



1977

Convergent and Discriminant Validity of the Koppitz Scales and a Hiler-Nesvig Formula with Children's Human Figure Drawings

Roger William Semyck
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses



Part of the [Psychology Commons](#)

Recommended Citation

Semyck, Roger William, "Convergent and Discriminant Validity of the Koppitz Scales and a Hiler-Nesvig Formula with Children's Human Figure Drawings" (1977). *Master's Theses*. 2941.
https://ecommons.luc.edu/luc_theses/2941

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).
Copyright © 1977 Roger William Semyck

CONVERGENT AND DISCRIMINANT VALIDITY OF THE
KOPPITZ SCALES AND A
HILER-NESVIG FORMULA WITH
CHILDREN'S HUMAN FIGURE DRAWINGS

by

Roger Semyck

A Thesis Submitted to the Faculty of the Graduate School
of Loyola University of Chicago in Partial Fulfillment
of the Requirements for the Degree of
Master of Arts

February

1977

ACKNOWLEDGMENTS

The author wishes to express his sincere gratitude to the members of his thesis committee, Dr. James Johnson, chairman, and Dr. Emil Posavac. These psychologists have given generously of their knowledge, interest, and time toward the completion of this research. The author also wishes to thank Dr. Patricia Barger, Dr. William Hunt, Dr. Thomas Petzel, Dr. Ron Walker, and especially Dr. Frank Kobler for encouragement and critical comments on the proposal. The author is indebted to the help of the principals and teachers of four schools: Mr. Ackley, Mrs. Christopher, Mr. Galetano, and Mrs. Gross of Feehanville School, School District 26; Mr. Bonhivert, Mrs. Lamagdaline, Mrs. Berg, Miss Mullen, Mrs. Peaco, and Mrs. Mannisto of Mechanics Grove School, School District 75; Sister Clarice, Mrs. Burke, and Mrs. Young of St. Joseph's School, Libertyville; and Nancy Buckler, Master Teacher of Loyola University Day School. In addition, the author wishes to express his appreciation to the students who acted as drawing raters: Brenda Gelman, Eileen Guardalabene, John Hinkle, Judy Lechert, Jacob Messing, and Matthew Zarantonello. While the author gratefully acknowledges the help of these several individuals, he recognizes that he is solely responsible for any shortcomings in the work.

VITA

Roger William Semyck was born on September 3rd, 1945 in Chicago, Illinois. The author is the son of Richard William Semyck and Carol (McAvoy) Semyck.

His elementary school education was obtained at St. Catherine of Siena, Oak Park, and at St. Mary of the Woods, Chicago. Upon graduation from Loyola Academy in June of 1963, he enrolled in an engineering program at the University of Notre Dame. He graduated from this university in June of 1967 with an A.B. in history.

From 1967 through 1971, Mr. Semyck taught mathematics and science and, later, social studies at River Trails Junior High School in Mount Prospect, Illinois. Mr. Semyck entered the program in clinical psychology at Loyola University in 1971. He completed his practicum and internship training at Hines V. A. Hospital and Loyola University Guidance Center.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
VITA	iii
LIST OF TABLES	vi
CONTENTS OF APPENDICES	vii
INTRODUCTION	1
Diagnostic Utility	2
Factor Analysis	4
Artistic Ability	6
REVIEW OF RELATED LITERATURE	9
Hiler-Nesvig Formula	9
Koppitz Emotional Indicators	12
Koppitz Expected and Exceptional Items	23
Hypotheses	25
METHOD	37
Subjects	37
Procedure	37
Scoring	40
RESULTS	43
Reliability	43

TABLE OF CONTENTS (continued)

	Page
Convergent Validation	45
Discriminant Validation	46
Comparison of the EI Scale and the MHN Formula	49
Item Overlap	49
Sex Differences	50
Internalizer/Externalizer Hypotheses	50
DISCUSSION	52
Koppitz EI Scale	52
Koppitz EE Scale	54
The MHN Formula	57
Sex Differences	60
Internalizer/Externalizer Hypotheses	61
Conclusion	63
SUMMARY	66
REFERENCES	68
APPENDIX A	77
APPENDIX B	82
APPENDIX C	86

LIST OF TABLES

Table	Page
1. Illustration of modified multitrait-multimethod matrix	32
2. Modified multitrait multimethod matrix	44

CONTENTS OF APPENDICES

	Page
Appendix A Scoring Chart for Expected and Exceptional Items	78
Appendix B Description of Emotional Indicators	83
Appendix C Modified Hiler-Nesvig Scoring System	87

INTRODUCTION

Several years ago Sundberg (1961) revealed that drawings of human figures were the second most frequently employed psychological technique for personality assessment in this country. Psychologists have utilized the drawings of human figures to assess both general intellectual development (e.g., Buck, 1948; Goodenough, 1926; Harris, 1963) and personality adjustment (e.g., Buck, 1948; Hammer, 1958; Mackover, 1949).

Reviews of the research on the use of drawings in evaluating intellectual maturity (e.g., Anastasi, 1972; Dunn, 1972) have considered the Goodenough-Harris Drawing Test (Harris, 1963) a fairly reliable and valid measure of the mental maturity of children. Zimmerman and Woo-Sam (1972) stated that correlations of the scores from drawing tests and the WISC ranged from .43 to .81 with a variety of samples including retarded, normal, and bright groups. It can be noted that these correlations are comparable to those obtained between the WISC and the Peabody Picture Vocabulary Test or various group tests of intelligence. Overall, it appears that human figure drawings can provide a quick and fairly valid measure of children's mental ability.

However, the employment of human figure drawings in personality appraisal has often been questioned in the literature. For instance, Swensen (1957) surveyed eight years of research on Machover's (1949) Draw-A-Person (DAP) test and concluded that ". . . the DAP is of

doubtful value in clinical work" (p. 461). Yet he noted the increased value of a series of drawings over just one or two drawings, the possible accuracy of Machover's interpretations in an individual case, the value of the DAP as part of a test battery, and the successful utilization of the test as an indicator of general "level of adjustment." At a later time, Roback (1968) attempted to evaluate eighteen years of research on Machover's hypotheses and concluded that they were largely unsupported by the research. He stated pessimistically that the ultimate fate of the DAP may be of "a rough screening device." But Swensen (1968) reviewed the same literature and drew a more optimistic conclusion. He stated: ". . . there has been substantial increase in empirical justification for the use of the DAP as a clinical tool" (p. 40). His opinion rested on recent studies of test reliability, extended drawing techniques, serial production of drawings by one subject over a period of time, patterns of signs of psychopathology, specific hypotheses, the process used in clinical judgment, and drawing results due to manipulations of the emotional state of the subjects.

Diagnostic Utility. More recent research on the employment of human figure drawings in assessing personality adjustment has included a consideration of the diagnostic utility of the test, factor analysis, artistic ability, and scaling devices based on specific signs of pathology. The diagnostic value of the DAP was tested in one especially noteworthy study by Wanderer (1969). Having consulted with Machover in planning the study, Wanderer carefully matched samples of five groups of adult subjects: mental defectives (only group unmatched for intelligence and education), schizophrenics, neurotics, homosexuals,

and normals. The obtained drawings were judged by a pool of the 20 highest ranked and cooperating experts with the DAP. Wanderer found that with all five of the subject groups the experts did significantly better than theoretical chance. However, if the mental defectives, who were correctly labeled by 95% of the judges, were removed from the list, then the experts did not do significantly better than chance in classifying the drawings. In considering the results, the researcher thought that the DAP may be popular merely because it occasionally reinforces the clinician in his use of the test or that the clinician may attribute knowledge from an interview and other extra-test conditions to the DAP.

Another possible interpretation of Wanderer's results is that the judges were influenced by an "illusory correlation." Chapman and Chapman (1967) postulated that entirely naive observers who view psychodiagnostic materials would report the same but erroneous correlates of patients' symptoms due to variables inherent in the stimuli observed. In a series of experiments, these investigators discovered that naive undergraduates, who viewed DAP drawings randomly paired with contrived symptoms statements about the patients who drew them, "rediscovered the same relationships between drawing features and symptoms as employed by practicing clinicians despite the fact that these relationships did not exist in the task materials.

Yet Hammer (1969) suggested other interpretations of Wanderer's study. He thought the results may be due to the narrowness of the range of judges, the fact that the method used by Wanderer compelled a judge to make a second mistake if he made one, and the artificial collapsing of the number of correct judgments into three categories:

0, 1, or 2 and more. Furthermore, Hammer noted that Wanderer's groupings of subjects may well have overlapped. For example, all outpatients at a psychiatric clinic were considered "neurotics" and may well have included other kinds of patients. In addition, Hammer felt that human figure drawings are best viewed as part of a test battery and thought it was unreasonable to expect one five-minute test to yield a diagnosis by itself. Moreover, the author stressed the importance of extending the number of drawings obtained and including the verbal comments made by the subject in any evaluation.

Various other studies have investigated the relationship between certain drawing characteristics and specific diagnoses or traits (e.g., Carlson, Quinlan, Tucker, & Harrow, 1973; Cauthen, Sandman, Kilpatrick, & Deabler, 1969; Gravitz, 1969; Johnson, 1971; Lapkin, Hillaby, & Silverman, 1968; Reznikoff & Dies, 1969; Wilkinson & Schnadt, 1968). These investigations will not be discussed herein. Most of them appear to present some useful data, yet they leave many questions unresolved and have various methodological limitations.

Factor Analysis. Perhaps of greater importance than studies of specific diagnoses or traits is the factor-analytic approach in evaluating human figure drawings. In an early study, Nichols and Strumpfer (1962) had included five global scales, height measures, and fourteen specific details in evaluating drawings made by male college students and VA patients. Their orthogonal simple structure contained four main factors: a broad factor which may be interpreted as reflecting psychological adjustment, or drawing ability, or both; an age factor; a size factor; and an aggression factor. The authors

proceeded to select certain scores from the VA sample and obtained oblique factor loadings; a quality-of-drawing factor, a "big-bosomy figure" factor, a defensiveness factor, and a "gross-behavioral-adjustment" factor. Overall, the researchers interpreted their one major factor as reflecting quality of drawing. This view was adhered to on the basis of the gross behavioral adjustment scale employed in the study. Yet, their reasoning does not seem compelling in that the "normals" of the VA sample were hospitalized patients, perhaps suffering from psychologically related psychosomatic problems or trauma situations. Consequently, their behavioral adjustment scale based on these "normals" may well have allowed an overlapping of "disturbed" groups, and therefore does not appear to be a sound basis upon which to interpret the general factor as being "drawing ability" rather than "psychological adjustment." It is also noteworthy that these authors indicated that the "drawing ability" did not reflect "artistic ability" because ratings by an artist failed to correlate with this factor, which had, however, been termed by other psychologists as "artistic quality."

In a more recent factor-analytic study, Adler (1970) attempted to extend the research of Nichols and Strumpfer. He employed 32 scoring categories which had been associated with pathology in the literature. He chose a four-factor solution following Varimax rotation as being the most compelling. The factors were labeled as follows: 1) formal accuracy of the drawn figure; 2) size and placement; 3) bizarreness or internal inconsistency; and 4) failure of behavioral control or lack of concern. Adler interpreted his results as indicating that the major

valid use of figure drawings is in evaluating cognitive maturity and hence felt that this must be controlled in any clinical assessment. But Adler seems to have contradicted himself in that he pointed out that the three other factors are relatively independent of each other and of the first factor and suggested therefore that they may bear a significant relationship to personality variables and diagnostic categories. It would appear that if certain drawing characteristics are related primarily to personality variables then the use of these features may be developed into a valid means of personality assessment.

Adler has authoritatively claimed that figure drawings are "essentially a one-factor test." One wonders if he is aware of the numerous subjective judgments that he had made in deriving his results and conclusions. Biases are present in his choice of scoring categories, method of rotation, choice of factor structure, and labeling of factors. One is especially curious as to why Adler did not obtain an oblique factor structure as Nichols and Strumpfer did. Mental maturity and personality adjustment may be inter-correlated variables yet independently measurable and modifiable, just as human height and weight are. Consequently, an oblique factor analysis would seem more appropriate for a consideration of such variables.

Artistic Ability. Besides studies involving the diagnostic utility or the factor analysis of human figure drawings, another research topic has been the influence of artistic ability upon clinical evaluation of drawings (Roback, 1968; Swensen, 1969). Ever since Whitmyre (1953) found a significant correlation between the ratings of personality adjustment by clinicians and of artistic ability by

artists, some critics have suggested that person drawings reflect "nothing but" artistic skill. Swensen (1968) mentioned several studies which indicated that artistic ability was in fact a contributing factor in the clinical evaluation of drawings. More recently, Solar, Bruehl, and Kovacs (1970) obtained correlations of .74 between Witkin's (1954) short-form scale (based on Machover's ideas) and artists' ratings and .76 between a global rating of sophistication of body concept and artists' ratings. In another study, Young (1970) derived results which he interpreted as indicating that art quality is the major factor influencing clinicians' adjustment ratings of DAP tests. However, he also found that patient-nonpatient status of the subject is a significant factor in the evaluation of adjustment by clinicians.

Thus, one may conclude that art quality appears to be a contributing factor to a clinician's global judgment of drawings. Consequently, the artistic skill involved in a drawing should be taken into account in globally evaluating a drawing for personality adjustment. However, the degree of the relationship between ratings of artistic quality and personality adjustment may have been exaggerated by the artificial nature of the research studies in that they required the clinicians to rate adjustment for all drawings whereas in practice they may obtain useful information from drawings only in some cases. Just as a medical doctor often finds an X-ray "unremarkable," so too a clinician may find that a single drawing does not contribute much to his understanding of the patient. In those cases in which the drawings may have been "unremarkable," the clinician, forced to make a judgment, may have relied upon artistic quality or cognitive accuracy

in evaluating the protocols. But in actual practice the clinician would have relied on other information. Furthermore, another possible view is that the clinician in practice adjusts his interpretations of the drawings depending upon the patient's educational background, intellectual ability, and artistic interest and skill of which he learns in interviewing and testing the patient.

REVIEW OF RELATED LITERATURE

Such confounding factors as artistic skill, mental maturity, and illusory observation could be minimized in the assessment of person drawings if the clinician were to rely only upon a cluster of drawing signs or features that could be reliably scored and validly related to psychopathology. Such an orientation would avoid the global assessment of drawings which allows the clinician to depend upon his so-called intuition and favorite personality theory and to be influenced by the artistic quality and the cognitive accuracy found in a drawing.

Hiler-Nesvig Formula. An important study along this line of reasoning was made by Hiler and Nesvig (1965). These authors uncovered the criteria which were successfully employed by clinicians in judging the drawings of adolescent patients and nonpatients. The criteria that discriminated beyond the 1% level of significance were used in forming a prediction formula: "definitely bizarre" and "major part omitted" were scored "-1" while "nothing pathological" and "happy or pleasant facial expression" were rated "+1." Those subjects receiving minus scores were predicted to be patients.

With a cross-validation sample of similar adolescents, Hiler and Nesvig found that, whereas psychologists and non-psychologists working without the formula were 64% and 65% accurate respectively, three graduate students utilizing the formula were on the average 79% accurate in judging patient or nonpatient status. Furthermore, the

mean biserial correlation coefficient between formula scores and the patient-normal dichotomy was .72, while the mean inter-judge reliability coefficient was .71. Thus, with a fair level of reliability, the specific-sign prediction formula developed by these researchers appeared to improve the accuracy of assessment with human figure drawings.

Stricker (1967) argued that Hiler and Nesvig had pitted actuarial prediction against naive clinical prediction. Stricker attempted to compare actuarial, naive clinical, and sophisticated clinical assessment, distinctions made by Holt (1958). Stricker formed three groups of "sophisticated" evaluators by providing them with the results of Hiler and Nesvig's research before evaluating the drawings. The groups were composed of six experienced clinicians, ten third-year and twelve first-year clinical graduate students. Some of the information provided to the judges included the criteria of the Hiler-and-Nesvig prediction formula: patients often revealed bizarreness and omissions of major parts of their drawings, while normals tended to sketch figures with a happy, pleasant facial expression and had nothing pathological. But, in addition, Stricker included two other indicators of pathology which were found by Hiler and Nesvig to be significant at about a 5% level: distortions (especially of head or arms) and transparencies. Furthermore, Stricker told the judges that some signs were not of value in making their judgments: certain conflict and anxiety indicators, size and line pressure, absence of clothing, proportion between body parts, and motion and posture of figure.

The three groups of evaluators all viewed the same drawing protocols which were used in the Hiler and Nesvig study. Stricker

learned that, while the "sophisticated" clinicians were 66% correct, the first- and third-year students were 72% and 73% accurate. The combined student group was statistically superior to the clinicians. Thus, the clinicians appeared to discount the provided information. Also, Stricker's students' 73% level of accuracy appeared poorer than the 78% modal level of the so-called actuarial judges in Hiler and Nesvig's study. However, Stricker argued that, since 23% of his sophisticated evaluators were better than the 78% modal level obtained with the formula, some sophisticated judges could do better than the actuarial judges. Nevertheless, Stricker's reasoning is not very cogent. First, using a modal level of accuracy for comparison of groups is inappropriate. A statistical test is needed. It can be noted that while 23% of Stricker's sophisticated students were superior to the 78% level, in fact 33% (one in three) of Hiler and Nesvig's raters did better than the 78% modal level. Secondly, Stricker gave more information to his judges and so the two groups are not directly comparable in terms of "actuarial" and "sophisticated" prediction. One wonders if a prediction formula which included "distortions" and "transparencies" as "-1" scores would have improved upon the 78% level of accuracy found by Hiler and Nesvig.

In a related study, Young (1970) provided clinicians with the research information which Stricker gave his judges. Young also included two more statements: first, that clinicians often rated drawings only on the basis of art quality; and secondly, that a global analysis was more reliable and valid than an atomistic approach. The author predicted that informed clinicians would do better than an

uninformed group of clinicians. However, contrary to his hypothesis, Young found that the informed clinicians did no better than the uninformed ones on the protocols from college students and VA patients. Apparently then, the clinicians in this study, like those in Stricker's experiment, discounted or ignored the useful information and relied upon their own methods of evaluating human figures.

Koppitz Emotional Indicators. It can be noted that all of the foregoing recent studies concerning the usefulness of human figure drawings in personality assessment obtained test protocols from adolescents or adults. The present writer believes that these studies have overlooked the subjects for whom the utility of person drawings is greatest. It is thought by this investigator that figure drawings are an especially good test of the personal characteristics of children. First, the test is relatively quick and thus suited to the short attention span of young children. So too, most youngsters enjoy drawing so the technique fosters rapport between child and examiner, which is an important factor often overlooked in assessment. Furthermore, the nonverbal nature of the test allows for assessment of taciturn and very shy children. Moreover, since children's reading skills are limited, their ability to use other forms of standard tests, such as questionnaires, is restricted. Also to be considered in this era of reduced financial budgets is the fact that drawings are a very inexpensive method of assessment.

As Koppitz (1968) indicated, the foremost proponents of the projective approach to figure drawings have been Machover (1949, 1953, 1960), Levy (1958), Hammer (1958, 1960) and Jolles (1952), all of whom

have worked mainly with adolescents and adults and only to a limited extent with children of the elementary school age. Machover (1953, 1960) attempted to offer hypotheses concerning the drawings of children, but offered neither a scoring system nor controlled research data. What is more important, tests of Machover's hypotheses have tended to be inconclusive (Koppitz, 1968; Roback, 1968; Swensen, 1968). But this may be due to the Freudian orientation of her interpretations, rather than to the lack of relationships between certain drawing features and various criteria of pathology.

For children's human figure drawings (HFDs), Koppitz (1968) has developed lists of signs to evaluate not only personality adjustment but also mental maturity. Initially, Koppitz, Sullivan, Blythe, and Shelton (1959) designed a tentative scoring system to be used along with the Bender-Gestalt in screening school beginners. Twelve drawing characteristics were thought to indicate emotional upset and/or lack of mental ability. Six other items were believed to reflect need for achievement and/or aggressive striving, while three others were understood as indicating above-average intelligence. All of these characteristics were combined into one scale. The researchers found that the drawing scores and the Bender-Gestalt scores measured primarily different factors and supplemented each other, in accurately predicting school achievement.

In a further study, Koppitz (1965) tried to compare drawings made by crayon with others made by pencil. The drawings were scored for the presence or the absence of twenty-two "developmental items" and eighteen "emotional indicators." On the developmental items, the

results were generally equivalent with pencil and crayon methods; however, girls tended to do better than boys. As for the emotional indicators, differences were obtained in comparisons of pencil and crayon methods and of boys and girls. However, clear conclusions are difficult to draw from this study because drawing methods were confounded with task instructions and group and individual administrations.

From these initial investigations, and her clinical experience, Koppitz (1966b) made a list of thirty emotional indicators (EIs) and tested the scale in distinguishing between a group of children from a guidance center and another group of public school children, matched for age and sex. The students were asked to "draw a whole person" on a blank sheet of paper with a No. 2 pencil. Koppitz and another psychologist independently scored the drawings and obtained a 95% level of agreement. The author discovered that four items (poor integration, shading of body and/or limbs, slanting figure, and tiny figure) were significant at the .01 level and that four other characteristics (big figure, short arms, cut-off hands, and omission of neck) were significant at the .05 level. Koppitz also thought that four scale features (shading of hands and/or neck, asymmetry of limbs, transparencies, and big hands) significant at the .10 level were noteworthy. In addition, while two of the thirty items did not occur in the protocols used in the study, all of the other scale characteristics tended to be in the predicted direction. Furthermore, Koppitz pointed out that while only 5% of the well adjusted group had two or more EIs in their drawings, about 74% of the clinical group had two or more EIs. Consequently, Koppitz thought that two or more EIs in an HFD of a child

between the ages of 5 and 12 suggest that the child is maladjusted.

In evaluating this study, two difficulties are quite apparent. First, the children from the two groups were not tested in the same location. It is possible that children tested at the clinic were more anxious and produced more EIs than children tested at school simply because of the difference in testing situations (e.g., Handler & Reyher, 1964). Also, although figure drawings are used to assess intelligence (Harris, 1963), Koppitz did not match the groups on intelligence. It may be that, if the groups were to be matched for intelligence, the disturbed group may actually be more "potentially intelligent" since emotional maladjustment would be expected to lower intellectual performance. Hiler and Nesvig (1965) followed this line of reasoning. Yet intelligence or mental maturity does stand in the present study as a possible confounding factor.

Another investigation using a psychological referral as the criterion of pathology was performed by Fuller, Preuss, and Hawkins (1970) to cross-validate the utility of the 30 EIs. These authors picked 80 normal public school children (five boys and five girls at each age from 5 to 12) and compared their HFDs to those of emotionally disturbed children (of similar ages) referred to either a guidance clinic or a school psychologist. Three judges scored the protocols and obtained inter-rater reliabilities of .84 for the normal group and .71 for the disturbed group. Fuller et al. found that nine EIs appeared more frequently among the disturbed group: poor integration, gross asymmetry of limbs, hands cut off, long arms, tiny head, three figures, no neck, no nose, and no feet. Thus, four of the items were

significantly found by both Koppitz and Fuller et al. in the HFDs of disturbed children: poor integration, gross asymmetry of limbs, hands cut off, and no neck. Furthermore, Fuller et al. stated that if Koppitz's method of predicting normal adjustment were used in their study, 58% of the disturbed children would have been incorrectly diagnosed as normal while 82% of the normal group would have been properly classified. The researchers thought that three or more EIs would better predict maladjustment.

In this study, as in Koppitz's (1966b) research, the possible confounds of intellectual maturity and of situational administration effects are present. Furthermore, it can be noted that both validity studies employed the criterion of referral to a clinic or school psychologist as an index of maladjustment. Also, both Koppitz and Fuller et al. have suggested that a certain number of EIs may be understood as indicating maladjustment. However, there was some difficulty in establishing the specific number of EIs which could be understood as indicating pathology. Part of the difficulty appears to be related to the dichotomous nature of their criterion of maladjustment. Perhaps adjustment or lack of it may better be conceived as a continuous variable. Consequently, the establishment of a definite number of EIs to be used as a "cut-off point" for maladjustment may be an artificial task. Additional research may profitably explore the relationship of the EI scale to other criteria of pathology such as objective psychological inventories which provide one score of "general level of adjustment" or one general score along with subscores indicating patterns of experienced symptoms.

In two studies Koppitz attempted to discover whether various EIs are differentially related to certain traits of children with problems. In one of these investigations, Koppitz (1966c) compared the HFD protocols of shy youngsters with those of aggressive children. Thirty-one pairs of children who were patients at a child guidance clinic were matched for age, sex, and WISC IQ score. In her results Koppitz believed that she demonstrated that "tiny figures" and the "omission of nose, mouth, or hands" were associated with shy children while "gross asymmetry of limbs," "teeth," "long arms," "big hands," and "genitals" were produced more often by aggressive youngsters. But the author's findings seem inconclusive because she used an inappropriate comparison group for each of the two groups. Instead of comparing each group with the other as Koppitz did, it appears that each group should have been compared to a normal control group. Thus, the characteristics noted by Koppitz are merely relative to the other group utilized, rather than to a normal population.

In similarly faulted research, Koppitz (1968) compared the HFDs of children with psychosomatic complaints and of those in trouble for stealing. Koppitz thought that an equally high number of the following features appeared in the drawings of both groups: shading of body and limbs, poor integration, hands cut off from arms, tiny figure, slanting figure, and omission of feet. Moreover, whereas the children with psychosomatic complaints revealed more "short arms," "clouds," and "no nose," the youngsters who stole produced more "big hands" and "no neck." But again, to infer that these signs are associated with the particular group, a normal control

group should have been utilized. Koppitz's results may be viewed as providing suggestions for further research.

In the last two studies by Koppitz (1966c, 1968), children who internalize their conflicts (shy and psychosomatic) have been compared with children who act out in response to their conflicts (aggressive and delinquent). Other research has compared the drawings of children with similar traits. McHugh (1966) compared children with neurotic traits with a group characterized by conduct disturbance. The groups were matched for age, sex, and mental ability. None of the characteristics of the Koppitz EI list discriminated between the two groups. Unfortunately, McHugh compared the two groups together rather than with an appropriate control group of normals.

With fourth- through sixth-grade students, Starkey (1970) obtained data supporting the convergent, but not discriminant validity of a list of EIs thought to reflect aggressive tendencies in children. With the exception of "no neck," all of the Koppitz (1966c) "aggressive" items were included on a list as well as "big figure," "general transparencies," and "omission of arms." Starkey found support for this list of EIs against two criteria: an "aggressive" factor on the Children's Personality Questionnaire and a checklist of aggressive responses from the Behavior Problem Checklist. A similar list of EIs purported to reflect anxiety were not validated. However, a list of items termed "emotional instability" received convergent validity, but not discriminant validity, with one criterion, the checklist. His "emotional instability" items were the following: poor integration of parts, shading of entire face, shading of neck,

slanting figure, tiny head, omission of body, clouds, and omission of neck.

Handler and McIntosh (1971) utilized drawing items emphasized by Koppitz (1966c) and McHugh (1966) in evaluating HFDs of aggressive, withdrawn, and normal 8- to 10-year olds. Categorization of subjects was made on the basis of teacher and peer judgments. The authors failed to find the aggression or the withdrawal items helpful in discriminating significantly between groups. However, they noted that the drawing items allowed a higher rate of correct classification than self-classification or a brief behavioral observation.

Another study attempted to replicate Koppitz's (1966c) findings for shy and aggressive children. Lingren (1971) matched pairs of 5- to 12-year-old children, considered to be either aggressive or shy. Contrary to Koppitz's results, she failed to find any significant differences on the Koppitz EIs between the two groups.

Thus in several studies there appears to be inconsistent support for the usefulness of drawing items in discriminating between children who internalize their conflicts and those who act out or externalize them. Some of the inconsistency may relate to differences in ages and backgrounds of subjects or the criteria used to evaluate the traits under consideration.

In other related research, Koppitz (1966a) studied the relationship of the 30 EIs to school achievement in the first two school grades. Prior to this work, Vane and Eisen (1962) tried to validate the use of drawings by kindergarten children in assessing adjustment. Using a list of 11 characteristics of pathology, these

authors found that four items were related to teachers' ratings of adjustment at the .01 level of significance: "grotesque," "no body," "no mouth," and "no arms." Furthermore, to assess whether the relationship between drawing characteristics and rated adjustment, the researchers matched two pairs of groups of children for adjustment and IQ (one pair with a vocabulary IQ and the other with the Goodenough IQ). In both matched groups, none of the children in the good-adjustment groups had any of the four signs of maladjustment, while those in the poor-adjustment groups had a significant number. In addition, the four characteristics were helpful in predicting later adjustment ratings in the first grade.

In her study with primary-school students, Koppitz (1966a) asked children to draw a whole person at the beginning of the school year; later at the end of the year, an achievement test was administered. On the basis of the achievement test results some students were classified as good or poor pupils, and the drawings were evaluated for the 30 EIs. Koppitz discovered that five EIs significantly distinguished the groups: poor integration of parts, slanting figure (15° or more), omissions of body and/or arms, and three or more figures spontaneously drawn. In interpreting her results, Koppitz thought that these items can be used as indicators of special learning problems among primary grade children.

It is noteworthy that both Vane and Eisen (1962) and Koppitz (1966a) found omissions of body and/or arms to be associated with children with troubles. It can also be pointed out that the differences between the results of the two investigations may have been due

to the differences in the ages of the subjects, in the instructions given, and/or in the kind of criteria employed. Furthermore, it is unfortunate that Koppitz did not check on the influence of mental maturity upon the value of the obtained items, as Vane and Eisen had done. In any case, some HFD items do seem associated with difficulty in learning primary school work.

Two studies on EIs have considered cultural and socioeconomic variables. Koppitz and DeMoreau (1968) matched two groups of school children age 5 to 11 years for age, sex, and mental maturity. One group was comprised of lower-class Mexican children from Guadalajara, while the other was made up of lower-class children from a small town in New York State. Both of the groups were divided into younger (ages 5 to 7) and older (ages 8 to 11) groups.

The authors found that six EIs differentiated the two subgroups of younger children while 11 EIs discriminated between the subgroups of older subjects. Whereas young Mexican children made drawings with more of two signs (tiny figures and slanting figures), young subjects from the United States drew more of four features (shading of hands, short arms, teeth, and clouds). Also while older Mexican children produced three of the characteristics (tiny figures, slanting figures, and transparencies) more frequently, the older United States subjects revealed more of eight items (shading of body and/or limbs, shading of hands and/or neck, short arms, teeth, clouds, omission of feet, big figures, and grotesque figures). The researchers thought that such signs suggested that Mexican children were more immature, insecure, timid, and concrete in their thinking than United

States children who experienced more anxiety, aggression, resentment, and inadequacy and had poorer self-concepts than their counterparts.

In another investigation, the occurrence of EIs in the HFDs of boys and girls from lower- and middle-class backgrounds in the United States were compared. In her sample, Koppitz (1969) discovered that three EIs appeared significantly more often in the drawings of lower-class students than in those of middle-class pupils: shading of hands and/or neck, legs pressed together, and omission of feet. Also middle-class children had more teeth and big figures in their drawings. Koppitz also drew a sample of 79 students from each group matched on age, sex, and mental maturity. With this sample she obtained no significant differences between the lower-class and middle-class groups. Koppitz interpreted this finding as indicating that many less advantaged youngsters are often overlooked as being adequate and capable children.

Regardless of the subjects' socio-economic background, Koppitz found that nine EIs appeared more frequently in the drawings of boys than in those of girls: poor integration, shading of face, shading of body/limbs, shading of hands/neck, transparencies, tiny figures, teeth, arms clinging, and grotesque figure or monster. Moreover, whereas only one-sixth of the girls showed two or more EIs, more than one third of the boys produced that number. Koppitz viewed the signs associated with males as pointing to the boys' impulsive, aggressive, anxious, and inadequate feelings and poor self-concept.

Judging from the outcome of the above studies, it is well to note that in any set of drawings such factors as age, sex, intelligence,

socio-economic level and cultural background may contribute to variation in the presence of EIs. Many of the inconsistencies in the results of research may be related to such factors. Of course, variation in the criteria of emotional adjustment or personality trait is another source of variability in the data.

Koppitz Expected and Exceptional Items. In both of the last two studies described above (Koppitz, 1966a; Koppitz & DeMoreau, 1968), the scale used to match children for mental maturity was a list of HFD "expected" and "exceptional" (EE) items developed by Koppitz (1967) on a sample of 1856 public grade school children. The investigator utilized various drawing characteristics that were related to mental development in children. Items which appeared on 86% or more of all HFDs of children at each age were considered expected items, while those which were present on less than 16% of the HFDs were termed exceptional items for that age. Koppitz devised a scoring system by giving each protocol an initial score of "5" and then scoring the omission of each expected item "-1" and the presence of each exceptional item "+1." The total number of EE items for any one age never exceeds 17; and, consequently, the scoring of an HFD for the EEs is much faster than the scoring of the Draw-A-Man (DAM) Test by Goodenough's 51-item system or Harris's 73-characteristic scale.

On a sample of 347 boys, aged 6 to 12, Koppitz obtained product-moment correlations between the HFD EE scores and either WISC or Stanford-Binet IQ scores. Correlations were significant at the .005 level for all ages and ranged from .45 to .80. Koppitz noted that these correlations were comparable to those obtained

between the Goodenough DAM test and other IQ test scores. Also, it can be pointed out that Koppitz provided a "level of mental maturity" interpretation to her scores rather than an IQ or mental-age scale.

The EE mental maturity scale was further evaluated on a group of 335 Mexican school children by DeMoreau and Koppitz (1968). Goodenough DAM scores were correlated with Koppitz EE scores. Correlations ranged from .64 to .77. Therefore, DeMoreau and Koppitz thought that the EEs provide a quick index of mental maturity that can be used with the children of Mexican culture, and perhaps other cultural groups as well.

Additional support for the use of the EE scoring system was derived by Snyder and Gaston (1970) in investigating the figure drawings of first-grade children. These authors found that with this age group essentially the same drawing characteristics could be termed "expected" and "exceptional" as defined by Koppitz. This result was obtained in spite of differences in instructions, methods of administration, and relationship to the examiner in the research of Koppitz and of Snyder and Gaston. However, unlike Koppitz, Snyder and Gaston caution against subjective interpretation of drawing signs because many characteristics appear as frequently as 30 to 60% of the time and therefore lack discriminative value. But this manner of reasoning may have limitations. For example, if a "sign of anxiety" appears among 30% of a sample of subjects it may be that 30% of the sample includes those with relatively greater anxiety, and consequently that sign, especially considered along with others, may be of value in assessing anxiety.

In developing her EI and EE scales, Koppitz hoped to discriminate between emotional adjustment and mental maturity. Yet a study by Hall and Ladriere (1970) raises a question as to the conceptual difference of the EI and the EE scales. These authors scored HFD protocols of emotionally disturbed, perceptually handicapped, and control groups. The groups were matched for age and WISC or SB IQ. The authors learned that both the 30-item EI scale and EE scoring system significantly discriminated between not only the emotionally disturbed group and control group, but also the perceptually handicapped group and the control group. Furthermore, neither scale discriminated between the perceptually handicapped and the emotionally disturbed groups. Consequently, the question whether the EI scale indicates anything different than what is reflected in the EE scale must be considered. However, if one considers means and standard deviations, it can be noted that the EI scale appeared to have a greater tendency toward discriminating between the perceptually handicapped and the emotionally disturbed group than the EE scale did.

Hypotheses. The purpose of the present study is to increase the amount of information available concerning the validity of several scoring systems for children's HFDs. One set of hypotheses considers the validity of labeling the scales as indices of mental maturity and emotional adjustment. The distinction between the Koppitz measure of mental maturity, the EEs, and her measure of emotional adjustment, the EIs, is questionable in light of the findings of Hall and Ladriere (1970). In addition, the Hiler-Nesvig formula was developed on the protocols of adolescents; and, the question whether the

formula indicates emotional adjustment among children may be raised.

In the present study, three scores, each purportedly related to either mental maturity or emotional adjustment, are obtained from children's HFDs. Two of the scores are derived from the scoring systems of Koppitz. The Koppitz (1967) EE scale is used to indicate mental maturity, and the Koppitz (1966b) 30-item EI list serves as a measure of emotional adjustment. The relevant parts of the Koppitz EE system and the EI list are presented in Appendices A and B, respectively. In addition, a Modified Hiler-Nesvig (MHN) prediction formula, derived from the studies of Hiler and Nesvig (1965) and of Stricker (1967), acts as an additional measure of children's emotional adjustment. The MHN scale scores each drawing "5" initially. A score of "+1" is added for two items: happy, pleasant facial expression and nothing pathological. So too, a score of "-1" is recorded for each of the following four items: bizarreness, omission of major parts of the body (head, body, arms, legs, hands, feet, eyes, nose, mouth, and hair), distortion of head or arms, and transparencies of the body, arms, or legs through the clothing. Thus MHN scores range from 1 to 7. (A more detailed description of the items is provided in Appendix C.)

As cross-validating criteria of mental maturity and emotional adjustment, so-called "objective" psychological tests are employed. This type of criterion, the objective test, is different than the criteria employed by Koppitz (1966b, 1966c, 1968), by Fuller, Preuss, and Hawkins (1970) and by Hiler and Nesvig (1965). In the present

research, two group-administered tests are utilized. The Otis-Lennon Mental Ability Test (Otis & Lennon, 1967) acts as a standardized test of mental maturity. The Otis-Lennon Mental Ability Test (OLMAT) yields one score reflecting "general intellectual ability." This test has been favorably reviewed (e.g., Milholland, 1972) and supported by a broad range of research findings (Otis & Lennon, 1969) on the reliability and validity of the test.

As a cross-validating standard measure of emotional adjustment, the Total Adjustment score of the California Test of Personality (Thorpe, Clark, & Tiegs, 1953) is utilized. The Total Adjustment score is composed of the fairly reliable Personal Adjustment and Social Adjustment scores, each of which is made up of six subscales. The California Test of Personality (CTP) has been generally accepted by reviewers (e.g., Sims, 1959) and supported by research results (e.g., Jackson, 1946; Semler, 1960; Smith, 1958; Thorpe, Clark, & Tiegs, 1953).

The use of a group intelligence test is a fairly accepted method of measuring the trait of mental ability, especially for research purposes. More discussion is perhaps needed concerning the present selection of a criterion for measuring emotional adjustment. As Fiske (1971) has stressed, in the area of personality information may be collected by various methods or modes. The CTP relies on the self-report method. One may question whether other methods might provide a preferable approach to evaluating emotional adjustment. Other methods in assessing emotional adjustment commonly include an evaluation of prior behavior (e.g., ratings by peers or associates)

and an observation of behavior (interview or individually administered projective test).

While all the research on the CTP cannot be reviewed herein, a few representative studies can be mentioned. Jackson (1946) compared the ratings of general adjustment obtained by the CTP with those from a standard clinical interview (including projective testing), from an observation by the experimenter, from teachers well acquainted with their students, and from parents. Jackson noted that, while ratings of general adjustment by teachers, the experimenter, or parents were most influenced by intelligence and school achievement, the ratings based on group paper-pencil testing or interview are less influenced by such factors. Jackson found that the correlation of evaluations by interview and ratings based on the CTP was .73. Jackson (1946) concluded that of the methods used the CTP was most effective. It can be noted, however, that Jackson's research was conducted on the 1939 version of the CTP. Yet this does not seem to hinder his conclusions for the 1953 version because several technical improvements were made in the 1953 edition.

Another noteworthy study on the CTP was conducted by Smith (1958). Groups of well adjusted, average adjusted, and very poorly adjusted boys were selected on the basis of a combined criteria of teacher nomination and peer evaluation. The accuracy of the groupings were in turn supported by four independent estimates of adjustment: referrals to school social workers, referrals to the Community Guidance Clinic, arrests and records with the Juvenile Police, and participation on school police patrols. Smith found that the differ-

ences between CTP group means were significant even when intelligence, reading achievement, and level of parental occupation were controlled.

In another study, Peak (1963) found that the CTP Total Adjustment (TA) score discriminated between a control group of "normal" ninth-grade males and a group of psychiatric patients. Peak also indicated that the Social Adjustment (SA) score differentiated between the control group and a group of delinquents incarcerated in a state industrial school. The test did not, however, distinguish between the control group and other groups of "leaders" and "problem students." These latter groups were selected by the principal, teachers, and a school psychologist. Yet recalling the study of Jackson (1946), it seems possible that the judgment of the principal and teachers were influenced by factors such as intelligence and school achievement. Thus, the groups of leaders and problem students may merely reflect bright students and slow learners.

One special issue for the self-report method is distortion or faking. Thorpe, Clark, and Tiegs (1953) indicated that as students mature and reach senior high school, the research evidence on test distortion becomes cloudy. Kimber (1947) attempted to evaluate the level of insight as to the "healthy" answers among college students. He found a significant difference between the tests of the students when instructed to answer the CTP as a well adjusted student might and when given standard instructions. However, the scores on the two testings correlated at .52 for men and .54 for women. Thus even with "fake good" instructions, the test measured much of the same quality as with normal instructions.

With younger subjects the test appears to be less susceptible to "faking good." King and Ross (1965) found no significant difference between scores of ninth-grade subjects when instructed to "fake happy" and when given by the usual set of instructions. This would seem to suggest that the students could not present themselves in a better light when they attempted to do so. A different interpretation was made by King and Ross, however, in that they thought that students usually "fake good" and thus could not do better than their usual attempt to do so. Their reasoning seems overly cynical of subjects' answers and their interpretation appears to enhance their own bias. They provide no support to their contention that subjects usually "fake good" independent of the CTP scores.

Overall then, the CTP seems to be as good a method for evaluating emotional adjustment for research purposes in a group setting as any other method. Test distortion is a possible factor yet available research has not conclusively shown that "faking" is a major factor. It must be admitted that, just as an individually administered full-scale intelligence test may be a better index of an individual's mental ability than a group test, so may be a combination of objective testing, interview, and case history provide a better assessment of an individual's emotional adjustment than one objective test. Such a combined assessment procedure is not practical for the present study.

An additional consideration is that in some regards the OLMAT and the CTP represent different methods. They differ in instructions ("mark the best answer" compared to "your answers will show what you

usually think, how you usually feel, or what you usually do about things") and format (five-item multiple choice vs. yes-or-no answers). However, in many ways the test may be considered similar. In both tests there are a limited number of possible answers to each test question, each test item has one answer scored as "correct" or "desirable," and one total score is derived from all the questions, reflecting either "ability" or "adjustment." Furthermore, just as one may consider the manner in which an individual solves personal and social problems as being learned, so too the solutions to "intelligence" problems can be thought of as learned through interaction and experience. Thus, one may interpret both tests as indicators of different kinds of adaptive learning (e.g., Wesman, 1968).

In the present study one set of hypotheses relates to the convergent and discriminant validity of the three HFD scoring systems. The hypotheses are presented in reference to a modified multitrait-multimethod matrix (Campbell & Fiske, 1959). The matrix has been modified by the inclusion of two monotrait-monomethod measures of HFD emotional adjustment, as evident in Table 1.

Insert Table 1 about here

In Table 1 two methods are employed: drawings and objective test. Three scores purportedly indicate the construct of emotional adjustment: Koppitz's 30 Emotional Indicators (EIs), the Modified Hiler-Nesvig (MHN) formula, and the California Test of Personality-Total Adjustment (CTPTA) score. Two scores purportedly indicate the

TABLE 1

Illustration of modified multitrait-multimethod matrix
providing cardinal numerals to represent the correlations of various
scale scores.

	<u>Drawings</u>			<u>Objective Test</u>	
	<u>EE</u>	<u>EI</u>	<u>MHN</u>	<u>OLMAT</u>	<u>CTPTA</u>
<u>EE</u>	1				
<u>EI</u>	2	3			
<u>MHN</u>	4	5	6		
<u>OLMAT</u>	7	8	9	13	
<u>CTPTA</u>	10	11	12	14	15

construct of mental maturity: the Koppitz Expected and Exceptional (EE) Items and the Otis-Lennon Mental Ability Test (OLMAT). The correlations to be derived from the various tests are represented by cardinal numerals in Table 1.

One group of hypotheses relates to convergent validity. Based on the discussion of Campbell and Fiske (1959), it was anticipated that the validity coefficients would be greater than zero. In this, it was predicted that 7, 11, and 12 would be each significantly greater than zero. It was expected that the common trait variance of mental maturity in 7 and of emotional adjustment in 11 and 12 would lead to such results.

Another group of matrix hypotheses related to discriminant validity. One aspect of discriminant validity is the expectation that the validity coefficients are greater than the heterotrait-heteromethod coefficients. Thus, it was hypothesized that 7, 11, and 12 would each be greater than 8, 9, and 10. This result was predicted on the basis of the common trait variance in the validity coefficients which is lacking in the heterotrait-heteromethod correlations. In addition, it was hypothesized that 5 would be greater than both 2 and 4. This result was anticipated on the basis of the common trait variance expected in the two HFD measures of emotional adjustment.

Another set of matrix predictions related to a second criterion of discriminant validity. It was predicted that the validity coefficients would exceed the heterotrait-monomethod coefficients because the trait variance was expected to exceed the method variance. It was hypothesized that 7 would be greater than 2 and 4



and that both 11 and 12 would be larger than 14. Furthermore, it was predicted that a similar pattern of trait interrelationships would be obtained in the heterotrait areas of both the monomethod and heteromethod groups. This last prediction is the third criterion for discriminant validity, as explicated by Campbell and Fiske (1959).

Two other matrix predictions were made. First, it was hypothesized that 11 would be greater than 12. In predicting this, it was contended that the Koppitz scale would serve as a better index of emotional adjustment than the MHN formula because the EI list includes several items thought to relate to maladjustment which are lacking in the MHN formula. For example, "teeth" and "big hands" have been associated with an aggressive adjustment not only by Koppitz (1966c), but also by others like Hammer (1960). These items are present in the EI list but lacking in the MHN formula. Secondly, it was hypothesized that 2, 4, and 5 would each be greater than 14. This was anticipated not only because of the instructional and format differences found in the two objective tests but also because of an item overlap in the EE, EI, and MHN scales. For example, omissions of arms, body, legs, eyes, nose and mouth are included in all three scoring systems; and most of the items of the MHN formula are a subset of those in the EI list.

Another set of hypotheses related to expected sex differences. Machover (1960) has described in general terms sex differences in the drawings of boys and girls. More specifically Koppitz (1969) obtained evidence to suggest sex differences in the presence of EIs in the protocols of boys and girls. Four hypotheses concerning sex differences were tested in the present study. First, it was predicted

that boys would produce more EIs than girls, as Koppitz (1969) had found. Secondly, to test the generality of the Koppitz findings, it was hypothesized that boys' protocols would make more of the following: poor integration, shading of hands/neck, transparencies, tiny figure, teeth, arms clinging to body, and grotesque figure or monster. It was also predicted that boys would score more poorly on the MHN prediction formula. However, no differences between boys and girls were expected in EE scores of boys and girls because the protocols are scored so as to equate any differences due to sex.

Besides the matrix hypotheses and the hypotheses concerning sex differences, another set of hypotheses attempted to test whether certain drawing signs related to children's tendencies to internalize or act out their conflicts. In her research, Koppitz (1966c; 1968) concluded that various drawing characteristics were indicative of shy, aggressive, psychosomatic, and delinquent (stealing) children. It is noteworthy that several of the signs of shy children overlap or tend to overlap with some of the drawing characteristics of psychosomatic children (Koppitz, 1968). A similar pattern seems evident for the drawings of aggressive youths and children who steal. The present researcher proposed the following interpretation: While both shy and psychosomatic children tend to internalize their conflicts, both aggressive children and those who steal tend to externalize their conflicts. The conceptual continuum relied upon in this reasoning is similar to that of "internalization" as formulated by Welsh (1952) in working with the MMPI.

If this thinking were accurate, a certain pattern might be

expected in the CTP scores of children. It was hypothesized that six signs of shy and psychosomatic children (tiny figures, short arms, hands cut off, clouds, no nose, and no mouth) would be negatively correlated to a greater extent with the CTP Personal Adjustment (PA) score than the Social Adjustment (SA) score. That is to say, it was expected that children showing these indicators would have a lower level of personal adjustment than social adjustment as measured by the CTP, because it is thought that they internalize their conflicts. In a similar manner of reasoning, it was hypothesized that the six characteristics in drawings associated with youth who steal or act aggressively (asymmetry, teeth, long arms, big hands, genitals, and no neck) would be negatively correlated to a greater extent with the SA score than the PA score. It can be noted that the relationships between the above-mentioned EIs and the PA or the SA scores were to be ascertained by means of point-biserial correlation coefficients.

METHOD

Subjects. The children who provided the main body of data for the study came from six fourth-grade classrooms in the Chicago area. All subjects were either 9 or 10 years of age, except for one who was 11 years old. A pair of classrooms was utilized from three different school systems so as to broaden the range of family backgrounds included in the study. Judging from parental occupations, the socio-economic level of the children's families ranged from upper working class to upper middle class. Nearly all the subjects were white. Four of the classrooms came from public schools, while two were from a Catholic parochial school. While 140 students took the HFD test, only 136 children were present for all three tests. Of the 136 students, 64 were males and 72 were females. All children had obtained parental permission to participate in the project. Only one parent had objected to his child participating in the project.

Procedure. Three tests were administered to the children. Each class was presented the three tests in one of the six possible orders so as to reduce any effects due to a particular order. The assignment of a test order to each class was done at random. In addition, the children were tested with only one test per day so that students were not overloaded on any one day. Also, each class took the tests on three consecutive school days. This was done to limit individuals' changes over time. Furthermore, the tests were administered at three different school times (two morning, one after-

noon) so as to reduce any interaction between a particular test and a certain time of day. In this, each test was administered twice at each of the three school times. Also, testing did not interfere with the accustomed recess and lunch periods of the classes. Children's desks were spaced apart so as to discourage any copying or interaction during testing.

The investigator acted as test administrator for all classes. The normal classroom teacher was present to assist in the administration and to provide reassurance to the children in the presence of the examiner who was a stranger to the children. The researcher introduced himself and explained his objectives as follows:

My name is Mr. Semyck. Over the next few days, I'm going to ask you to take a series of tests. For me, this is part of a research project. I'm what's called a graduate student in a doctoral program at Loyola University in Chicago. I study people's abilities, skills, attitudes, interests, and feelings about themselves and others. Your performance on the tests will help me learn more about how people your age can do these types of tasks. Some of the tests will be used by your teachers and myself to improve the way in which you personally are taught at school. That is, the tests will help us to provide things in school that meet better your interests and needs. How you do on the tests will be confidential and no one else in class will learn how you did or what you said. OK, let's start.

By this introduction it was hoped that a fair amount of rapport was established with the children. Cozby (1973), for instance, has noted that tester or interviewer self-disclosure tends to enhance subject cooperativeness.

It also can be noted that some results were shared with the school principals and teachers. They were provided with the scores on the OLMAT. They were also given a description of the meaning and

limitation of the scores. It was suggested that any children, about whom the teachers were concerned on the basis of school achievement, class performance, and the OLMAT scores, be referred for individual assessment. In doing this, it was hoped that the inappropriate use of the group mental ability test was avoided. The teachers hoped to use the scores as an additional piece of information in planning class and individual learning activities.

The HFD test was administered to each class in the following manner. First, as a warm-up task, the children were asked to "draw a tree" with a No. 2 pencil on an 8-1/2 x 11 inch blank sheet of paper. A tree was selected for the warm-up drawing for it is generally thought to be a less psychologically threatening task than the person drawing (e.g., Hammer, 1971). Next, the youngsters were requested to draw on a second sheet of paper "a whole person," as indicated in the testing directions provided by Koppitz (1968). Finally, the children were asked to fill out a one-page questionnaire on their interests and family backgrounds. Twelve minutes were allowed for each drawing. The total time of this procedure was roughly 40 minutes.

The CTP and the OLMAT were each given to the students in a class grouping. The children marked their responses to the items on separate standard answer sheets. Standard instructions for the OLMAT were provided and the test was given in the usual 40-minute period. Standard procedures for the CTP were followed with one exception. Because it was discovered in a pilot trial that below-average children were unable to read some CTP items yet were

generally able to understand the items when read to them, each CTP item was presented twice by tape recording while the students were asked to read along. This procedure lengthened the testing time to 1 hour, slightly longer than the usual 45-minute administration.

Scoring. Both the CTP and the OLMAT were scored by hand as described in the manuals of the tests. Three scores reflecting emotional adjustment (TA, PA, and SA) were recorded from the CTP. The OLMAT provided one score indicative of mental maturity.

As for the HFDs, after the "tree" drawings had been set aside, the HFD of each subject was evaluated by means of three measuring devices: the EIs, the EEs and the MHN formula. Each of these three scoring methods was used by a pair of undergraduates, who volunteered from advanced undergraduate psychology courses. The students were considered to be intelligent and highly motivated. Each pair of judges included a male and a female in an attempt to avoid sex-related scoring biases. Each rater of the pairs scored 85 drawings. This allowed for an overlapping set of 30 drawings scored by each pair of raters. Thus, a pair of judges scored the same group of 30 randomly selected drawings. The interjudge reliability of each scale was derived from the scores on this group of drawings.

Each pair of judges participated in a three-step training procedure: (a) study of the scales and drawings illustrating the various items; (b) practice scoring on a broad range of drawings and discussion of scoring between the judges; and (c) practice scoring on a range of drawings similar to those found in the main pool of drawings and extended discussion to clarify the meaning of the items.

In some cases additional scoring rules were adopted. These are given in brackets in the Appendices.

In working with the EI scale, the judges read the descriptions of the characteristics provided by Koppitz (1968) and reviewed the drawings which she included to illustrate the features. After this, the judges practiced their scoring on 25 drawings from another sample of subjects (roughly equal number of each sex). This first group of practice drawings included drawings obtained from severely disturbed children of the Loyola University Day School and from a seventh class of students (from one of the three schools cooperating in the study) whose test data were merely used to provide drawings for training. Thus the drawings in the first group came from a broad range of students. After comparing scoring of the first set of HFDs, the judges then scored a set of 25 drawings exclusively from the seventh class of students. It was thought that this group of drawings were obtained from students like those in the main pool of drawings. It was hoped that the practice scoring would minimize practice effects on the main body of drawings. In addition, each judge was presented the main set of drawings in a different random order to avoid any effects due to a particular order of HFDs.

With the MHN formula, a similar procedure was followed. First the judges were provided with the item descriptions of Hiler and Nesvig (1965) and of Stricker (1967) and were shown drawings among the Koppitz (1968) selections which illustrated those features. The present researcher picked drawings which, in his subjective judgment, showed "nothing pathological." Then the raters scored the two

sets of practice drawings and discussed their results before proceeding to the main body of HFDs.

For the EE scale, the descriptions of the items as given by Koppitz (1968) were studied as were the drawings which she thought depicted EE items. For the EE judges as for the EI and the MHN judges, the sample sets of drawings were scored and discussed. Then the main body of drawings were rated by each scorer separately.

RESULTS

The main findings of the study are presented in the modified multitrait-multimethod matrix in Table 2.

Insert Table 2 about here

In Table 2 two methods are represented: drawings and objective tests. Three scores purportedly indicate the construct of emotional adjustment: Koppitz's 30 Emotional Indicators (EIs), the Modified Hiler Nesvig (MHN) formula, and the California Test of Personality--Total Adjustment (CTPTA) score. Two scores purportedly indicate the construct of mental maturity: the Koppitz Expected and Exceptional (EE) Items and the Otis-Lennon Mental Ability Test (OLMAT). Except for the correlations in the reliability diagonals, all correlations are product-moment correlations based on the scores of the 136 children who took all three tests.

Reliability. Interjudge reliabilities were obtained for the three pairs of male and female raters. The product-moment correlations were .90, .78, and .76 for the EE, the EI, and the MHN scales respectively. These values are acceptable for this type of rating and indicate a fairly good agreement between judges as to the presence or absence of specific items. The EE scale value is a high value and may be a characteristic of the scale or the particular pair of individuals doing the ratings. In any case the level of agreement on item scoring was at an acceptable level.

TABLE 2

Modified multitrait-multimethod matrix providing correlations of various scale scores.

	<u>Drawings</u>			<u>Objective Test</u>	
	<u>EE</u>	<u>EI</u>	<u>MHN</u>	<u>OLMAT</u>	<u>CTPTA</u>
<u>EE</u>	.17*				
<u>EI</u> ^a	.23**	.18*			
<u>MHN</u>	.06	.46***	.43***		
<u>OLMAT</u>	.10	.09	.23**	(.93)	
<u>CTPTA</u>	.08	.01	.17*	.25**	(.94)

^a The direction of the EI scale has been reversed so that higher scores reflect better adjustment as with the MHN formula and the CTPTA.

* $p < .05$

** $p < .01$

*** $p < .001$

The reliability coefficients for the EI and MHN scales were computed using the Kuder-Richardson method (Bruning & Kintz, 1968). The EI and the MHN reliability coefficients were derived from the protocols of 140 students. The reliability coefficient for the EE scale is a weighted average of the four Kuder-Richardson coefficients for the male/female nine-/ten-year-old groups. The EE reliability coefficient was based on the protocols of 139 children. The results of one 11-year-old child were not included in computing the weighted-average coefficient for the EE scale. The reliability coefficients for the OLMAT and the CTPTA were obtained from the test manuals and were computed with the Kuder-Richardson formula.

The reliability coefficients for the Koppitz EE and EI scales are .17 and .18 respectively, significant at the .05 level. However, it can be noted that these are very low for this type of coefficient. This suggests little internal consistency in the scales. The MHN coefficient was .43, significant beyond the .001 level. This may be considered low for a coefficient of internal consistency. Usually values in the .70's or .80's are considered acceptable. The .93 and .94 values for the OLMAT and the CTPTA respectively, indicate high levels of internal consistency and suggest that these scales measure unitary traits.

Convergent Validation. Campbell and Fiske (1959) noted that in a multitrait-multimethod matrix the validity coefficients are expected to be greater than zero. This is understood as being indicative of convergent validity. In the present study, this is the case for the MHN scale, but not the EI and EE scales of Koppitz. The

monotrait-heteromethod coefficient of correlation between the MHN scale and the CTPTA, the criterion of emotional adjustment, was .17. This is significant beyond the .05 level. The validity coefficient of correlation between the EI scale and the CTPTA was a non-significant .01. Likewise, the validity coefficient of .10 for the EE scores and the OLMAT, the criterion index for mental maturity, was non-significant.

It can be noted that although the validity coefficient for the MHN scale is significant, the magnitude of the correlation must be considered rather low in comparison to the more ideal results presented by Campbell and Fiske (1959). Values in the .40's through the .60's are usually expected.

To some extent the correlation between the EI scale and the MHN formula can be considered a validity coefficient in that both scales purportedly measure the same trait assessed by a somewhat different method: a different scale. In part, however, the methods are the same in that both scores were derived from the drawings. The obtained correlation of .47 is significant beyond the .001 level. This lends some additional support for the notion that the two scales measure the same trait. However, the relationship is not very strong in light of the fact that the measures were derived in part from the same method.

Discriminant Validation. Campbell and Fiske (1959) presented three criteria of discriminant validity. First, it is expected that the validity coefficients are greater than the heterotrait-heteromethod coefficients. Thus, it was expected in this study that the correlation of the EE scale and the OLMAT scores would exceed the correlations

of the EIs and the OLMAT, of the MHN formula and the OLMAT, and of the EE items and the CTPTA. Employing tests for differences between dependent correlations (Bruning & Kintz, 1968) no significant differences were found for the data of the present study. Thus the correlation of EE mental maturity with the OLMAT mental maturity was not greater than the relationships of HFD emotional adjustment and OLMAT mental maturity or the relationship of EE mental maturity and CTPTA emotional adjustment.

Likewise, it was expected that the correlation of the EIs and the CTPTA scores would exceed the correlations of the EIs and the OLMAT, of the MHN formula and the OLMAT, and of the EE scale and the CTPTA. Utilizing tests for differences between dependent correlations, the correlation between the EIs and the CTPTA was not significantly greater than the correlations of the EIs and OLMAT or of the CTPTA and the EE scores. Multiple regression analysis (Kerlinger & Pedhazur, 1973) was employed to convert the correlations of the EIs and the CTPTA and of the MHN formula and the OLMAT to partial correlations. These were then compared as independent correlations (Bruning & Kintz, 1968). The results were opposite of those predicted. The partial correlation of the MHN scores and the OLMAT was significantly greater than that of the EIs and the CTPTA, $z = -2.25$, $p < .05$.

In a similar fashion, it was anticipated that the correlation of the MHN formula and the CTPTA was greater than the correlations of the MHN formula and the OLMAT, of the CTPTA and the EE scale, and of the OLMAT and the EIs. Tests of the differences between correla-

tions yielded no significant results.

Overall then, the hypotheses based upon the first criterion of discriminant validity found no significant support with the present data. That is, none of the validity coefficients exceeded the heterotrait-heteromethod coefficients.

The second criterion of discriminant validity proposed by Campbell and Fiske (1959) requires the comparison of the values of the validity coefficients with those of the heterotrait-monomethod coefficients. The former were hypothesized to exceed the latter in this study. Following this line of reasoning, the correlation of the EE scale and the OLMAT was compared to the EE scale and the MHN formula and of the EE scale and the EIs. In statistical tests neither of these comparisons yielded significant results. In addition, the strength of the relationship of the EIs and the CTPTA was compared to that of the OLMAT and the CTPTA. The results were contrary to the predicted outcome. The correlation of the OLMAT and the CTPTA was greater than that of the EIs and the CTPTA, $t(133) = -2.06$, $p < .05$. It was also predicted that the correlation of the MHN formula and the CTPTA would exceed that of the CTPTA and the OLMAT. This was not found to be the case by statistical test.

Insofar as the different rating scales may be seen as partially different methods, it was expected that the correlation of the EIs and the MHN formula would exceed those of the EIs and the EE items and of the MHN formula and the EE items. As predicted, the correlation of the EIs and the MHN formula was greater than that of the EIs and the EE items, $t(133) = 2.33$, $p < .05$. So too, the

correlation of the EIs and the MHN formula was greater than that of the MHN formula and the EE items, $t(133) = 4.22$, $p < .01$.

Overall, however, by the second criteria of discriminant validity the results provide little support for the validity of the drawing scales. Only within the drawing method do we find some suggestion that two distinct traits are being assessed by the HFD scales.

Campbell and Fiske (1959) mentioned a third criterion for discriminant validity--the presence of a similar pattern of trait interrelationships in the monomethod and heterotrait groups. No clear pattern can be said to emerge with the present data.

In summary, little if any support for the discriminant validity of the drawing scales was found with the present data according to the three criteria of Campbell and Fiske.

Comparison of the EI Scale and the MHN Formula. Contrary to prediction, the Koppitz EI scale was not more strongly correlated with the objective criterion than was the MHN formula. If any effect were present, there was a tendency for the MHN formula to surpass the EI scale, $t(133) = -1.80$, $p < .10$.

Item Overlap. An additional prediction was that, because of the item overlap of the three drawing scales, the correlations among the drawing scales were expected to exceed the correlation between the objective tests. However, in the present study this was not found to be consistently the case. The correlation of the EIs and the MHN formula was greater than that of the OLMAT and the CTPTA, $z = 2.42$, $p < .05$. This test, as well as the others, was made by means of multiple regression analysis and comparison of partial

correlations as independent correlations. The correlation of the EE scale and the MHN formula was not significantly greater than that of the objective tests. Furthermore, contrary to prediction, the correlation of the OLMAT and the CTPTA was larger than that of the EE scale and the EIs, $z = 2.36$, $p < .05$.

Sex Differences. Contrary to prediction, the boys in this study did not produce significantly more EIs than the girls. However, as predicted, males scored more pathologically on the MHN scale, $t = 3.48$, $p < .01$. A rather surprising result was found in the fact that males scored higher on the EE scale than females, $t(138) = 2.00$, $p < .05$. It had been expected that boys and girls would be equivalent because drawing protocols are scored differently for each sex to render them equivalent.

It was also predicted on the basis of Koppitz's (1969) research that boys would produce seven specific items more frequently than girls. However, in the present study only two of the items were made more frequently by boys. Males drew more "teeth" ($t = 2.66$, $p < .01$) and "monster/grotesque" figures ($t = 2.80$, $p < .01$). The production of "poor integration," "shading of hands/neck," "transparencies," "tiny figure," and "arms clinging to body" was not greater in the drawings of either boys or girls.

Internalizer/Externalizer Hypotheses. Six drawing items were predicted to correlate more with pathological scores on the CTPTA than with those on the CTPSA. Of these items only one, "short arms," was found to correlate more with pathological scores on the CTPPA than the CTPSA, $t(137) = 1.99$, $p < .05$. The item, "short arms," was

present in the drawings of 16% of the subjects. Tests for differences of correlations for the following yielded nonsignificant results: tiny figure, hands cut off, clouds, no nose, and no mouth.

Six drawing items were expected to correlate more strongly with poorer scores on the CTPSA scale. Only one item, "gross asymmetry of limbs," was found to correlate more with pathological scores on the CTPSA than the CTPPA, $t(137) = 5.06$, $p < .01$. This sign was obtained in the protocols of 5% of the subjects. Tests for two characteristics yielded non-significant results: long arms and genitals. Contrary to prediction, "teeth" was found to correlate more with scores of better adjustment on the CTPPA than the CTPSA, $t(137) = 2.08$, $p < .05$. The item, "teeth," was scored in 10% of the drawings of the children. Another unexpected result was that one feature, "big hands," tended to correlate more strongly with the CTPSA than the CTPPA in the direction of healthy adjustment, $t(137) = 5.10$, $p < .01$. However, "big hands" was noted in only 1% of the drawings. In addition, one item, "no neck," correlated more with negative scores on the CTPPA rather than the predicted CTPSA, $t(137) = 1.99$, $p < .05$. The item, "no neck," was obtained in the protocols of 12% of the students.

Thus the hypotheses that certain drawing items tend to relate to internalization as measured by the CTPPA or to externalization as assessed by the CTPSA did not receive broad support in the present study. Most predictions obtained non-significant results, and significant results were roughly as often in unpredicted directions as in predicted ones.

DISCUSSION

In their reviews of the literature, Swensen (1968) drew a more optimistic picture of the utility of HFDs in assessing personality than Roback (1968) did. Many of the studies, which Swensen considered as supportive of the validity of HFDs, were those of Koppitz (1966a, b, c; 1967; 1968) and of Hiler and Nesvig (1965). The results of the present research are not generally supportive of the validity of the Koppitz scales and only moderately of the formula of Hiler and Nesvig. In light of the present study, another look at the assumptions of Koppitz and of Hiler and Nesvig seems to be in order.

Koppitz EI Scale. The derived inter-judge correlation of .78 represents an acceptable level of agreement as to the presence of specific items. Also, it suggests that with sufficient training well motivated and intelligent undergraduates or teachers could learn to score the HFDs of children. This would be an important consideration if the HFDs were to be used as a screening device or part of one in making referrals to a counselor, psychologist, or guidance clinic for further assessment and treatment.

However, the results of the present study tend to indicate that the Koppitz EI scale represents a rather poor device for assessing emotional adjustment, at least of fourth-grade children from working- and middle-class backgrounds. To begin with, the internal consistency of the scale as reflected in the .18 Kuder-

Richardson coefficient is at a low level. This value suggests that the scale does not clearly measure a unitary trait. Consequently, the validity of the scale is limited because high reliability is a necessary though not sufficient requirement for high validity. In addition, the EI scale was supported by neither the main convergent nor discriminant patterns in the correlation matrix obtained from the results. Only the .47 correlation between the EI scale and the MHN formula may buttress the notion that the EI scale measures emotional adjustment. However, this support seems to be rather weak. First, the relationship accounts for only 22% of the variance. And secondly, since the MHN formula does not correlate strongly with the CTPTA criterion, it may well be argued that, although there is a moderate correlation between the EI scale and the MHN formula, it may not relate to emotional adjustment but merely some common variance of the scales.

This discussion has assumed, of course, that the CTPTA provides a very good criterion for emotional adjustment. As Fiske (1971) has noted, the utilization of several modes of gathering information is desirable. Although this is certainly true, within the limits of the present study it seems safe to say that the use of the CTP is roughly as good as any other single method (e.g., Jackson, 1946; Smith, 1958).

It is noteworthy that in several studies, the CTP demonstrated an ability to discriminate among criterion groups even when other related factors such as intelligence and school achievement were held constant. And it is possible that it is variables of this type

which provide a confound in several of the studies which were aimed at validating the EI scale. Koppitz (1966b) failed to match her criterion group for intelligence, thus intelligence or mental maturity is a possible confound in her study. Indeed Koppitz's "normal" comparison group were pupils selected by their teachers as "outstanding all around" pupils. Koppitz admitted that the students were probably of high average or superior intelligence. Thus, it is possible that the EI items which differentiated her disturbed group from her comparison groups were related to intellectual or artistic skills rather than emotional adjustment.

The results of Fuller, Preuss, and Hawkins (1970) may be accounted for in similar fashion. These authors failed to match their disturbed and normal groups for intelligence. Consequently, their results may be due to differences in intellectual or cognitive skills. Furthermore, if one assumes that the EI scale is confounded by intelligence, the results of Hall and Ladriere (1970) seem more understandable. As will be further explained in the next section, it is possible that both the EI and the EE scales reflect a blend of intellectual, emotional, and motor skills and that this explains the relative equivalence of the EI and the EE scales in distinguishing among emotionally disturbed, perceptually handicapped, and control groups as found by Hall and Ladriere.

Koppitz EE Scale. In the present study, the EE items were scored very reliably by the judges. This is reflected in the .90 inter-judge reliability correlation. However, the scale demonstrated little internal consistency as evidenced by a .17 weighted-average

Kuder-Richardson coefficient. This coefficient suggests that there is little reason to consider the scale as measuring a unitary trait such as mental maturity. In other comparisons of correlations, the main tests for neither convergent nor discriminant validity yielded support for the scale. Only within the drawing method was there some suggestion of a discriminant pattern. Thus, judging from the results of the present study, one has difficulty in seeing the value of the EE scale in assessing the drawings of fourth-grade pupils from cultural and socio-economic backgrounds like those of the present subjects.

These findings provide a new stimulus to re-examine the results of Koppitz (1968) and DeMoreau and Koppitz (1968). In her validating study for the EE scales, Koppitz (1967) obtained correlations between the EEs and the WISC of .68 and .45 for nine- and ten-year-old boys, respectively. However, because of the nature of the sample of subjects, it is possible that the correlations are artificially inflated. All the subjects had been referred for a psychological evaluation either in a psychiatric clinic or in a school. It seems possible that the relationship between the EE scale and the WISC may be exaggerated within an emotionally disturbed sample of subjects because of the selected nature of subjects in the validating group. Although Koppitz developed her "expected" and "exceptional" items from the protocols of normal classrooms, she failed to validate her scale on a normal sample of subjects. Within the emotionally disturbed sample which she employed, it is possible that an intelligent and disturbed student does much better than a dull and disturbed student on the EE drawing scale. If this were true, the relationship

between the EE scale and the WISC may be inflated over what would be obtained within a group with a broader range of subjects.

Two other aspects of the Koppitz validating study are worth noting. First, all the subjects are males. It is possible that the EE scale and the WISC correlate well for males but not for females. Koppitz did not include data on girls. Secondly, it can be admitted that the WISC tests a broader range of abilities than the OLMAT. The WISC included Performance scales which focus on perceptual and motor skills as well as the Verbal scales which focus mainly on verbal and numerical skills. The HFD task taps perceptual and motor skills as well as cognitive skills. Consequently, the correlation of the EE scale and the WISC may be greater than the correlation of the EE scale and the OLMAT because of the greater similarity in skills assessed by the HFD test and the WISC than by the HFD test and the OLMAT. This may account for some of the difference between the validity coefficients obtained by Koppitz and the one derived in the present study. Unfortunately, Koppitz (1967) did not present the correlations of the EE scale and the WISC Verbal scores and of the EE scale and the WISC Performance scores.

DeMoreau and Koppitz (1968) presented results comparing the HFD EE scores to Goodenough DAM IQ scores. For 9- and 10-year-olds, the obtained correlations were .67 and .72 respectively. The sample of subjects were Mexican lower-class children from Guadalajara. DeMoreau and Koppitz (1968) understood their results as indicating the validity of the EE scale in that moderately high correlations were obtained. However, as the authors pointed out, "It was also

found that the Goodenough IQ scores are not a valid measure of intelligence for children from under-developed countries or from areas not exposed to Western civilization" (p. 38). Thus, instead of concluding that the EE scale was a valid indicator like the Goodenough DAM IQ, it seems that DeMoreau and Koppitz should have concluded that the EE scale is perhaps as invalid as the Goodenough scores for the sample of children from an underdeveloped area. It is possible that correlations between the EE scale and the DAM IQ scores for a more typical sample of subjects would be lower. Further research is needed to clarify this issue.

Overall then, the present study does not provide much support for the use of the EE scale. Also, it seems that the validity studies by Koppitz (1967) and by DeMoreau and Koppitz (1968) may have faulty conclusions based on poorly chosen samples of subjects. More research is needed to assess whether results similar to those found in the present study will be obtained with children of other ages, socio-economic groups, and geographical areas.

The MHN Scale. Of the three drawing scales compared in the present study, the MHN scale seems the most supported by the data and appears to warrant the most exploration in future studies. The scale was reliably scored by a pair of judges as indicated by the obtained .78 coefficient of correlation. A moderately low homogeneity of .48 was derived using the Kuder-Richardson formula and suggests a fairly low level of internal consistency, yet one which is greater than that found for either the EI or the EE scales.

The convergent validity of the scales was demonstrated by the

.17 validity coefficient. Although significant, this is not a very strong relationship. One might hope for a correlation in the .50's or .60's. None of the main tests for discriminant validity were significant. Within the drawing method, however, some convergence of variance was apparent in the correlation of the EI and the MHN scales as well as some discriminant validity in that the correlation of the EI and MHN scales was greater than either that of the EIs and the EEs or of the MHN formula and the EEs. But the value of these results is questionable as noted in the discussion on the EI scale above. It seems that the validity of the scale is in part limited by its low level of internal consistency.

One reason for the rather poor support for the validity of the scale may be found in the original study by Hiler and Nesvig (1965). These researchers developed their scale by comparing the protocols of patients and non-patients. They reasoned that a group of adolescent patients would tend to have a lower mean IQ than a group of successful adolescents because the emotional problems would be expected to retard their learning. Thus the mean IQ of their patient group was 90.5 while the mean IQ of their successful group was 107.7. This does, however, allow for a possible confound in their study. It is possible that the criteria, which they found as being used by clinicians in making accurate judgments of patient/non-patient status, may reflect IQ differences in the comparison groups rather than differences in emotional disturbance. It seems that their patient and non-patient groups should have been equated for IQ or at least have differed by fewer than 17 points in mean IQ.

The samples upon which the formula was cross-validated were drawn in a similar fashion to the first and can be assumed to have similar IQ differences. The researchers did not report mean IQ for the cross-validation groups.

It is also possible that the original formula was "watered down" by the addition of the criteria of "distortion" and "transparency." The strength of these criteria in discriminating between the disturbed and successful groups was only at the .05 level. Also, it may be mentioned that Hiler and Nesvig found "transparency" to be a valid criterion only when it was "very obvious." It must be admitted that the judges employed in the present study scored any transparency of the arms, legs, or body which was clearly visible. Thus, the criterion for "transparency" may have weakened the strength of the relationship between emotional disturbance and the CPTA. However, it can be noted that in a recent study of several drawing items Prytula and Thompson (1973) found only transparencies to be significantly more frequent in the drawing of low self-esteem subjects than in those of high self-esteem children. Thus, the Hiler-Nesvig formula may not have been weakened by the addition of "transparency."

More research is warranted for the Hiler-Nesvig formula. Different ages, sexes, and socio-economic groups may be explored. Variations in the presence of scale items has been found for different ages (e.g., Koppitz, 1966b, 1967; Vane & Eisen, 1962), sexes (e.g., Machover, 1960; Koppitz, 1966b, 1967, 1969), and socio-economic groups (Koppitz & DeMoreau, 1968; Koppitz, 1969) with other scales. Consequently, similar variations may be anticipated with the Hiler-Nesvig

formula. In addition, other criteria beside the CTP need to be employed in evaluating the MHN formula. However, it needs be admitted that it is possible that the drawing task by its very nature is too amorphous to act as a good test of emotional adjustment.

Sex Differences. Boys received more pathological scores than girls on the MHN formula, as predicted, but not on the EI scale, contrary to prediction. In consideration of the greater apparent validity of the MHN scale in evaluating emotional adjustment, it may be said that the results tend to agree with the typical depiction of elementary school boys as less well adjusted in a female-dominated school environment. Machover (1960) noted in her observations of 9- and 10-year-olds that boys experienced a mixture of dependency and anger and strained toward manliness in a female-controlled school while girls were more passive and concerned with practical feminine virtues and the development of their bodies.

The difference between the MHN scale and the EI scores for boys and girls may be related to specific items present on one scale but not on the other. For example, "pleasant, facial expression" is part of the MHN scale, but not part of the EI scale. It may be the case that girls more frequently depict a happy expression or other items than boys.

In using her scale, Koppitz (1969) had found that boys illustrated more EIs. She thought that this indicated that boys were more impulsive, aggressive, anxious, and inadequate than girls. It is noteworthy that teachers acted as administrators in her study. Assuming that most, if not all, of the teachers in the Koppitz (1969)

study were females and pointing out that the administrator in the present study was male, may lead one to consider a possible interaction between administrator sex and the sex differences. Other sources of possible influence include cultural background, family background, and socio-economic level.

It had been predicted that girls and boys would have equivalent average scores on the EE scale. This was not found to be the case. The boys performed at a higher level on the test. Of course, the exact meaning of this result appears to be as cloudy as the meaning of the scale. This result does, however, suggest that the normative data employed by Koppitz (1968) needs to be broadened.

Of the specific items from the EI scale which were predicted to occur more frequently in the protocols of boys only two did so. The items "teeth" and "monsters" were more frequent in boys' drawings. This finding is consistent with Machover's (1960) observation that boys are more concerned with dependency and aggression as reflected in larger mouths and the presence of teeth. Koppitz (1968) did not consider the presence of teeth to be a sign of serious pathology and thought their presence indicated aggressiveness. The presence of monsters may suggest a greater degree of difficulty for boys to attain a clear self-concept in the school environment. Koppitz (1968) interpreted monsters as reflecting intense feelings of inadequacy and a very poor self-concept. Of course, different interpretations of "teeth" and "monsters" may be made from other theoretical viewpoints.

Internalizer/Externalizer Hypotheses. The results as regards

the predictions generated by the internalizer/externalizer concept were mixed. Only two predictions found support in the data of the present study. "Short arms" appeared in the drawings of children with more pathological scores on the CTPPA than on the CTPSA. It had been reasoned that children with more pathological scores on the CTPPA than on the CTPSA would internalize their conflicts rather than act out in response to them. The CTPPA scale includes subscales entitled "Sense of Personal Worth," "Withdrawing Tendencies," and "Nervous Symptoms" which can be expected to be accentuated in the tests of shy and psychosomatic children (e.g., Jackson, 1946). Thus some support is given to the notion that children who draw "short arms" tend to internalize their conflicts.

It was also found that "gross asymmetry of limbs" correlated more strongly with pathological scores on the CTPSA scale than on the CTPPA scale. The CTPSA scale included subscales entitled "Anti-social Tendencies," "School Relations," and "Community Relations." It was reasoned that children who are aggressive or delinquent (caught stealing) tend to externalize their conflicts and would receive relatively higher scores on the CTPSA scale. Such an inference would be consistent with the findings of Jackson (1946) and Peak (1963). Thus some support is given to the notion that children who draw "gross asymmetry of limbs" tend to externalize their problems. One item, "no neck," tended to correlate more strongly with negative scores on the CTPPA than on the CTPSA. This suggests that this sign may be more associated with children who internalize their problems rather than externalize them.

But these conclusions are best made cautiously, for other predictions based on the same reasoning were not supported. In fact, the presence of two items tended to correlate with healthier adjustments. One item, "teeth," tended to correlate more strongly in the direction of better adjustment on the CTPPA than the CTPSA. Another, "big hands," tended to correlate more with healthy adjustment on the CTPSA than on the CTPPA. However, this relationship may be spurious because "big hands" appeared in only 1% of the drawings. In any case, "teeth" and "big hands" may be seen as signs of positive personal and social adjustment, respectively, at least for the sample of subjects used in this study. Consequently, a question can be raised as to the appropriateness of the items on the Koppitz scale. It is noteworthy that neither "teeth" nor "big hands" appeared in the HFDs of the disturbed groups at a .05 level of significance in either the study by Koppitz (1966b) or the one by Fuller, Preuss, and Hawkins (1970). Hence, the appropriateness of "teeth" and "big hands" on an adjustment scale may be questioned. The absence of these items on the MHN formula may have contributed to the greater relative validity of the MHN formula over the EI scale.

Conclusion. Of the three scales evaluated in the present study, the MHN scale tended to receive the most support for its validity. Yet, strong signs of discriminant validity were lacking for all three scales. In addition, the research of Koppitz and of Hiler and Nesvig appears to have several possible flaws. This appears to render many of their conclusions questionable. For all three HFD scales, more research which balances the effects of intelligence in

criterion groups is needed. The multitrait-multimethod matrix is one design that may be employed to sort out the effects of several variables. More research is needed for an evaluation of the drawings of children of different ages, geographical and cultural backgrounds, socio-economic levels and family backgrounds. It can be recalled that the study by Vane and Eisen (1962) found that four items (grotesque, no body, no mouth, and no arms) appeared more frequently in the drawings of poorly adjusted children than in those of fair and well adjusted children, even when matched for IQ. Also, Starkey (1970) found support for the convergent validity of a cluster of "aggressive" EIs against two criterion methods and for a cluster of "emotionally unstable" EIs against one criterion method. Furthermore, Hall and Ladriere (1970) did find with groups matched for IQ that the EIs and the EE scale distinguished between the control group and the emotionally disturbed group as well as between the control group and the perceptually handicapped group. So further research does seem warranted for the HFD scales.

In future studies, other objective tests may well be used for assessing the traits of emotional adjustment and intelligence. In addition, the approach of Starkey (1970) may profitably be utilized to see whether certain clusters of drawing characteristics relate to certain traits which make up the global "emotional adjustment." These might include traits such as aggressiveness, anxiety, withdrawal, depression, and so on. Such an approach would be more consistent with the manner in which Koppitz (1968) tended to interpret drawing characteristics. The use of the internalizer/externalizer

dimension in the present research was one attempt to survey subgroups of EI items which reflect personality variables. Moreover, as Fiske (1971) emphasized, other methods or modes beside self-report can be explored.

However, at present the usefulness of HFDs is unclear. Only the MHN formula was moderately supported by the data reviewed herein. Nevertheless, it can be stated that drawings are an enjoyable task for children and their use tends to establish rapport. Possibly some combination of "free response" and "inquiry" approach might improve upon the useful information derived from the drawings. Or perhaps, some combination of drawing and story telling may provide information that most validly relates to personality variables. It may eventually be thought that drawings are best relied upon during the primary grades before children have acquired the skills with which to take more specific verbal or motor tests. In any case, it is likely that drawings will continue to fascinate the practicing clinician because of the lure of a rare "find" in the sketch by a client.

SUMMARY

This research employed a multitrait-multimethod design to investigate the convergent and discriminant validity of human figure drawings of children as indicators of emotional adjustment and mental maturity. The Koppitz (1966b) EI scale and a Modified Hiler-Nesvig (1965) formula were used to assess emotional adjustment in the drawings, while the Koppitz (1967) EE items were utilized to assess mental maturity. After training, a pair of opposite-sexed advanced undergraduates acted as judges for each scale. The self-report criterion for emotional adjustment was the California Test of Personality, whereas the standard criterion for mental maturity was the Otis-Lennon Mental Ability Test. One hundred thirty-six fourth-grade children from three school systems acted as subjects. The main pattern of convergent validity was apparent only for the MHN scale. None of the scales were supported by the main expected patterns of discriminant validity. Only within the drawing method, if one considers the separate scales as partly different methods, was some convergence of variance for the EI and the MHN scales and some support for the discriminant validity of the scales found. Yet these patterns are open to other interpretations. Although the pairs of judges demonstrated reliable agreement in their scoring of items on the three scales, the internal consistency coefficients were low, especially for the Koppitz EI and EE scales. This low internal consistency was thought to limit the validity of the scales.

The predictions of sex differences for the total EI scale and the EE scale were not supported. However, boys scored more pathologically on the MHN formula. Boys only produced two EI items more frequently than girls: teeth and monster/grotesque figures. Predictions of the relationships between drawing items and the internalizer/externalizer construct obtained mixed results. Support was found for "short arms" as an internalizer item and for "gross asymmetry of limbs" as an externalizer characteristic.

In light of the results of this study, it was thought that the validating research of Koppitz (1966b) and of Hiler and Nesvig (1965) may be confounded by a lack of control for intelligence. The research supporting the EE scale, Koppitz (1967) and DeMoreau and Koppitz (1968) may well have employed inappropriate sampling groups.

Further research is needed to clarify these conjectures and may profitably explore the relationships among personality traits and criterion methods.

REFERENCES

- Adler, P. Evaluation of the figure drawing technique: Reliability, factorial structure, and diagnostic usefulness. Journal of Consulting and Clinical Psychology, 1970, 35, 52-57.
- Anastasi, A. Goodenough-Harris Drawing Test. In O. K. Buros (Ed.), The Seventh Mental Measurement Yearbook. Highland Park, N.J.: Gryphon Press, 1972.
- Buck, J. The H-T-P technique: A qualitative and quantitative scoring manual. Journal of Clinical Psychology, 1948, Monograph Supplement, No. 5.
- Bruning, J., & Kintz, B. Computational handbook of statistics. Glenview, Ill.: Scott, Foresman and Company, 1968.
- Campbell, D., & Fiske, D. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.
- Carlson, K., Quinlan, D., Tucker, G., & Harrow, M. Body disturbance and sexual elaboration factors in figure drawings of schizophrenic patients. Journal of Personality Assessment, 1973, 37, 56-63.
- Cauthen, N., Sandman, C., Kilpatrick, D., & Deabler, H. DAP correlates of Sc scores on the MMPI. Journal of Projective Techniques and Personality Assessment, 1969, 33, 262-264.
- Chapman, L., & Chapman, J. Genesis of popular but erroneous psychodiagnostic observations. Journal of Abnormal Psychology,

- 1967, 72, 193-204.
- Cozby, P. Self-disclosure: A literature review. Psychological Bulletin, 1973, 79, 73-91.
- DeMoreau, M., & Koppitz, E. Relationship between Goodenough Draw-A-Man Test IQ scores and Koppitz Human Figure Drawing scores. Revista Inter-americana de Psicologia, 1968, 2, 35-40.
- Dunn, J. Goodenough-Harris Drawing Test. In O. K. Buros (Ed.), The seventh mental measurement yearbook. Highland Park, N.J.: Gryphon Press, 1972.
- Fiske, D. Measuring the concepts of personality. Chicago, Ill.: Aldine, 1971.
- Fuller, G., Preuss, M., & Hawkins, W. The validity of the HFDs with disturbed and normal children. Journal of School Psychology, 1970, 8, 54-56.
- Goodenough, F. Measurement of intelligence by drawings. New York: Harcourt, Brace & World, 1926.
- Gravitz, M. Figure drawing size as an index of depression and MMPI depression scores in normal adults. Journal of Clinical Psychology, 1969, 25, 77-79.
- Hall, L., & Ladriere, M. A comparative study of a diagnostic potential and efficiency of six scoring systems applied to children's figure drawings. Psychology in the Schools, 1970, 7, 244-247.
- Hammer, E. The clinical application of projective drawings. Springfield, Ill.: Charles C. Thomas, 1958.
- Hammer, E. The House-Tree-Person (H-T-P) Drawings as a projective technique with children. In A. Rabin & M. Haworth (Eds.),

- Projective techniques with children. New York: Grune & Stratton, 1960.
- Hammer, E. DAP: Back against the wall? Journal of Consulting and Clinical Psychology, 1969, 33, 151-156.
- Hammer, E. The Horse-Tree-Person Projective Drawing Technique: Content interpretation. In E. Hammer (Ed.), The clinical application of projective drawings. Springfield, Ill.: Charles C. Thomas, 1971.
- Handler, L., & McIntosh, J. Predicting aggression and withdrawal in children with the Draw-A-Person and Bender-Gestalt. Journal of Personality Assessment, 1971, 35, 331-335.
- Handler, L., & Reyher, J. The effects of stress on the Draw-A-Person Test. Journal of Consulting Psychology, 1964, 28, 259-264.
- Harris, D. Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace & World, 1963.
- Hiler, E., & Nesvig, D. An evaluation of criteria used by clinicians to infer pathology from figure drawing. Journal of Consulting Psychology, 1965, 29, 520-529.
- Holt, R. Clinical and statistical prediction: A reformation and some new data. Journal of Abnormal and Social Psychology, 1958, 56, 1-12.
- Jackson, J. The relative effectiveness of paper-pencil test, interview, and ratings as techniques for personality evaluation. Journal of Social Psychology, 1946, 23, 35-54.
- Johnson, J. Note on the validity of Machover's indicators of anxiety. Perceptual and Motor Skills, 1971, 33, 126.

- Jolles, I. A study of the validity of some hypotheses for qualitative interpretation of the H-T-P for children of elementary school age: I. Sexual identification. Journal of Clinical Psychology, 1952, 8, 113-118.
- Kerlinger, F., & Pedhazur, E. Multiple regression in behavioral research. New York: Holt, Rinehart, and Winston, 1973.
- Kimber, M. The insight of college students into the items on a personality test. Educational and Psychological Measurement, 1947, 7, 411-420.
- King, P., & Ross, D. Test transparency as related to test response. Personnel and Guidance Journal, 1965, 43, 669-73.
- Koppitz, E. A comparison of pencil and crayon drawings of young children. Journal of Clinical Psychology, 1965, 21, 191-194.
- Koppitz, E. Emotional indicators on human figure drawings and school achievement of first and second graders. Journal of Clinical Psychology, 1966, 22, 481-483. (a)
- Koppitz, E. Emotional indicators on human figure drawings of children: A validation study. Journal of Clinical Psychology, 1966, 22, 313-315. (b)
- Koppitz, E. Emotional indicators on human figure drawings of shy and aggressive children. Journal of Clinical Psychology, 1966, 22, 466-469. (c)
- Koppitz, E. Expected and Exceptional items on Human Figure Drawings and IQ scores of children age 5 to 12. Journal of Clinical Psychology, 1967, 23, 81-83.
- Koppitz, E. Psychological evaluation of children's Human Figure

- Drawings. New York: Grune & Stratton, 1968.
- Koppitz, E. Emotional Indicators on Human Figure Drawings of boys and girls from lower- and middle-class backgrounds. Journal of Clinical Psychology, 1969, 25, 432-434.
- Koppitz, E., & DeMoreau, M. A comparison of Emotional Indicators on Human Figure Drawings of children from Mexico and from the United States. Revista Inter-america de Psicologia, 1968, 2, 41-48.
- Koppitz, E., Sullivan, J., Blyth, D., & Shelton, J. Prediction of first grade school achievement with the Bender Gestalt Test and Human Figure Drawings. Journal of Clinical Psychology, 1959, 15, 164-168.
- Lapkin, B., Hillaby, T., & Silverman, L. Manifestations of the schizophrenic process in the figure drawings of adolescents. Archives of General Psychiatry, 1968, 19, 465-468.
- Levy, S. Projective figure drawing. In E. Hammer (Ed.), The clinical application of projective drawings. Springfield, Ill.: Charles C. Thomas, 1958.
- Lingren, R. An attempted replication of emotional indicators in human drawings by shy and aggressive children. Psychological Reports, 1971, 29, 35-38.
- Machover, K. Personality projection in the drawing of the human figure. Springfield, Ill.: Charles C. Thomas, 1949.
- Machover, K. Human figure drawings of children. Journal of Projective Techniques, 1953, 17, 85-91.
- Machover, K. Sex differences in the developmental pattern of children

- as seen in Human Figure Drawings. In Rabin and Haworth (Eds.), Projective techniques with children. New York: Grune and Stratton, 1960.
- McHugh, A. Children's figure drawings in neurotic and conduct disturbances. Journal of Clinical Psychology, 1966, 22, 219-221.
- Milholland, J. Otis-Lennon Mental Ability Test. In O. K. Buros (Ed.), The seventh mental measurement yearbook. Highland Park, N.J.: Gryphon Press, 1972.
- Nichols, R., & Strumpfer, D. A factor analysis of Draw-A-Person test scores. Journal of Consulting Psychology, 1962, 26, 156-161.
- Otis, A., & Lennon, R. Otis-Lennon Mental Ability Test: Manual for administration. New York: Harcourt, Brace & World, 1967.
- Otis, A., & Lennon, R. Technical handbook. New York: Harcourt, Brace & World, 1969.
- Peak, B. The California Test of Personality: A study of validation. (Doctoral dissertation, Florida State University, 1963). Dissertation Abstracts, 1963, 24, 1281-1282. (University Microfilms No. 63-6360)
- Prytula, R., & Thompson, N. Analysis of emotional indicators in Human Figure Drawings as related to self-esteem. Perceptual and Motor Skills, 1973, 37, 795-802.
- Reznikoff, M., & Dies, R. The use of clothing in human figure drawings. Journal of Clinical Psychology, 1969, 25, 80-81.
- Roback, H. Human figure drawings: Their utility in the clinical psychologist's armamentarium for personality assessment. Psychological Bulletin, 1968, 70, 1-19.

- Semler, I. Relationships among several measures of pupil adjustment. Journal of Educational Psychology, 1960, 51, 60-64.
- Sims, V. California Test of Personality, 1953 revision. In O. K. Buros (Ed.), The fifth mental measurement yearbook. Highland Park, N.J.: Gryphon Press, 1959.
- Smith, L. The concurrent validity of six personality and adjustment tests for children. Psychological Monographs, 1958, 72, 1-30.
- Snyder, R., & Gaston, D. The figure drawings of the first grade child--Item analysis and comparison with Koppitz norms. Journal of Clinical Psychology, 1970, 26, 377-383.
- Solar, D., Breuhl, D., & Kovacs, J. The Draw-A-Person Test: Social conformity or artistic ability? Journal of Clinical Psychology, 1970, 26, 524-525.
- Starkey, T. Convergent and discriminant validity of Human Figure Drawings of children as indicators of psychological traits. Unpublished doctoral dissertation, Loyola University of Chicago, Ill., 1970.
- Stricker, G. Actuarial, naive clinical and sophisticated clinical prediction of pathology from figure drawings. Journal of Consulting Psychology, 1967, 31, 492-494.
- Sundberg, N. The practice of psychological testing in clinical services in the United States. American Psychologist, 1961, 16, 79-83.
- Swensen, C. Empirical evaluations of human figure drawings. Psychological Bulletin, 1957, 54, 431-466.
- Swensen, C. Empirical evaluations of human figure drawings: 1957-

1966. Psychological Bulletin, 1968, 70, 20-44.
- Thorpe, L., Clark, W., & Tiegs, E. Manual: California Test of Personality. Monterey, California: McGraw-Hill, 1953.
- Vane, J., & Eisen, V. The Goodenough Draw-A-Man Test and signs of maladjustment in kindergarten children. Journal of Clinical Psychology, 1962, 18, 276-279.
- Wanderer, Z. Validity of clinical judgments based on human figure drawings. Journal of Consulting and Clinical Psychology, 1969, 33, 143-150.
- Welsh, G. An anxiety index and an internalization ratio for the MMPI. Journal of Consulting Psychology, 1952, 16, 65-72.
- Wesman, A. Intelligent testing. American Psychologist, 1968, 23, 267-274.
- Whitmyre, J. The significance of artistic excellence in the judgment of adjustment inferred from human figure drawings. Journal of Consulting Psychology, 1953, 17, 421-424.
- Wilkinson, A., & Schnadt, F. Human figure drawing characteristics: An empirical study. Journal of Clinical Psychology, 1968, 24, 224-226.
- Witkin, H., Lewis, H., Hertzman, M., Machover, K., Meissner, P., & Wapner, S. Personality through perception. New York: Harper, 1954.
- Young, J. An investigation of the relationship between art quality and clinical judgment on the Draw-A-Person Test. Unpublished doctoral dissertation. Loyola University of Chicago, 1970.
- Zimmerman, I., & Woo-Sam, J. Research with the Wechsler Intelligence

Scale for Children: 1960-1970. Psychology in the Schools,
1972, 9, 232-271.

APPENDIX A

APPENDIX A

Chart of the EE items relevant to the drawings of boys (Bs) and girls (Gs) in the fourth grade is presented below. Each protocol is initially scored "5"; then "+1" is scored for each exceptional item and "-1" for each missing expected characteristic. Scored features for each age and sex are marked with an "x."

	<u>Age 9</u>		<u>Age 10</u>		<u>Age 11</u>	
	<u>Bs</u>	<u>Gs</u>	<u>Bs</u>	<u>Gs</u>	<u>Bs</u>	<u>Gs</u>
<u>Expected Items:</u>						
Head	x	x	x	x	x	x
Eyes	x	x	x	x	x	x
Nose	x	x	x	x	x	x
Mouth	x	x	x	x	x	x
Body	x	x	x	x	x	x
Legs	x	x	x	x	x	x
Arms	x	x	x	x	x	x
Feet	x	x	x	x	x	x
Arms 2 dimensions	x	x	x	x	x	x
Legs 2 dimensions	x	x	x	x	x	x
Hair		x	x	x	x	x
Neck		x	x	x	x	x
Arms down			x	x	x	x
Two clothing items				x		x

	<u>Age 9</u>		<u>Age 10</u>		<u>Age 11</u>	
	<u>Bs</u>	<u>Gs</u>	<u>Bs</u>	<u>Gs</u>	<u>Bs</u>	<u>Gs</u>
Arm at shoulder						x
<u>Exceptional Items:</u>						
Knee	x	x	x	x	x	x
Profile	x	x	x	x	x	x
Elbow	x					
Two Lips	x		x			
Nostrils	x					

Description of the EE items as provided by Koppitz (1968) is given below. Additional rules are in brackets.

Expected Items:

Head: Any representation, clear outline of head required.

Eyes: Any representation.

Nose: Any representation.

Mouth: Any representation.

Body: Any presentation, clear outline necessary.

Legs: Any presentation; in case of female figure in long skirts this item is scored if distance between waist and feet is long enough to allow for legs to be present under the skirt.

Arms: Any representation.

Feet: Any representation.

Arms in two dimensions: Both arms presented by more than a single line.

Legs in two dimensions: Both legs presented by more than a single line.

Hair: Any presentation or hat or cap covering head and hiding hair.

Neck: Definite separation of head and body necessary.

Arms pointing downward: One or both arms pointing down at an angle of 30° or more from horizontal position or arms raised appropriately for activity figure is engaged in: arms extending horizontally from body and then turning down some distance from the body is not scored.

Arms correctly attached at shoulders: Indication of shoulder necessary for this item, arms must be firmly connected to body.

Exceptional Items:

Knee: Distinct angle in one or both legs (sideview) or kneecap (frontview); round curve in leg not scored. [A bend in leg with rounded outline scored as knee if hypothetical midlines of leg parts bend by more than 15°. Designs on pants are not scored as knees.]

Profile: Head drawn in profile even if rest of figure is not entirely in profile.

Elbow: Distinct angle in arm required; rounded curve in arm is not scored. [As with knee, if the hypothetical midlines of the upper and lower parts of the arm bend by more than 15° the arm is scored for elbow.]

Two Lips: Two lips outlined and separated by line from each other; two rows of teeth only are not scored. [Open mouth by itself

is not scored as two lips. For frontal view, a line across the mouth is needed in addition to the top and bottom lines.

For profile, protruding edges are scored as two lips.]

Nostrils: Dots or nostrils show in addition to presentation of nose.

[Dots separated from outline of nose are scored as nostrils.]

APPENDIX B

APPENDIX B

A list of the 30 EIs and their description as provided by Koppitz (1968) is given below. All items are scored by boys and girls ages 5 to 12 except as indicated. All protocols are initially scored "0" and then "1" for each characteristic.

Quality Signs

1. Poor integration of parts (Boys, 7, Girls, 6): One or more parts not joined to rest of figure, part only connected by a single line, or barely touching.
2. Shading of face: Deliberate shading of whole face or part of it, including "freckles," "measles," etc.; an even, light shading of face and hands to represent skin color is not scored.
3. Shading of body and/or limbs: (Boys, 9, Girls, 8): Shading of body and/or limbs.
4. Shading of hands and/or neck (Boys, 8, Girls, 7).
5. Gross asymmetry of limbs: One arm or leg differs markedly in shape from the other arm or leg. This item is not scored if arms or legs are similar in shape but just a bit uneven in size.
6. Slanting figures: Vertical axis of figure tilted by 15° or more from the perpendicular.
7. Tiny figure: Figures two inches or less in height.
8. Big Figure: (Boys and Girls, 8): Figure nine inches or more in height.

9. Transparencies: Transparencies involving major portions of body or limbs; single line or lines if arms crossing body not scored.

Special Features

10. Tiny head: Height of head less than one-tenth to total figure.
11. Crossed eyes: Both eyes turned in or turned out; sideway glance of eyes not scored.
12. Teeth: Any representation of one or more teeth.
13. Short arms: Short stubs for arms, arms not long enough to reach waistline.
14. Long arms: Arms excessively long, arms long enough to reach below knee or where knee should be.
15. Arms clinging to body: No space between body and arms.
16. Big hands: Hands as big or bigger than face of figure.
17. Hands cut off: Arms with neither hands nor fingers; hands hidden behind back of figure or in pocket not scored.
18. Legs pressed together: Both legs touch with no space between, in profile drawings only one leg is shown.
19. Genitals: Realistic or unmistakably symbolic representation of genitals.
20. Monster or grotesque figure. Figure representing nonhuman, degraded or ridiculous person; the grotesqueness of figure must be deliberate on part of the child and not the result of his immaturity or lack of drawing skill.
21. Three or more figures spontaneously drawn: Several figures shown who are not interrelated or engaged in meaningful activity;

repeated drawing of a boy and a girl or the child's family is not scored.

22. Clouds: Any representation of clouds, rain, snow, or flying birds.

Omissions

23. No eyes: Complete absence of eyes; closed eyes or vacant circles.
for eyes not scored.
24. No nose: (Boys 6, Girls 5).
25. No mouth.
26. No body.
27. No arms: (Boys 6, Girls 5).
28. No legs.
29. No feet: (Boys 9, Girls 7).
30. No neck: (Boys 10, Girls 9).

APPENDIX C

APPENDIX C

The Modified Hiler-Nesvig (1965) scoring system is described in the following. The description of each item is that used by Stricker (1967). Each drawing is scored "5" initially; then "-1" is scored for each "patient" item and "+1" for each "normal" characteristic. Additional rules are in brackets.

Patient Items:

1. Bizarreness: This category includes such impressions as "schizy," "grotesque," "inhuman," "sinister," "sick," "ghoulish," "weird," and "gnomelike," but not simply "peculiar" or "distorted."
2. Omission of major parts of the body: The omission of major parts of the body, such as head, body, arms, legs, hands, feet, eyes, nose, mouth, and hair, and particularly of arms, hands, and torso, was characteristic of "patients" more often than of "normals."
3. Distortions: This category was particularly effective if distortion of head or arms was present. [Head distortion: The height of head is either less than one-tenth or greater than one-third of the total figure; head is clearly distorted in shape. Arm distortion: short stubs for arms, arms not long enough to reach waistline; arms excessively long, arms long enough to reach below knee or where knee should be; one arm differs markedly in shape from the other arm; arms are clearly distorted in shape. In judging the length of the arm the outside line

was measured as its length. Also, if the figure had a "belt," then the center of the belt was used as the waistline.]

4. Transparencies: This category referred particularly to transparency of the body or legs through the clothing. [Transparency of arms through clothing also scored.]

Normal Items:

1. Happy, pleasant facial expression. [The overall total facial expression was evaluated. Specifically, vacant circles for eyes were judged as unhappy. Also, a turned-up mouth was generally considered as happy, while a turned-down mouth was usually thought to be neutral. Thus, a face with vacant circles for eyes and a turned-up mouth would be considered to be neutral and not a "happy, pleasant facial expression."]
2. Nothing pathological: The subjective impression that there was nothing pathological in a drawing. [Modified to "certainly nothing pathological."]

APPROVAL SHEET

The thesis submitted by Roger W. Semyck has been read and approved by the following committee:


Dr. James Johnson, Director
Associate Professor, Psychology, Loyola

Dr. Emil Posavac
Associate Professor, Psychology, Loyola

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the Committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

1-5-77
Date


Director's Signature