts; likewise, it is not safe to assume that drawings from emotionally disturbed children will reflect conflicts.

Selfe (1983) expressed this viewpoint in noting that individual drawing features may be the result of instruction, imitation of cartoons, or other influences, instead of arising from the emotional state of the subject alone. She stated that "the problem is to bring the emotional, expressive motivation for drawing into some significance, while taking account of cognitive and developmental explanations" (p. 23).

In conclusion of this discussion on the issue of validity, several points may be emphasized. First, the recommendation of using many drawings from each child in validity research is not appropriate because this procedure is not used in actual practice and does not seem feasible for future use. On the other hand, it appears that the use of one drawing or a few drawings in actual practice is such a small sample that it is an inherently unreliable, and thus invalid technique. Second, "blind" interpretation seems to be necessary for research on HFD validity in order

to avoid "illusory correlation" or bias in interpretation. But in actual practice, illusory correlation may occur when psychologists "see" in drawings what they already know about the subject or seek to confirm later. Third, due to these problems and other difficulties in interpretation (e.g., accounting for the variability within ED vs. non-ED children's HFDs), the projective use of children's HFDs as a measure of emotional adjustment appears to be invalid. This conclusion is supported by the findings of this study.

Limitations of This Study

The present study involved a sample of ten children diagnosed as emotionally disturbed and ten normal children, each of whom produced two human figure drawings. These children were male, in the average range of intelligence, and ranged in age from eight to ten years. Therefore, the findings of this study cannot be generalized to the drawings of females and those of children of different ages and ranges of intelligence.

Another possible limitation was that significant differences may have existed between the two drawings of each child in the study. One drawing was rated for emotional adjustment, while the other drawing was rated for artistic quality. This procedure was used in order to help prevent certain features of a drawing from biasing the opposite rating. It is unlikely that significant differences between two drawings from the same child existed or influenced results, since a counterbalancing procedure was

used for rating drawings and a relatively high degree of interrater reliability was found.

The issue of the validity of the emotional adjustment status of the children making the drawings should also be considered a limitation of this study. In randomly selecting the emotionally disturbed children, care was taken to select those who had been diagnosed as emotionally disturbed and placed in ED classes. This procedure does not necessarily mean the children were actually emotionally disturbed (the validity of the diagnoses could not be determined), but it can be concluded that the behavior of these children was significantly different from others to warrant placement in a special class. Likewise, it cannot be totally ruled out that the randomly selected normal children were not ED to a degree, but they showed no evidence of emotional disturbance as revealed from their teachers and cumulative records.

Suggestions for Future Research

Based on the results of this study and the review of related literature, several suggestions may be made for future research. First, the possible influence of artistic quality on HFD interpretations for emotional adjustment cannot be ruled out by this one study. Further research is needed in order to determine if artistic quality of drawings can be eliminated as a hypothesized confounding influence on projective HFD interpretations. In addition to replication of this study using the same age group,

intelligence level of subjects, etc., research needs to be conducted using drawings by females and by those in other age and intelligence ranges.

Second, the focus of additonal research on the projective uses of HFDs needs to change from that of proving or disproving validity through diagnostic sorting tasks. It seems evident from the present study and past research that, in general, HFDs appear to be invalid for indicating level of emotional adjustment. Thus, future research needs to focus on whether the use of HFDs is in any way helpful as part of an assessment battery. For instance, further studies need to be conducted on the usefulness of HFDs through Kinetic Family Drawings (e.g., does drawing his or her family stimulate a child to talk about the family?). It may be that the only appropriate uses of HFDs are a rapport building activity (Martin, 1983) and a screening instrument for those of below average intelligence (Scott, 1981), but these conclusions should be substantiated by methodologically sound research using children's drawings rather than those of adults.

Third, another possible avenue for future research is the determination of why the continued projective use of HFDs is so popular in the face of nonsupportive or inconclusive research on this technique. Research on Wanderer's (1969) contention that the use of HFDs partially reinforces clinician's beliefs about children might prove

interesting. Analyses of current formal training methods on projective uses of HFDs (e.g., courses and workshops) might also help in determining why HFDs are used inappropriately.

Summary

Research on the validity of the projective use of children's human figure drawings as a measure of emotional adjustment has been inconclusive. As a reason for the inconsistent findings, researchers have hypothesized that the artistic quality of HFDs may be a confounding influence on interpretation of adjustment from the drawings (e.g., drawings of low artistic quality might erroneously be perceived as having been drawn by maladjusted children). Past research on this issue has been inadequate due to the fact that adults' drawings were used instead of children's, and other judges (i.e., artists) were used to rate artistic quality rather than psychologists.

The major purpose of this study was to determine if a relationship existed between psychologists' judgments of artistic quality and judgments of emotional adjustment from children's human figure drawings. Two human figure drawings each were collected from a randomly selected group of emotionally disturbed children and normal children. These children were matched according to age (eight to ten years), sex (all were male), and IQ (average range). The ED children were diagnosed by their school systems or in clinical settings and were placed in ED classes, whereas

the normal children showed no evidence of emotional disturbance.

Twelve psychologists (six male and six female) were randomly selected to rate the drawings for artistic quality and emotional adjustment without knowledge of the children's emotional adjustment status. These psychologists reported using children's HFDs projectively in an average of 63 percent of their assessments. Their average length of use of children's HFDs was 7 years. All of the psychologists indicated that they placed at least some emphasis on children's HFDs in making decisions about emotional adjustment.

A counterbalancing procedure was employed for the order in which the psychologists rated the drawings (i.e., half rated artistic quality first and half rated emotional adjustment first). Artistic quality was defin l as the technical accuracy of the drawing, rather than its aesthetic appeal. The definition of emotional disturbance from Public Law 94-142 was suggested as a guideline for the emotional adjustment ratings. Otherwise, the psychologists were allowed to employ the methods of interpetation they use in their practice. Each psychologist was also asked to list or describe the methods or criteria used in rating each drawing. In addition, one month after the original ratings were made, the psychologists were asked to re-rate a random sample of the drawings in order to determine the intrarater reliability.

The analyses of these ratings addressed (a) the relationship between the artistic quality and emotional adjustment ratings, (b) the similarities and differences between the criteria used for making both types of ratings, (c) the relationship between the emotional adjustment ratings and actual emotional adjustment status, (d) the relationship between the artistic quality ratings and actual emotional adjustment status, and (e) the degree of agreement of ratings among judges (interrater reliability) and the degree of stability of ratings (intrarater reliability).

It was hypothesized that (a) there would be a significant positive relationship between psychologists' judgments of artistic quality and judgments of emotional adjustment from children's HFDs, (b) the criteria or methods used for ratings of artistic quality would be similar to those used for ratings of adjustment, and (c) ratings of artistic quality and emotional adjustment would have a low level of identification of actual level of adjustment.

A positive, but nonsignificant, point biserial correlation ($r = .37$) was found between the artistic quality ratings and emotional adjustment ratings, indicating that the two ratings may be measuring different dimensions of children's HFDs. The psychologists' perceptions of artistic quality of the HFDs in this study did not influence their ratings of emotional adjustment to a significant

degree. Thus, the first hypothesis of a high positive relationship was not supported.

A cumulative frequency analysis of the criteria used by the psychologists in classifying drawings indicated that the same types of criteria were frequently cited for both types of ratings, lending support to the second hypothesis. Since the ratings were not highly correlated, the criteria were presumably interpreted differently in the two types of ratings. For example, some criteria, such as "detail" and "size," were listed more frequently in one type of rating than the other and may have been given more weight by the judges.

A relatively high degree of interrater reliability was found for the artistic quality ratings (intraclass r = .86) and the emotional adjustment ratings (intraclass r = .75). The intrarater reliability was also relatively high (the average correlation coefficent for the artistic quality ratings was .90, and the average percent agreement for the emotional adjustment ratings was 88 percent).

However, the emotional adjustment ratings were not significantly related to the actual emotional adjustment status of the children making the drawings (average percent agreement = 58 percent). This finding indicated that the validity of the methods used by the psychologists in interpreting the HFDs for emotional adjustment was questionable. The artistic quality ratings were also not significantly related to the children's actual adjustment (r =

.09), indicating that emotionally disturbed children's HFDs were not necessarily perceived as having low artistic quality.

In addition to these analyses, other unexpected findings from this study included the inappropriate use of the Goodenough-Harris scoring system as a projective technique, and diagnoses such as obsessive/compulsive tendencies made on the basis of the drawings by some of the judges in this study. These findings suggest that HFDs are frequently used inappropriately as a projective technique.

Overall, the results of this study indicate that the validity of emotional adjustment interpretations from children's HFDs appears questionable because (a) the emotional adjustment ratings were not related to actual emotional adjustment status, (b) a proposed confounding influence (artistic quality of HFDs) was not supported, and (c) inappropriate techniques for interpretation were found to be used. The conclusion of questionable validity was further emphasized by the frequency of use and emphasis placed on HFDs in making decisions about emotional adjustment reported by the psychologists participating in this study. Thus, this study contributes to the body of research suggesting that children's HFDs are not valid or appropriate for indicating level of emotional adjustment of elementary school aged children.

References

Adler, P. (1970). Evaluation of the figure drawing technique: Reliability, factoral structure, and diagnostic usefulness. Journal of Consulting and Clinical Psychology, 35, 52-57.

Albee, G.W., & Hamlin, R.M. (1949). An investigation of the reliability and validity of judgments of adjustment inferred from drawings. Journal of Clinical Psychology, 5, 389-392.

Albee, G.W., & Hamlin, R.M. (1950). Judgment of adjustment from drawings: The applicability of rating scale methods. Journal of Clinical Psychology, 6, 363-365.

Arkell, R.N. (1976). Naive prediction of pathology from children's human figure drawings. Journal of School Psychology, 14, 114-117. (a).

Arkell, R.N. (1976). Adults' representation of children's human figure drawings. Perceptual and Motor Skills, 43, 458. (b)

Bolander, K. (1977). Assessing personality through tree drawings. New York: Basic Books, Inc.

Buck, J. (1948). The H-T-P technique: A qualitative and quantitative scoring manual. Journal of Clinical Psychology, 4, 317-396.

Burns, R.C., & Kaufman, S.H. (1970). Kinetic family drawings (KFD). New York: Brunner-Mazel.

Burton, A., & Sjoberg, B. (1964). The diagnostic validity of human figure drawings in schizophrenics. Journal of Clinical Psychology, 57, 3-18.

Cauthen, N.R., Sandman, C.A., Kilpatrick, D.G., & Deabler, H.L. (1969). DAP correlates of Sc scores on the MMPI. Journal of Projective Techniques and Personality Assessment, 33, 262-264.

Chapman, L.J., & Chapman, J.P. (1967). Genesis of popular but erroneous psychodiagnostic observations. Journal of Abnormal Psychology, 72, 193-204.

Cressen, R. (1975). Artistic quality of drawings and judges' evaluations of the DAP. Journal of Personality Assessment, 39, 132-137.

Diffenbach, E.W. (1978). Koppitz' human figure drawing test: The reliability and clincal validity of its emotional indicators. Dissertation Abstracts International, 38, 6053.

Eno, L., Elliott, C., & Woehlke, P. (1981). Koppitz emotional indicators in the human figure drawings of children with learning problems. Journal of Special Education, 15, 459-470.

Falk, J.D. (1981). Understanding children's art. Journal of Personality Assessment, 45, 465-472.

Feldman, M.H., & Hunt, R.G. (1958). A relation of difficulty in drawing and ratings of adjustment based on human figure drawings. Journal of Clinical Psychology, 22, 217-220.

Fisher, S., & Fisher, R. (1950). Tests of certain assumptions regarding figure drawing analysis. Journal of Personality Assessment, 9, 132-137.

Frankenburg, W.K., & Dodds, J.B. (1975). The Denver Developmental Screening Test-Revised. Denver: LA DOCA.

Fuller, G.R., Preuss, M., & Hawkins, W.F. (1970). The validity of the human figure drawing with disturbed and normal children. Journal of School Psychology, 3, 54-56.

Gittelman-Klein, R. (1978). Validity of projective tests for psychodiagnosis in children. In R.I. Spitzer & D.F. Klein (Eds.). Critical issues is psychiatric diagnosis. Raven Press: New York.

Goh, D.S. & Fuller, G.B. (1983). Current practices in the assessment of personality and behavior by school psychologists. School Psychology Review, 12, 240-249.

Goh, D.S., Teslow, C.J., & Fuller, G.B. (1981). The practice of psychological assessment among school psychologists. Professional Psychology, 12, 696-706.

Goldman, R.K., & Velasco, M.M. (1980). Toward the development of a rational scale in the use of human figure drawings as a kindergarten screening measure. Perceptual and Motor Skills, 50, 571-577.

Goldman, R.K., & Warren, D. (1976). Preliminary analyses of human figure drawings as a kindergarten screening measure. Perceptual and Motor Skills, 43, 58.

Goodenough, F.L. (1926). Measurement of intelligence by drawings. Chicago: World Book Co.

Hall, L.P., & Ladriere, M.L. (1970). A comparative study of diagnostic potential and efficiency of six scoring systems applied to children's human figure drawings. Psychology in the Schools, 7, 244-247.

Hammer, E.F. (1954). Guide for qualitative research with the H-T-P. Journal of Genetic Psychology, 51, 41-60.

Hammer, E.F. (1969). DAP: Back against the wall? Journal of Consulting and Clinical Psychology, 33, 151-156.

Hammer, E.F. (1981). Projective drawings. In A.I. Springer (Ed.). Assessment with projective techniques. New York: Springer Publishing Company.

Hammer, M., & Kaplan, A.M. (1966). The reliability of children's human figure drawings. Journal of Clinical Psychology, 22, 316-319

Harris, D.B. (1963). Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace & World, Inc.

Hendricks, R., Balzer, W., and Sheehy, J. (1980, April). Comparison of methods for estimating the reliability of nominal data. Paper presented at the 51st Session of the Eastern Psychological Association, Hartford, CT.

Hiler, E.W., & Nesvig, D. (1965). An evaluation of criteria used by clinicians to infer pathology from figure drawings. Journal of Consulting Psychology, 29, 520-529.

Hobbs, N. (1975). The futures of children. San Fransisco: Jossey-Bass.

Holzberg, J.D., & Wechsler, M. (1950). The validity of human form drawings as a measure of personality deviation. Journal of Projective Techniques, 14, 343-361.

Hulse, W.C. (1951). The emotionally disturbed child draws his family. Quarterly Journal of Child Behavior, 3, 152-174.

Jolles, I. (1952). A catalogue for the qualitative interpretation of the H-T-P. Beverly Hills, Calif.: Western Psychological Services.

Jue, R.W. (1976). The interpretation of normal and maladjusted children's drawings by psychologists and non-psychologists. Dissertation Abstracts International, 37, 78.

Kay, S.R. (1978). Qualitative differences in human figure drawings according to schizophrenic subtype. Perceptual and Motor Skills, 47, 923-932.

Klepsch, M., & Logie, L. (1982). Children draw and tell: An introduction to the projective uses of children's human figure drawings. New York: Brunner-Mazel, Inc.

Klopfer, W.G., & Taulbee, E.S. (1976). Projective tests. In M.R. Rosensweig & L.M. Border (Eds.). Annual review of psychology (Vol. 27). Palo Alto: Annual Reviews, Inc.

Koppitz, E. (1966). Emotional indicators on human figure drawings of children: A validation study. Journal of Clinical Psychology, 22, 313-315. (a)

Koppitz, E. (1966). Emotional indicators on human figure drawings of shy and aggressive children. Journal of Clinical Psychology, 2, 481-483. (b)

Koppitz, E. (1966). Emotional indicators on human figure drawings and school achievement of first and second graders. Journal of Clinical Psychology, 23, 81-83. (c)

Koppitz, E. (1968). Psychological evaluation of children's human figure drawings. New York: Grune & Stratton.

Leichtman, S.R. (1979). Adults' simulations of preschoolers' human figure drawings. Perceptual and Motor Skills, 49, 18.

Lewinsohn, P.M. (1965). Psychological correlates of overall quality of figure drawings. Journal of Consulting Psychology, 29, 501-512.

Lindzey, G. (1961). Projective techniques and cross-cultural research. New York: Appleton-Century-Crofts.

Lingren, R.H. (1971). An attempted replication of emotional indicators in human drawings by shy and aggressive children. Psychological Reports, 29, 35-38.

Lubin, B., Wallis, R.P., & Paine, C. (1971). Patterns of psychological test usage in the United States: 1935-1969. Professional Psychology, 2, 70-74.

Machover, K. (1953). Human figure drawings of children. Journal of Projective Techniques, 17, 53-92.

Machover, K. (1949). Personality projection in the drawing of the human figure. Springfield, Ill.: Charles C. Thomas.

Maloney, M.P., & Glasser, A. (1982). An evaluation of the clinical utility of the Draw-A-Person test. Journal of Clinical Psychology, 38, 183-190.

Martin, R.P. (1983). The ethical issues in the use and interpretation of the draw-a-person test and other similar projective techniques. The School Psychologist, 38, 6-8.

McCarthy, D. (1972). McCarthy scales of children's abilities. New York: The Psychological Corporation.

McIntosh, E.M. (1981). Sources of diagnostic error in detection of maladjusted children using human figure drawings. Dissertation Abstracts International, 42, 1182.

McNemar, Q. (1969). Psychological statistics. New York: John Wiley and Sons, Inc.

Murray, D.C., & Deabler, H.F. (1958). Drawings, diagnoses, and the clinician's learning curve. Journal of Projective Techniques, 22, 415-420.

Nichols, R., & Strumpfer, D. (1962). A factor analysis of draw-a-person test scores. Journal of Consulting Psychology, 26, 156-161.

Pihl, R.O., & Nimrod, G. (1976). The reliability and validity of the draw-a-person test in IQ and personality assessment. Journal of Clinical Psychology, 32, 470-472.

Reiznikoff, N.A. & Nichols, A.L. (1958). An evaluation of human figure drawing indicators of paranoid pathology. Journal of Consulting Psychology, 22, 395-397.

Ribler, R.I. (1957). Diagnostic prediction from emphasis on the eye and ear in human figure drawings. Journal of Consulting Psychology, 21, 223-225.

Ries, H., Johnson, M.H., Armstrong, H., & Holmes, D. (1966). The draw-a-person test and process reactive schizophrenia. Journal of Projective Techniques and Personality Assessment, 30, 184-186.

Roback, H. (1968). Human figure drawings: Their utility in the clinical psychologist's armamentarium for personality assessment. Psychological Bulletin, 70, 1-10.

Royal, R.E. (1949). Drawing characteristics of neurotic patients using a drawing-of-a-man-and-woman technique. Journal of Clinical Psychology, 5, 392-395.

Sattler, J.M. (1982). Assessment of children's intelligence and special abilities. Philadelphia: W.B. Saunders.

Schaeffer, R.W. (1964). Clinical psychologists' ability to use the draw-a-person test as an indicator of personality assessment. Journal of Consulting Psychology, 28, 382.

Schmidt, L.D., & McGowan, J.F. (1959). The differentiation of human figure drawings. Journal of Consulting Psychology, 23, 129-133.

Scott, L.H. (1981). Measuring intelligence with the Good-enough-Harris drawing test. Psychological Bulletin, 83, 483-505.

Selfe, L. (1983). Normal and anomalous representational drawing ability in children. New York: Academic Press.

Sherman, L.J. (1958). The influence of artistic quality on judgments of patient and nonpatient status from human figure drawings. Journal of Projective Techniques and Personality Assessment, 22, 338-340.

Snyder, R.T. & Gaston, D.S. (1970). The figure drawing of the first grade child - item analysis and comparison with Koppitz norms. Journal of Clinical Psychology, 26, 377-383.

Springer, N.N. (1941). A study of the drawings of maladjusted and adjusted children. Journal of Genetic Psychology, 58, 131-138.

Stolz, R.T. & Coltherp, F.C. (1961). Clinical judgments and the draw-a-person test. Journal of Consulting Psychology, 25, 43-45.

Stricker, G. (1967). Actuarial, naive clinical, and sophisticated clinical prediction from figure drawings. Journal of Consulting Psychology, 31, 492-494.

Strumpfer, D. & Nichols, R. (1962). A study of some communicable measures for the evaluation of human figure drawings. Journal of Projective Techniques, 26, 342-353.

Sundberg, N.D. (1961). The practice of psychological testing in clinical services in the United States. American Psychologist, 16, 79-83.

Sundberg, N.D. (1977). Assessment of persons. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Swensen, C.H. (1957). Empirical evaluations of human figure drawings. Psychological Bulletin, 54, 431-466.

Swensen, C.H. (1968). Empirical evaluations of human figure drawings - 1957-1966. Psychological Bulletin, 70, 20-44.

Szasz, C.W., Baade, L.E., & Paskewicz, C.W. (1980). Emotional and developmental aspects of human figure drawings in predicting school readiness. Journal of School Psychology, 18, 67-73.

Vane, J.R. & Eisen, V.W. (1962). The Goodenough draw-a-man test and signs of maladjustment in kindergarten children Journal of Clinical Psychology, 18, 276-279.

Wagner, M.E. & Shubert, H.J. (1955). D.A.P. quality scale for late adolescents and young adults. Kenmore, N.J.: Delaware Letter Shop.

Wanderer, Z.W. (1969). Validity of clinical judgments based on human figure drawings. Journal of Consulting and Clinical Psychology, 33, 143-150.

Watson, C.G. (1967). Interjudge agreement of draw-a-person diagnostic impressions. Journal of Projective Techniques and Personality Assessment, 31, 42-45. (a)

Watson, C.G. (1967). Relationship of distortion to draw-a-person diagnostic accuracy among psychologists. Journal of Consulting Psychology, 31, 142-146. (b)

Whitmyre, J. (1953). The significance of artistic excellence in the judgment of adjustment inferred from human figure drawings. Journal of Consulting Psychology, 17, 421-424.

Winer, B.J. (1971). Statistical principles in experimental design. New York: McGraw-Hill.

Ysseldyke, J.E. & Algozzine, B. (1982). Critical issues in special and remedial education. Boston: Houghton Mifflin Company.

Appendixes

Appendix A

Correspondence to Judges

WESTERN KENTUCKY UNIVERSITY

BOWLING GREEN, KENTUCKY 42101

Department of Psychology

In our recent telephone conversation, I asked for your assistance in rating children's human figure drawings for my Master's thesis at Western Kentucky University. Enclosed is a series of children's human figure drawings which you are requested to rate according to the instructions given. The drawings are divided into Section 1 and Section 2. Please complete Section 1 before beginning Section 2. After completing both sections, please return the drawings and ratings to me in the enclosed postage paid envelope.

Please try to return the drawings to me in two weeks if possible. If you have questions at any time, please call me at (502) 745-2695 or 782-0551. A brief follow-up to the ratings and information for, which are necessary to complete the study, will be sent to you later. A copy of the purpose and results of this study will also be sent to you upon completion.

Thank you very much for your cooperation. I really appreciate your time and effort in helping me with my thesis.

Sincerely,

Karen Collier
Psychologist-in-training

Dr. William Pfohl
Supervising Psychologist

WESTERN KENTUCKY UNIVERSITY

BOWLING GREEN, KENTUCKY 42101

Department of Psychology

Thank you for your cooperation in rating children's human figure drawings for my Master's thesis. I greatly appreciate the time and effort you put forth in rating the drawings.

In order to complete the study, I need to obtain a measure of the consistency of each psychologist's ratings. Enclosed is a sample of drawings which should be rated according to the same instructions as the original ratings. I also need you to fill out the enclosed information form, which will complete you role in the data collection for this thesis. After completing the ratings and form, please return them to me as soon as possible in the enclosed postage paid envelope.

Again, if you need to contact me for any reason, please call (502) 745-2695 or 782-0551. A copy of the purpose and results of the study will be sent to you immediately upon completion.

Thanks again.

Sincerely,

Karen Collier
Psychologist-in-training

Appendix B

*Instructions for Section 1

A series of 20 children's human figure drawings is attached. All of the children who made these drawings are male and are in the average range of intelligence. The age of each child is listed at the top of each drawing. A rating sheet is also attached to each drawing.

Please rate each drawing individually according to the level of emotional adjustment best reflected in the drawing. The definition of emotional disturbance from Public Law 94-142 should be used as a guideline.

Please indicate on the rating sheet if the drawing best reflects normal emotional adjustment or maladjustment (emotional disturbance). Then, list or describe on the rating sheet, as you rate each drawing, the methods or criteria you used for determining categorization. For example, this could be a scoring system, your own unique system of interpreting drawings, global impressions, and content, detail or other elements in the drawing. Please be as precise as possible in indicating the criteria used.

Please rate each drawing and list the criteria used without referring to the other drawings, i.e., complete the ratings and listings for the first drawing before going on to the second and so on. Please do not make comparisons between drawings.

After completing the ratings in this section, please place the drawings and the attached rating sheets in the

enclosed return envelope and go on to Section 2.

*Instructions for Section 1 and 2 were reversed for those psychologists making artistic quality ratings first and emotional adjustment ratings second.

Appendix C

Instructions for Section 2

Twenty different children's human figure drawings make up Section 2. These children are also male and in the average range of intelligence. The age of each child is listed at the top of the drawings, and a rating sheet is attached to each drawing.

Please rate each drawing individually according to the artistic quality reflected in the drawing. Artistic Quality is defined as the goodness or technical accuracy of the drawing, i.e., how well it represents a person, rather than its aesthetic appeal. Please indicate if each drawing is of high, medium or low artistic quality on the rating sheet attached to each drawing. Next, further differentiate the artistic quality of each drawing by rating the drawing on a scale of 1 to 7, which a rating ot 1 being the lowest quality and a rating of 4 being the highest quality. (Ratings of 1 and 2 correspond to low quality, ratings 3, 4, and 5 correspond to medium quality, and ratings 6 and 7 to high quality.)

Then, as you rate each drawing, please list or describe on the rating sheet the methods or criteria you used in determning each categorization. For example, this could be content, detials, or other elements in the drawing, or global impressions. Please be as precise as possible in listing or describing what criteria from each drawing resulted in its classification in a particular

category.

Please rate each drawing and list the criteria used without referring to the other drawings, i.e., complete the ratings and listings for the first drawing before going on to the second drawing and so on. Please do not make comparisons between the drawings.

When you have completed this section, please place it in the return envelope containing Section 1 and mail it to me.

Appendix D

Information Form

*Name _____

Title or Position _____

Please describe any training (for example, coursework or workshops) you may have received in the use of children's human figure drawings (HFDs) as a projective technique

_____

_____

How do you use children's HFDs projectively as part of your assessments? _____

_____

_____

How long have you used children's HFDs as a projective technique? _____

In what percentage of assessments do you use children's HFDs projectively? _____

How much emphasis do you place on HFDs in making decisions regarding level of emotional adjustment?

1 2 3 4 5 6 7

No Emphasis Great Emphasis

Comments: _____

_____

_____

*Your name will be kept totally confidential

Appendix E

Artistic Quality Rating Sheet

Drawing No.___

Please circle

low			medium		high	
1	2	3	4	5	6	7

Please list or describe methods or criteria used in classifying this drawing_____

_____

_____

_____

_____

Appendix F

Emotional Adjustment Rating Sheet

Drawing No.___

Please circle

normal adjustment maladjustment

Please list or describe methods or criteria used in classifying this drawing_____

_____

_____

_____

_____

Appendix G

Individual Judges' Correlations Between Artistic Quality and Emotional Adjustment Ratings

Judge	r	Judge	r
1	.39	7	.71
2	-.22	8	.09
3	.41	9	.64
4	.21	10	.60
5	.46	11	.35
6	.33	12	.17

Appendix H

Cumulative Frequency of Criteria Common to Both Ratings

Criteria	Emotional adjustment *n	Artistic quality *n
Detail	49	110
Global impressions	59	43
Proportion	18	74
Size	33	7
Placement	22	6
Content	21	9
Quality	20	10
Facial expression/features	15	17
Immaturity	13	7
Line quality	3	21
Activity/action content	3	11
Integration	6	13
Shading	9	6
Non-human	4	9
Clothing	8	7
Primitive	5	7
Pencil control	1	7

table continues

Criteria	Emotional adjustment n	Artistic quality n
General Mood	1	2
Symmetry	1	4
Rigid Stance	10	5
Transparency	5	3
Erasures	7	3
Horizon	2	2
Cartoon figure	2	2
Realistic/unrealistic	1	7
Stick figure	2	2
Partial profile	1	1
All parts present	8	4
Form	1	25
Good for age	1	2
Poor for age	1	2
Simplified	1	1
Arms	21	9
Teeth/mouth	6	1
Hands	7	4
Neck	6	3

table continues

Criteria	Emotional adjustment n	Artistic quality n
Hair	1	8
Head	9	5
Trunk	7	4
Shoulders	4	3
Feet	7	3
Ears	4	2
Legs	1	1
Fingers	1	1

*This number represents the total number of times the criterion was listed out of a total of 240 opportunities (12 psychologists x 20 drawings)

Appendix I

Cumulative Frequency of Criteria Unique to Each Rating

Emotional Adjustment		Artistic Quality	
Criteria	*n	Criteria	*n
Distortion	10	Structure	13
Midline emphasis	6	Basic features	5
Grotesque	4	Composition	5
Message	2	Sureness of stroke	4
Bizarre content	2	Shape	4
Opposite sex drawing	2	Texturing	3
Body concept	2	Creativity	3
Violent scene	2	Unappealing	2
Age of character	1	"Blob" body	2
Eyes	4	Originality	1
Nose	3	Contour of body	1
Muscles	2	Compressed	1
Emotional indicators	1	One dimensional	1
Omissions	1	Body parts misplaced	1
Tense	1	Eyebrows	1
Accessories (flowers)	1	Strange	1
Rear view	1	Hatch marks	1
Base line	1	Sketching	1

table continues

Emotional Adjustment Criteria	n	Artistic Quality Criteria	n
Object in hand	1	Poor closure skills	1
Constriction	1	Balance	2
Boy or girl?	1		

*This number represents the total number of times the criterion was listed out of a total of 240 opportunities (12 judges x 20 drawings)

Appendix J

Individual Judges' Correlations Between Artistic Quality Ratings and Actual Emotional Adjustment Status

Judge	r	Judge	r
1	.14	7	.16
2	.22	8	.10
3	.12	9	-.06
4	.22	10	.03
5	.20	11	-.18
6	-.03	12	.10

Appendix K

Individual Judges' Percent Agreement Between Emotional Adjustment Ratings and Actual Emotional Adjustment Status

Judge	Percent Agreement	Judge	Percent Agreement
1	60	7	60
2	35	8	65
3	60	9	45
4	60	10	60
5	65	11	65
6	65	12	55

Appendix L

Intrarater Reliability: Correlations Between Artistic Quality Ratings

Judge	r	Judge	r
1	.94	7	.87
2	.62	8	.71
3	.90	9	.98
4	.89	10	.85
5	.95	11	.76
6	.91	12	.98

Appendix M

Intrarater Reliability: Percent Agreement Between Emotional Adjustment Ratings

Judge	Percent Agreement	Judge	Percent Agreement
1	70	7	90
2	70	8	80
3	100	9	100
4	90	10	100
5	100	11	90
6	90	12	70