



1982

Assessing Psychopathology in Children's Bender Gestalt Performance: A Comparison of Two Objective Systems

Edward D. Rossini
Loyola University Chicago

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



Part of the [Psychology Commons](#)

Recommended Citation

Rossini, Edward D., "Assessing Psychopathology in Children's Bender Gestalt Performance: A Comparison of Two Objective Systems" (1982). *Master's Theses*. 3303.

https://ecommons.luc.edu/luc_theses/3303

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 © 1982 Edward D. Rossini

ASSESSING PSYCHOPATHOLOGY IN CHILDREN'S BENDER
GESTALT PERFORMANCE: A COMPARISON OF TWO OBJECTIVE SYSTEMS

by

Edward D. Rossini

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

October

1982

ACKNOWLEDGMENTS

I sincerely wish to thank Dr. J. Clifford Kaspar for his direction and counsel throughout this project, for the generous use of his clinical files and for introducing me to the interesting field of clinical child assessment. I also thank Dr. Patricia A. Rupert for her support and excellent suggestions. Their patience and participation on this committee will never be forgotten.

Generous support was offered without limits by the following colleagues and friends: Janice Kowalski and Linda Vest Klein. I again thank each of these fine psychologists. Martin McDonald provided the generous use of his students. I thank him and his staff for their interest and cooperation.

VITA

The author, Edward D. Rossini, is the son of August J. Rossini and the late Tina (Herman) Rossini. He was born on March 29, 1954, in Chicago, Illinois.

His elementary education was obtained in the Chicago parochial school system (St. Sylvester School), and secondary education at Lane Technical High School, Chicago, Illinois, where he graduated in 1968.

In September, 1972, he entered Loyola University of Chicago, and in February, 1976, received a Bachelor of Science, Magna Cum Laude, with a major in psychology and a minor in theology. While attending Loyola University of Chicago, he was an officer in Psi Chi.

In September, 1976, he entered the doctoral program in clinical psychology at Loyola University of Chicago. His clinical training was at Lakeside Veteran's Hospital (clerkship, 1978), and the Doyle Center of Loyola University (internship, 1979-1981). In October, 1982, he will begin an internship in adult clinical psychology at Hines Veteran's Hospital, Hines, Illinois.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
VITA	iii
LIST OF TABLES	iv
CONTENTS OF APPENDICES	v
INTRODUCTION	1
REVIEW OF RELATED LITERATURE	5
The Hutt & Briskin (1960) System	14
Psychopathology Scale	17
Koppitz Developmental Scoring System and Emotional Indices . .	21
Statement of Problem and Hypotheses	31
METHOD	34
Subjects	34
Materials	36
Procedure	38
RESULTS	40
DISCUSSION	62
SUMMARY	73
REFERENCES	75
APPENDIX A	79
APPENDIX B	88
APPENDIX C	91

LIST OF TABLES

Table	Page
1. Demographic Characteristics of the Sample	37
2. Analysis of Variance on the Intelligence Quotients of the Two Clinical Groups	41
3. Means and Standard Deviations of the Intelligence Quotient Data	42
4. Analysis of Variance on the Koppitz Developmental Score	44
5. Descriptive Data on the Koppitz Developmental System . .	46
6. Analysis of Variance on the Koppitz Emotional Indices Sum Score	48
7. Means and Standard Deviations of the Sum of Emotional Indices	49
8. Analysis of Variance on the Psychopathology Scale Score	51
9. Psychopathology Scale Means and Standard Deviations . . .	52
10. Pearson Product-Moment Correlation Matrices	55
11. Pearson Product-Moment Correlation Matrices	56
12. Analysis of Covariance on the Emotional Indices Sum With the Koppitz Developmental Score a Covariate	58
13. Analysis of Covariance on the Psychopathology Scale Score With the Koppitz Developmental Score a Covariate	59

CONTENTS OF APPENDICES

	Page
APPENDIX A Psychopathology Scale Scoring Criteria	80
APPENDIX B Emotional Indices Scoring Criteria	89
APPENDIX C Parental Consent Form	92

CHAPTER I

INTRODUCTION

Psychological assessment has long been accepted as a valuable procedure for understanding troubled children. Recently there has been increased attention given to training psychologists to understand the special needs of children, especially as it related to the unique difficulties of childhood personality assessment (Cass, 1978; Gries, 1978). Nathan (1979) concluded that the majority of children seen for diagnostic evaluation appeared to be assessed by professionals who have not adequately demonstrated their competence to deliver valid clinical services to children. This is particularly troublesome since the two most frequently administered psychological tests are those developed for use with children, the Weschler Intelligence Scale for Children-Revised and the Bender Visual Motor Gestalt Test.

Divergent trends seem to characterize the field of clinical child assessment. Psychodiagnostic assessment has come under intense scrutiny and criticism in recent years. Criticisms have focused primarily on the inferred validity of the tests themselves; that is, on the appropriateness of inferences drawn from test scores. Many psychological tests are being administered to populations of children well beyond the standardization of the test, rendering the interpretation of the results invalid. Criticisms of psychological assessment

also include economic considerations. In many situations, projective assessment appears to be a cost-inefficient clinical service. Test data do not appear to yield sufficient information to justify their professional expense and time. Despite these criticisms, the number of psychological tests administered and marketed for professional use has risen significantly throughout the last decade (Brown & McGuire, 1976).

Among clinical child psychologists, the proper use of psychological assessment remains a critical problem. There appears to be a considerable need for valid diagnostic services that is currently being unmet. The President's Commission on Mental Health (1978) cited diagnostic services to children to be among the most understaffed service areas in mental health delivery services at the community level. Its recommendations called for more research into the validation of instruments to better identify and direct troubled children into remediation programs.

Blau (1979) highlighted a discrepancy in the effective use of psychological tests. He suggested that children were routinely overtested in educational settings, where their impact on curriculum and academic remediation was minimal, and seriously undertested in clinical settings where individual therapeutic programs are the hallmark of the profession. He proposed a research strategy that would obviate one aspect of the critique of psychological test validity. Blau mandated concurrent "situation specific re-validation" of major psychological tests to increase their external validity and to apply the

power of psychometric precision to specific sub-populations of disturbed children. By increasing the efficacy of a test, the service becomes more cost-efficient. Blau's suggestion offered timely and comprehensive remedies to the pervasive criticism of psychological assessment.

This study was partially anticipated by Blau's challenge for focused research on the validity of specific tests to increase their diagnostic efficacy. This study addresses the validity of the Bender Visual Motor Gestalt Test as a measure of personality assessment among school-aged children. The Bender Gestalt Test is among the most important, most frequently administered and versatile clinical tests available. It provides a very brief, and therefore inexpensive, non-verbal perceptual-motor test ideally suited for children.

The goal of this study is to assess the validity of the Bender Gestalt Test as a personality test in the identification of psychologically disturbed children. Specifically, this project examined and compared the validity of two systems for assessing psychopathology manifested on children's Bender Gestalt records. Koppitz' (1975) series of Emotional Indices, discrete Bender Gestalt distortions suggesting psychopathology, were compared to the objective Psychopathology Scale developed by Hutt (1977). This study compared the Emotional Indices sum and the Psychopathology Scale score of children of normal intelligence at two age levels representing three diagnostic groups. These children differ only in terms of their degree of psychological impairment, from normal to chronically disturbed, in three well

established diagnostic groups. This study was designed to offer an important contribution to the literature on children's Bender Gestalt performance, and was among the first to adhere to Achenbach's (1976) criteria of methodological rigor in the selection and definition of clinical subjects.

CHAPTER II

REVIEW OF THE RELATED LITERATURE

The Bender Visual Motor Gestalt Test (Bender, 1938) was developed to assess the maturational level of the hypothesized "visual motor gestalt function" in normal school aged children. Its purpose was to introduce a novel clinical instrument deduced from a specific theoretical model. The "gestalt function" was a seminal construct of the European school of Gestalt psychology. A "gestalt function" was defined as an active biologically based perceptual process which encodes a given constellation of stimuli as a phenomenological whole. The response to the stimuli was similarly defined as an integrated whole or "gestalten," gestalt being its English transliteration.

The Gestalt school of psychology sought to discover universal processes which governed cognition and perception. While the theoretical influence of Gestalt psychology has waned, studies on visual perception, visual memory and the verbal report of sensory phenomena remain the major contributions of this school to contemporary experimental psychology. Within clinical psychology, the Visual Motor Gestalt Test, introduced in a monograph by Laretta Bender (1938), continues to be a salient contribution of Gestalt psychology to contemporary clinical assessment procedures. The test consisted of nine figures, mostly familiar geometric designs, which were presented one

at a time. The usual procedure required the person to copy these figures, with no time limit, as accurately as possible on one or more sheets of blank paper. In the reproduction of visually presented stimulus cards, it was assumed that the geometric properties of each figure would be completed or reorganized by the person into a whole gestalt more complex than the simple stimulus properties of the design.

Bender's (1938) monograph documented the maturational development of certain aspects of the "gestalt function," among them the ability to reproduce contiguous figures. Her normative system yielded approximate age equivalents for the successful reproduction of each figure. An overall developmental level was not determined, nor was there a useful summary score derived from the person's performance on the Bender Visual Motor Gestalt Test. From 1938 until 1946 when the standard figures were first published for general use, most of the literature on the Bender Gestalt Test was descriptive, emphasizing a developmental approach of interpretation. Little or no attention was given to the psychometric properties of the test such as reliability and validity. Tolor and Schulberg (1963) criticized Bender's evaluative system as being "primarily subjective and based upon global impressions rather than upon careful analysis of specific factors which determine the accuracy of the reproduction" (p. 30).

As a clinical manual, Bender's monograph offered extensive case material drawn from the standardization sample of eight hundred children. Bender tentatively suggested the application of the Visual

Motor Gestalt Test to psychiatric patients, especially those with normal visual-motor functioning. Her position with regard to the test's possible value in assessing personality disturbances was, initially, ambiguous. She offered little hope that the test procedure would be useful in the differential diagnosis of neurotic patients yet, almost contradictory, she maintained that "the final product is a visual motor pattern which reveals modifications in the original pattern by the integrating mechanism of the individual who has experienced it" (p. 12). In the same paragraph she noted that ". . .it would not be surprising to find that some such Gestalten might become the symbol of the individual's unsatisfied infantile drives" (p. 12). She acknowledged the idiographic nature of the task, and hinted at potential psychoanalytic interferences with the task demands, yet her general position seemed to indicate little faith that her test could be used effectively in the identification or classification of personality disturbances.

Bender maintained this essentially negative evaluation of the clinical use of the Visual Motor Gestalt Test throughout her long career. In a retrospective evaluation of the research spanning twenty-five years, Bender concluded that "The application of this test to various categories of personality deviations is a futile task until such time as these categories have some definition" (Tolor & Schulberg, 1963, p. xvi). She expressed disbelief that such accurate categorization of psychological disorders would ever be possible.

The systematic use of the Bender Visual Motor Gestalt Test for

personality assessment resulted from the extensive research of Max Hutt in the 1940s with military and veteran populations. Hutt (1950) proposed a psychodynamic or projective method for evaluating a Bender Gestalt Test protocol based on a series of objectively scored criteria, ostensibly independent of maturational criteria. Using psychoanalytic theory and psychoanalytic symbolism as referents, Hutt hypothesized the relationship between specific distortions on the Bender Gestalt Test and specific clinical syndromes and diagnoses in an adult population.

During the period from 1938 to 1955, the two theoretical positions of Bender and Hutt challenged each other, with Bender maintaining a developmental and richly deductive strategy for evaluating Bender Gestalt protocols and Hutt offering inductive and objective data suggesting the clinical efficacy of Bender Test distortions in clinical diagnosis. Empirical substantiation of Bender Gestalt Test interpretations continued to be overlooked, as did the psychometric qualities of the test.

Although it was standardized as a test for school aged children, the use of the Bender Gestalt Test with children was markedly delayed. This was evident in the dearth of published research following the publication of Bender's monograph. In a comprehensive review of the literature on the use of the Bender Gestalt Test with children, Koppitz (1963) cited only twenty-seven studies of children's Bender responses in the three decades following the introduction of Bender's normative data. Of these few studies, only a small proportion

considered the use of the test as a personality assessment measure. Tolor and Schulberg (1963) commented on the paradoxical nature of evaluating children's Bender performances; individual clinicians undoubtedly made psychodynamic interpretations from children's Bender drawings without evidence for their validity and in the absence of a reliable scoring system. The latter reason made nomothetic comparisons impossible.

The first published projective application of the Bender Gestalt Test with school aged children was attributed to Greenbaum (1955). He simply asked children to free associate to each figure, saying what each figure reminded her/him of. These associations were added to a list of standard word associations. Without supporting data, Greenbaum reported that approximately two-thirds of the elicited responses contained significant psychodynamic material, anecdotally confirming Bender's (1938) speculation that the stimulus figures could arouse "infantile drives" in a person.

Several studies conducted in the late 1950s adduced cautious optimism regarding the clinical efficacy of the Bender Gestalt Test for the identification of psychologically disturbed children. Two early studies attempted to validate several of the objective factors proposed by Hutt (1950) in childhood samples by comparing the performance of a well adjusted sample with maladjusted children. Byrd (1956) compared the Bender Gestalt performance of typical children, labelled well adjusted, with the performance of peers being seen in outpatient psychotherapy. The experientnal and control groups

consisted of 200 children each, equally divided into four age groups ranging from eight to fifteen years. The contrasting subgroups consisted of 50 "well-adjusted" and 50 "maladjusted" children each. Byrd found that his data ". . . support the validity that the majority of test factors selected from Hutt are signs of personality adjustment" (p. 132). Among children ages seven to ten years, nine of the fifteen factors significantly differentiated between groups, each in the hypothesized direction. The "maladjusted" sample scored significantly higher on such factors as: chaotic sequence, rotation, changes in curvature, closure difficulty, change in size, overlapping difficulty, and changes in angulation. The "well-adjusted" group scored higher on two factors: orderly sequence and normal placement of figure A. Definition and scoring criteria are listed in Appendix A.

The Byrd (1956) study was significant as the first investigation of the potential utility of the test factors proposed by Hutt (1950) as indicative of childhood psychopathology. The findings are noteworthy since the proposed test factors were defined and scored from adult records and their applicability to children, especially in the younger age range, had not been proposed.

Simpson (1958) hypothesized that Bender Gestalt distortions in children result from affective causes rather than from immaturity in perceptual or motoric functioning. Twenty-five "well-adjusted" and a similar number of "emotionally disturbed" first graders were administered three tasks: a visual discrimination test, a test of fine

motor copying skill and the Bender Gestalt Test. The scores of the two groups did not differ on either the copying or the discrimination tests. However, the Bender Gestalt performance of the "emotionally disturbed" group was significantly poorer than the control group. Simpson hypothesized an affective impairment of "visual motor integration" among disturbed children, antedating Koppitz' (1963) hypothesis that "visual-motor integration" was the fundamental process assessed on the Bender Gestalt Test, and the most common source of Bender distortions.

Clawson (1959) extended Byrd's (1956) investigation to encompass a more definitive study of the projective uses of the test for childhood personality assessment. Whereas Byrd (1956) sought to identify which of Hutt's proposed factors would differentiate between normal and disturbed children, a pathognomonic sign approach, Clawson attempted to validate specific psychodynamic relationships between specific Hutt factors and projective test data. Her study was limited to children ranging in age from 7 to 12 years. At each age level 40 boys and 40 girls composed the control and experimental groups. Mal-adjusted children were clients of the Wichita Guidance Center with established psychological clinical diagnoses. The control group, matched individually on age, sex and I.Q., and matched on a group basis for socioeconomic status, consisted of children judged by their teachers to be normally developing and typical for their age group. Clawson utilized three types of statistical analyses to validate aspects of Hutt's test factors: differences in frequencies of

pathologic factors between the school and the clinic groups; chi square tests between test factors and behavioral symptoms in the clinic cases; and chi square test comparisons between selected Rorschach variables and specific, analogous test factors. Each hypothesis was partially confirmed. Of 48 chi square comparisons done overall, 19 were significant at the .01 level. More important, four of Hutt's factors were significantly related to specific behavioral problems: expansive organization on the Bender with acting out (in behavior), compressed organization with withdrawn behavior, decreased figure size with withdrawn behavior and horizontal placement with behavioral acting out. Four specific test distortions significantly differentiated the groups in the hypothesized direction: regression, closure, joining difficulty and erasures. Each was more frequent in the clinical group at the .001 level. While Byrd (1956) tested the "sign" approach method of interpretation, Clawson's investigation explored the full projective potential of the Hutt factors in clinical diagnosis with generally favorable results.

These three studies (Byrd, 1956; Clawson, 1959; Simpson, 1958) offered initial support for the extension of Hutt's test factors of assessing psychopathology to children's records. However, each author cautioned against making individual clinical decisions on the basis of nomothetic mean differences. It was concluded that while there appeared to be several Bender Gestalt Test distortions associated with adult psychopathology that occurred more frequently in the records of maladjusted children than in those of their normal peers, other signs

did not differentiate between groups. No single distortion appeared exclusively in the clinical samples; none could be considered valid pathognomonic signs. Similarly, psychodynamic interpretations of specific signs or patterns remained tentative at best. Tolor and Schulberg (1963) reviewed these studies and reported total dissatisfaction with the clinical significance of these studies. They recommended a return to research in the Bender tradition; exploration of the effects of age and maturation in the reproduction of Bender Gestalt designs, especially among children heterogenously labeled "emotionally disturbed."

The advent of the 1960s marked a turning point for the clinical interpretation of children's Bender Gestalt Test performance. Two major systems were proposed attempting to integrate disparate research findings regarding the Bender Gestalt's valid place as an instrument of personality assessment with children. Koppitz (1963, 1975) developed a series of projective Emotional Indices for the global identification of maladjusted children between five and ten years of age. These were considered independent of age of the child and her/his visual motor maturation. In addition, Hutt and Briskin (1960) proposed a system of objectively defined and scored test factors which quantified the type and severity of Bender Test distortions. This system was a revision of their previously cited tentative factors (Hutt, 1950) and was later revised into an objective Psychopathology Scale (Hutt, 1969, 1977) for the interpretation of adult and adolescent Bender Gestalt performance. Each system will be traced

historically and research will be presented focusing on investigations relevant to school aged children.

The Hutt & Briskin (1960) System

As previously stated, research studies dealing with the projective potential of the Bender Gestalt Test had appeared in the literature since 1950. A descriptive, inductive approach characterized many of these studies. Many authors proposed "signs" based on their clinical experience that seemed to appear more frequently in the protocols of psychiatric patients than in normal adult's records. Psychoanalytic inferences guided a search for specific test distortions that reflected characterological disturbances; for example, certain figures were given phallic or oral meanings and distortions on that figure were considered suggestive of conflicts in that area of psychosexual functioning. In one sense, Bender Gestalt distortions became nonverbal parapraxes; seemingly trivial errors of test performance with potent psychodynamic significance. This speculative method of interpretation was ideally suited to individual case studies but not subject to empirical validation techniques. Reviews of the literature on adult Bender Gestalt performance repeatedly cited the idiographic and subjective aspects of this method of evaluation (Hutt, 1960; Tolor & Schulberg, 1963). What was needed was a reliable scoring system to test these clinical hypotheses.

The culmination of attempts to develop a reliable scoring method for evaluating Bender Gestalt productions projectively appeared in the book of Hutt and Briskin (1960). This work was essentially a scoring

manual and an extended speculative work deduced from psychodynamic principles. No normative or nomothetic comparative data was presented to support their system of "inferential analysis." Their method relied simultaneously on the objective scoring of 25 factors and the covariation of these factors within a psychodynamic model. Psychodynamic interpretations were proposed for each factor and its types of distortions similar to Rorschach determinants and content categories.

The historical significance of this book continues to be the first wide-spread use of the 25 proposed factors and their scoring criteria. Hutt's 25 factors on the Bender Gestalt Test fell into five major categories: Size, Movement, Gross Distortion, Changes in Form, and Organization. A definitive list of Bender Gestalt distortions was finally available for validation.

In a critique of the Hutt and Briskin system, Tolor and Schulberg (1963) seriously questioned the validity of the "inferential analysis" of Bender Gestalt Test records. They were chagrined by the seemingly contradictory rationale in the development of this system. They noted that while specific test factors were classified, rigidly defined and given general circulation for research purposes, there appeared to be a "total abandonment of the need to validate their (psychodynamic) interpretation of specific test factors" (p. 46). In the development of the "inferential analysis" of Bender Test data there was an absence of psychometric validity. Hutt and Briskin (1960) proposed a system ostensibly based on their extensive clinical experience alone.

Several studies have used the Hutt and Briskin scoring categories attempting to extend their validity to childhood populations in the identification of maladjusted children. Mogin (1966) compared the frequencies of five, seemingly arbitrarily selected, factors with a behavior rating among normal second and third grade children. Significant positive correlations were reported between the total Bender Gestalt errors and high behavior ratings (indicative of maladjustment) for both sexes and both grades. This study of 211 children offered support for at least five of the factors: Placement, Sequence, Closure Difficulty, Overlapping Difficulty and Second Attempt, as a quick group screening of young children for maladjustment.

Flint (1967) investigated the convergent validity of several of Hutt and Briskin's (1960) factors with traits analogously named, but assessed on a standard personality inventory (Children's Personality Questionnaire). In a sample of normal ten to twelve year old children, the symbolic interpretation of seven factors was either wholly or partially confirmed with the objective subscales of the Children's Personality Questionnaire. Brannigan, Barone and Margolis (1978) investigated the relationship between impulsivity-reflectivity as measured by the Matching Figure Test and Hutt's hypothesized Bender Gestalt indices of impulsivity, specific distortions within six factors: Collision, Curvature Difficulty, Change in Angulation, Regression and Simplification and Elaboration. Subjects were 60 normal children between ten and eleven years of age. Children rated impulsive as measured by the Matching Figures Test produced significantly

more errors than reflective children on the Bender Gestalt Test. A majority of the specific hypotheses could not differentiate between reflective and impulsive children.

Finally, Trahan and Stricklin (1979) found no relationship between 15 of Hutt's Bender Gestalt signs of aggressive behavior and teacher rated aggressive acting out among 93 normal children with age ranging from five to eleven years. Neither the total number of scored indicators, nor the presence of any specific indicator, correlated significantly with scores on the Behavior Rating Scale. This negative finding held for males and females in each age group: five to eight and eight to twelve.

In sum, the limited research to date offers equivocal support for the projective analysis of Hutt's factors; each of the four cited studies demonstrated the need for more focused research with an emphasis on more stringent criteria for the selection of clinical populations.

Psychopathology Scale

In an attempt to quantify the amount of psychological disturbance manifested in Bender Gestalt performance of adults, Hutt (1969) introduced a 19-factor preliminary scale of psychopathology. Each of the previous Hutt and Briskin (1960) factors was assigned a linear scaled score "weighted" for increasingly distorted reproductions of the stimulus designs. Psychopathology was hypothesized to vary from a relatively small amount (normal) to an extreme amount (psychotic).

Scoring on the Scale was similarly incremental and linear; high scores indicated greater, more severe, psychopathology.

The Psychopathology Scale, and its validation, was based on empirical nomothetic comparisons between groups of individuals presumed to differ on a priori levels of assessed psychopathology. Normative data were presented in Hutt (1969). Three levels of psychopathology were compared with unequal numbers of subjects in each group: normal college students (N=80), outpatient adult psychoneurotics (N=80) and hospitalized, heterogeneously diagnosed schizophrenic patients (N=95). Normative groups were marginally matched for intelligence and equivalent ages. Mean age of the standardization sample was approximately 25 years old. Analyses of group mean differences indicated that the three groups were significantly different from each other in the hypothesized direction: Normals ($\underline{M} = 33.8$, $\underline{SD} = 3.2$), Neurotics ($\underline{M} = 57.8$, $\underline{SD} = 5.6$) and Schizophrenics ($\underline{M} = 100.3$, $\underline{SD} = 10.1$). No significant relationships were reported between scores on the Psychopathology Scale and sex, highest level of educational attainment, or intelligence as measured by the Otis Self-Administered Test.

Test-retest and inter-judge reliabilities were exceptionally high in the normative sample. Over two weeks, the retest reliability was reported to be significantly high ($\underline{r} = .85$) in a mixed clinical sample. Inter-scorer reliability was reported to be significantly high between moderately experienced scorers ($\underline{r} = .96$).

The Psychopathology Scale was revised in 1977, with two marginal factors eliminated and revised normative data presented. Hutt (1977) presented data indicating that the revised Psychopathology Scale significantly differentiated among six levels of psychopathology: Normals, Outpatient Neurotics, Inpatient Neurotics, Unipolar Depressives, Outpatient Schizophrenics and Chronic Schizophrenics. Mean group differences were significant at the .001 level or better with one exception; i.e., the mean difference between Inpatient Neurotics and Unipolar Depressives was significant at the .01 level of probability. Correlational analyses revealed no significant relationship between the Psychopathology Scale score and sex, chronological age and intelligence in non-retarded adolescents and adults.

The current Psychopathology Scale consists of seventeen factors each scored on a scale from one to ten and one factor scored between one and 3.25. Scores on the entire scale may range from a minimum of 17.0 to a maximum of 163.25, the total range being 146.25 points. Scoring of the Revised Psychopathology Scale was greatly facilitated by the use of a Scoring Template and standardized Scoring Sheets (Appendix A).

Recent studies have extended the use of the Psychopathology Scale to adolescent male populations. Hutt and Dates (1977) examined the differential effectiveness of individual versus group psychotherapy for male delinquents and their relative effect on psychopathology assessed on the Psychopathology Scale. Group mean differences in scores were significantly lower for both treatment modalities

following treatment but unchanged among control peers. Each group of delinquents continued to score significantly higher on psychopathology than normal peers and "emotionally disturbed" peers following the completion of their treatment contract. Hutt, Dates and Reid (1977) reported significant Pearson correlations between scores on the scale and other indices of delinquency. The Psychopathology Scale had significant predictive ability for rates of recidivism defined as the number of repeat offenses in two years. The authors cautioned against using the Psychopathology Scale score to diagnose an individual even though the data indicate the value of the scale in differentiating among groups given the significant but moderate correlations among measures.

Hutt (1977) reported little normative data on children's performance on the Psychopathology Scale; none with children younger than ten years of age. Preliminary data with children between ten and twelve suggested that "disturbed" children performed as a group most similar to outpatient neurotic adults. These children had mean scores significantly higher than normal peers, normal adolescents and adults. Yet no discernable age or maturational trend appeared in the normative data. Hutt (1977) extrapolated that the Psychopathology Scale was roughly applicable for children older than six years of age given intelligence in the normal range and age appropriate visual-motor functioning. The validity of the Hutt Psychopathology Scale with children younger than ten years has yet to be demonstrated.

The Koppitz System: Developmental Scoring
System and Emotional Indices

While the Bender Gestalt Test had been well established as a clinical instrument with impaired adults, its wide-spread use with children began with the publication of the Developmental Bender Test Scoring System (Koppitz, 1963). Koppitz proposed that visual-motor integration, the theoretical construct assessed by the test, reached a mature level by the middle elementary school years. Through the early elementary school years, roughly five to ten, this ability undergoes rapid development in several related areas: fine motor dexterity, perceptual orientation and spatial ability. The Developmental Bender Scoring System offered the first objective scoring system standardized to assess normal maturational changes in children's performance.

The Developmental Bender Scoring System has well established psychometric qualities (Koppitz, 1963, 1975; Keogh, 1969; Buckley, 1977). The majority of research studies with children in the past fifteen years have used this system as the standard criteria of Bender Gestalt Test performance. To date, the Developmental Bender Scoring System remains the most effective system for evaluating the performance of children between five and ten years of age.

The Developmental score is the sum of thirty possible errors. Errors are objectively defined and can be grouped under four categories: distortion of shape, rotation, integration and perseveration.

Higher scores imply more immature (poorer) performances. Normative data are provided in six-month intervals from five years to ten years six months, from a sample of 975 children. The normative data show no significant sex differences in the Developmental score. Among children of normal intelligence, scores are relatively uninfluenced by sex, ethnocultural or socioeconomic variables.

Reliability coefficients in test-retest situations were reported to be consistently above .80, quite acceptable for this measure. Interscorer reliability ranged from .79 to .99 in the twenty-four published studies reviewed by Koppitz (1975).

Koppitz (1963) proposed two independent methods of interpreting Bender Gestalt Test data. Distortions in the reproduction of the stimulus designs were attributed primarily to immaturity in "visual-motor integration" rather than to specific developmental delays in visual perception or motoric coordination. Koppitz (1975) concluded from thousands of administrations of the Bender Gestalt Test to children that youngsters with poor Bender Gestalt performances do not have poor visual perception, nor do they demonstrate difficulties with fine motor coordination. Rather, they tend to have problems in perceptual-motor integration; a higher level integrative function analogous to Bender's (1938) "gestalt function" and foreshadowed by Simpson (1958). This construct was ostensibly assessed by the Developmental Scoring System.

In addition, Koppitz (1963, 1975) suggested that a child's mode

of working on the task, his/her planning and organizational strategies as well as critically assessing the performance, appeared to relate to personality characteristics and learned attitudes rather than to maturational skills. This aspect of Bender Gestalt Test performance warranted a clinical evaluation and led to the development of the Emotional Indices.

Children with "emotional disturbances" were hypothesized to reveal significantly more Bender Gestalt errors not attributable to visual-motor immaturity relative to their normal peers. It was suggested that disturbed children draw more to satisfy their own needs than to comply with the task demands to reproduce the presented design. Following this reasoning, Koppitz (1963) introduced a series of discrete emotional indices, or signs, thought to be independent of visual-motor integration ability, that would differentiate between groups of normal and maladjusted children. Her initial list of emotional indices included: Confused Order, Wavy Line, Dashes for Circles, Progressive Increases in Size, Large and Small Size, Fine Line, Overwork, Second Attempt, Expansion, and Constriction. The revised list (Koppitz, 1975) added two rare Emotional Indices: Boxed Drawings and Spontaneous Elaboration. These twelve signs were adapted from the Hutt (1960) system. Each of these Emotional Indices is defined and scoring criteria cited in Appendix B.

Normative data for the Emotional Indices were collected on two groups of children, ages five to ten years, matched for age and sex as well as selected from children with normal intelligence. Half of

the sample (N = 136) were children referred for psychological assessment because of "serious psychological problems." The control group consisted of children with no known history of emotional maladjustment. Chi square analyses on the incidence of each Emotional Indicator revealed that ten of the eleven proposed indices occurred more frequently among the disturbed children; six of the Indices were significantly more frequent in the protocols of the disturbed children: Wavy Line, Overwork, Confused Order, Small Size, Second Attempt. The Emotional Indices appeared not entirely independent of the age factor in school aged children. Several of the Indices were found to be significantly related to emotional problems in younger children between five and seven years old. These included: Wavy Line, Overwork and Expansion. Other Indices were significantly related to maladjustment in the older children, ages eight to ten, notably: Confused Order, Small Size, Second Attempt and Expansion. Only Expansion was diagnostically significant at each age level. One Indicator, Constriction, was eliminated because of relatively equal frequency within each group.

The sum of the Emotional Indices present in a protocol appeared to be related to the severity of the emotional disturbance. False positive assignments, normal children categorized as emotionally disturbed on the basis of their Bender Gestalt performance, accounted for 35% of subjects with three or more Emotional Indices, 20% at a four Indicator level and no normal child's record had five or more Emotional Indices. The results of the normative study revealed that

subjects with emotional problems scored from 0 to 6 on the Emotional Indices while the control subjects did not reveal more than a maximum of 4 Emotional Indices. The mean number of Emotional Indices in the group of emotionally disturbed children was 2.32 with a standard deviation of 1.59. Control children averaged 1.27 Emotional Indices with a standard deviation of 1.04. The mean difference was significant, $t(270) = 6.41$, $p = .001$. For the eight to ten year olds, the mean number of Emotional Indices was 1.9 for the clinical group and 0.9 for the control children. Among the five to seven year old subjects, the clinical group scored 2.9 Indices while the control group score 1.7. Each age level difference was statistically significant. Koppitz (1963) concluded that the total incidence of Emotional Indices appeared to be related to the seriousness of the emotional disturbance and that the Emotional Indices could differentiate between children with psychiatric problems and normal peers. The normative data did not examine the differential diagnostic efficacy of the Emotional Indices; it did not demonstrate validity in the differentiation between neurotic and psychotic children, nor between other diagnostic categories.

Koppitz (1963) hypothesized that the two systems she described, the Developmental Bender Scoring System and the Emotional Indices, were independent and provided normative data to support this notion. However, a closer examination of the presented data suggested that a complex relationship between systems existed, one not explained sufficiently by Koppitz. Normative data revealed that children with

poor visual-motor perception had significantly higher incidences of emotional disturbance than children with good visual-motor perception. Koppitz therefore hypothesized that children with poor developmental Bender Gestalt performances would show a higher incidence of Emotional Indices than children with good visual-motor perception. Chi square analyses revealed a significant relationship between the quality of Bender Gestalt performance (good or poor) and the relative incidence of Emotional Indices (0 to 2 or 3+). The Chi-square statistic was significant at each age level and for the entire normative sample (N = 272). Children with two or fewer Emotional Indices were most likely to have good Developmental Bender records. However, the relationship among the pooled normative data, while statistically significant, revealed a complicated relationship not adequately explained by Koppitz with her cursory and ambiguously worded conclusion that ". . . it seems safe to assume that the Developmental score and the emotional indicators on the Bender Test measure different aspects of a child's functioning. . ." (p. 142).

A large proportion of children with poor Developmental Bender records also revealed 2 or fewer Emotional Indices. Few children (N = 19) demonstrated good Developmental Bender records with three or more Emotional Indices. In independent systems the relative cell frequencies in a 2 x 2 factorial table would be approximately equal. This does not occur in the normative sample. One powerful relationship (good Developmental Bender/two or less Emotional Indices) appeared to mask the interpretative significance of the other cells.

Koppitz (1963) conceded that poor scores on each system were found more often on the records of emotionally disturbed children than on those of their normal peers. This overlooked the relationships found in the full sample Chi-square table.

Reanalysis of the normative data for the present study suggested a strong relationship existed between scores on the Developmental Test and the frequency of Emotional Indices. The relationship was not as simple as suggested by Koppitz' original conclusions. Among the normative sample 61% of the children with 2 or less Emotional Indices had good Bender records, 39% had poor Developmental records. Yet Koppitz maintained that the systems were relatively independent. The normative data indicate a strong, yet confusing, relationship.

The hypothesized interaction between the Developmental scores and the incidence of Emotional Indices remains equivocal. Dibner and Korn (1966) reported a nonsignificant relationship between the two systems among 492 normal school children between five and one-half years old and ten and one-half years of age. In a clinical sample confounded by an age range from five to twenty-five years, McConnell (1967) concluded that emotionally disturbed subjects' Developmental scores were unrelated to the incidence of Emotional Indices. The normative data offered by Koppitz (1963) do not directly address the question of the amount shared variation that exists between the scores of the two systems. Koppitz (1975) later concluded "that the Emotional Indicators are not primarily related to age and maturation. . . even though many children with immature visual-motor perception also reveal

Emotional Indicators on their Bender Test records" (p. 81). She suggested that children experiencing impaired visual-motor functioning tended to be children high in vulnerability for developing emotional difficulties secondary to their neuropsychological deficits.

Following Koppitz' (1963) normative study on the validity of the Emotional Indices, little research support has been adduced to cross-validate her system. Naches (1967) found that the Bender Gestalt records of institutionalized acting-out children contained significantly more signs of psychopathology than normal school children matched for age, sex and intelligence. Five modified Emotional Indices differentiated between groups; two of these, Confused Order and Large Size, were derived from the Koppitz (1963) list of Emotional Indices. This study cross validated one Koppitz Indicator, Confused Order, in a sample of eight to eleven year old children. McConnell (1967) attempted to use the Emotional Indices to differentiate among four levels of psychopathology: Situational Reactions, Neurotic/Characterological Disorders, Borderline Psychotics and Psychotic children. There was no significant difference in the mean number of Emotional Indices between the Neurotic Group (N = 10) and the Situational Disorder Group (N = 10). Each of these groups had significantly fewer Indices than both the Borderline Psychotic Group (N = 10) and the Psychotic Group (N = 10), which were also significantly different. The external validity of this factorial study is limited; the age range of these subjects (five to twenty-five years) and their limited intellectual abilities limits the meaning of these

results and confounds their generalizability. The research attempting to cross validate the Emotional Indices has been equivocal. Methodological problems in the two reported studies render a definitive evaluation of the emotional indices difficult, if not impossible at this time.

Similar controlled studies in academic settings have also offered little support for the Koppitz system of Emotional Indices in the identification of children with behavioral problems. Dibner and Korn (1966) evaluated the predictive as well as concurrent validity of the Emotional Indices compared with teacher ratings of academic and psychological adjustment. Among children five to ten years of age (N = 492), the sum of Emotional Indices was not significantly related to any of the teacher assessed categories of behavioral or academic problems. However, the number of Emotional Indices produced in the kindergarten group was significantly correlated with problems in reading, arithmetic and overall academic performance assessed in first and second grades. This study was noteworthy independent of its results. It offered the first normative data reported by age, grade level and sex among normal children. In a group administration of the Bender Gestalt Test, the mean number of Emotional Indices reported was 2.08 (N = 150) for boys and 2.10 (N = 164) for girls, each between five and ten years old.

Gregory (1977) compared the Emotional Indices with the objective subscales of the Devereux Child Behavior Rating Scale in a sample of children with mean age of 8 years 7 months (N = 47) referred for

assessment because of heterogeneous school difficulties. The total number of Emotional Indices correlated significantly with 5 of 17 DCB factors: Psychological Use of Senses, Poor Coordination and Body Tonus, Unresponsiveness to Stimulation, Anxious-Fearful Ideation and Inability to Delay. However, in the comparison of objective factors with their hypothesized Emotional Indicator equivalent, none of the hypothesized correlations between specific Indicators and the logically associated DCB factors were significant. The behavioral correlates of the specific Indicators was not supported by this research.

Finally, in another academic application of the Emotional Indices, Ackerman, Peters and Dykman (1971) reported no significant differences in the frequency of Emotional Indices present in records of normal and learning disabled children. Neither the total number of Indices nor any single Indicator differentiated the groups of boys between 8 and 12 years of age. The authors challenged Koppitz' contention that emotional problems develop secondarily to the handicap of visual-motor impairment in many children.

In a pragmatic overview of the related literature Keogh (1969) suggested that a child's emotional disturbance is often not evident in drawing tasks similar to the Bender Gestalt Test or that the proposed objective signs of psychopathology are not sensitive enough. Her clinical experience indicated that emotionally disturbed children perform idiosyncratically on the Bender and "far worse on the Bender than their objective Koppitz scores would indicate" (p. 19). Her

recommendation as to the use of the Bender Test in differential diagnosis was similar to Bender's (1938): a projective evaluation of Bender performance based on extensive clinical experience guided by, but not limited to, nomothetic comparisons of signs and indicators.

In sum, although the research to date offers little support for the continued use of the Emotional Indices in the assessment of psychopathology among school aged children, it continues to be a popular clinical test. Similarly, the Psychopathology Scale has not been validated in children younger than ten years of age. What can be concluded from the literature is that atypical children, usually labeled "emotionally disturbed," produce Bender Gestalt protocols more deviant from normal children than from each other. These differences have been classified, organized into systems and reliably scored yet their diagnostic significance remains elusive, even chimerical.

Statement of Problem and Hypotheses

This review of the relevant literature reveals that most of the work that has been done on the projective evaluation of children's Bender Gestalt performance has been exploratory, episodic and clearly lacking in an inductive theme. Available research has tended to have two serious deficiencies. The first concerns the arbitrary and undefined meaning of "emotionally disturbed" in subject selection and identification. Problems of nosology abound in the reported literature. Secondly, many studies do not appreciate the importance of maturation in children's performance. Age specificity has not been

adhered to in the selection of experimental groups. This is particularly important because the standard criterion of Bender Gestalt Test performance is age normed for children exclusively between five and ten and one-half years of age.

It was the intention of the present project to contribute to the resolution of these methodological deficiencies by using a factorial design to investigate the validity of the Psychopathology Scale and the Emotional Indices with a normally intelligent sample of children in a factorial design. The Bender Gestalt performance of three groups of children will be compared. Groups were determined by a current psychiatric diagnosis, either Normal, Adjustment Reaction or Chronic Disorder. Each diagnostic group had two age levels: Younger subjects range from seven to eight years of age, Older subjects were either nine or ten years old.

This factorial paradigm offered the most refined method for examining the relevant variables and their interaction in the sample of children most likely to benefit from an evaluation using the Bender Visual Motor Gestalt Test.

- 1) Each of the three dependent variables, the Psychopathology Scale score, the Developmental Bender Scoring System score, and the sum of Emotional Indices will demonstrate significant inter-scoring reliability.

- 2a) It is hypothesized that the Adjustment Reaction Group and the Chronic Disorder Group will produce a significantly higher

Emotional Indices sum than the Normal Group.

- 2b) The frequency of Emotional Indices will not be significantly different in the Adjustment Disorder Group as opposed to the Chronic Disorder Group.
- 3) There will not be a significant main effect of Age of subject as related to the frequency of Emotional Indices; no significant interaction of Age and Level of Psychopathology is hypothesized.
- 4) The Psychopathology Scale score will significantly differentiate among Levels of Psychopathology, with the Normal Group demonstrating the lowest score.
- 5) The Age of subjects will be a non-significant main effect relative to the Psychopathology Scale score; no significant Age by Level of Psychopathology interaction is hypothesized.
- 6) Within the normal range of intelligence, there will be no significant correlation between the intelligence quotients of subjects in each clinical group with the Psychopathology Scale score or with the sum of Emotional Indices.
- 7) There will be a significant correlation between the sum of Emotional Indices and the Psychology Scale score across the range of psychopathology.

CHAPTER III

METHOD

Subjects

Data for the two clinical groups defined in the present study were archival. Clinical subject's protocols were selected from the files of the Charles I. Doyle, S.J. Center of the Loyola University of Chicago. For selection, each subject's file met the following criteria: (1) A complete Bender Gestalt Test record was available. (2) A psychiatric diagnosis was available. (3) There was no indication of neuropsychological impairment, psychosis or pervasive developmental delay. (4) Intelligence as defined by criteria in the Diagnostic and Statistical Manual-II was normal. The intellectual level of all clinical subjects was documented by an intelligence quotient derived from either the Weschler Intelligence Scale for Children or the Stanford Binet (Form M).

Subjects were assigned to one of three groups determined by their psychiatric diagnosis: Normal (Control) Group, Adjustment Reaction Group or Chronic Disorder Group. These groups were defined as follows:

The Adjustment Reaction Group was defined by either a Diagnostic and Statistical Manual I or II diagnosis of: Transient Situational Disorder (307) or Adjustment Reaction of Childhood (307.1). This

diagnosis was the modal diagnosis among children seen in outpatient psychotherapy (Koss, 1980) and among disturbed children in general (Eme, 1979). The determination of this diagnosis included virtually any symptoms or set of symptoms that appeared precipitated by acute environmental stress.

The Chronic Disorder Group was defined by the diagnosis of: Neurosis (300), Behavior Disorder of Childhood (308), or Personality Disorder (301-304). Pfeffer and Plutchik (1982) reported the high prevalence of these chronic disorders in latency age children. This diagnosis resulted from the persistence of symptoms which appeared internalized, less responsive to environmental situations and relatively resistant to psychotherapeutic modification.

The Normal (Control) Group consisted of Bender Gestalt Test records randomly selected from data previously collected from a Chicago parochial grammar school. For selection into the present study, each child was age and grade appropriate and attending a normal classroom. Each Control subject participated with the signed consent of a parent and the child's informed consent (Appendix C).

Each diagnostic group was divided into two age levels. Older subjects in each diagnostic group were either nine or ten years of age. Younger subjects were either seven or eight years of age. Each group contained forty subjects, twenty at each age level. The total sample size was one hundred and twenty subjects.

The sex ratio of each group reflected the prevalence of males to

females in the clinical population. The sample sex ratio of approximately 3:1 males to females was consistent with previous research in the area and established epidemiological patterns of childhood psychopathology (Eme, 1979). There was no attempt to control for racial, cultural or socioeconomic factors in the assignment of subjects. Research among normally intelligent children had repeatedly documented the non-significance of these demographic factors in Bender Gestalt Test performance (Byrd, 1957; Clawson, 1959; Zuelzer, Stedman & Adams, 1976). Table 1 lists the average and intellectual characteristics of the sample and the actual sex ratios.

Materials

The Bender Gestalt Test performance of each subject was measured on three objectively scored scales: The Developmental Bender Scoring System (Koppitz, 1975), the Emotional Indices (Koppitz, 1975) and the Psychopathology Scale (Hutt, 1977). Scoring definitions and criteria are listed in Appendix A for the Psychopathology Scale and Appendix B for the Emotional Indices.

The Developmental Bender Test Scoring System represents the total of thirty possible errors present in the record, although in the age range of the present subjects the number of errors is typically less than eight. These errors are grouped into four categories: rotation, perseveration, poor integration and distortion of shape. The higher the Developmental score, the poorer the child's performance indicating a relatively immature visual-motor ability.

TABLE 1
Demographic Characteristics of the Sample

<u>Group</u>	<u>Number</u>	<u>Male</u>	<u>Female</u>	<u>Age (Months)</u>		<u>Intelligence</u>	
				<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
<u>Adjustment</u>	(40)						
Younger	20	15	5	99.1	6.56	110.6	11.49
Older	20	16	4	121.2	7.71	102.4	10.87
<u>Chronic</u>	(40)						
Younger	20	11	9	93.5	7.79	98.3	11.49
Older	20	15	5	122.3	8.70	102.2	12.60
<u>Control</u>	(40)						
Younger	20	15	5	98.5	5.89	*	*
Older	20	15	5	120.0	4.81	*	*

The Emotional Indices Sum is the total number of mutually exclusive distortions present not considered related to visual-motor integration processes. The Emotional Indices can range from zero to eleven. Koppitz (1975) suggested that these distortions appear to reflect the personality characteristics of the subject rather than simple rendering skills.

The third measure, the Psychopathology Scale, was one objective scale from the Hutt adaptation of the Bender Gestalt Test designed to quantify the amount of psychopathology manifested in the copied figures. The score is the sum of the scaled scores of seventeen individual factors. The score can range from seventeen to 164.25. Although the Psychopathology Scale was developed on Bender Gestalt Test cards slightly dissimilar to those recommended by Bender (1946), actual differences between decks are minimal and all scoring systems are applicable to each set.

Procedure

Subjects in the Control Group were individually administered the Bender Gestalt Test with the following instructions: "Here are some figures for you to copy. Just copy them the way you see them." Any subsequent questions were answered with a non-committal "You can do it any way that you like."

Each subject was supplied with several sheets of paper and a sharp pencil. Following the test each child was asked her/his birth-date and grade. Subjects were invited to ask any questions that might

have occurred to them. Each child was told that she/he did very well and was thanked for their participation.

Protocols from the Adjustment Reaction Group and the Chronic Disorder Group were selected from the files of the Charles I. Doyle, S.J. Center of the Loyola University of Chicago. The clinical data represent roughly one subject cohort; subjects tested between 1978 and 1981. Data was gathered which met the selection criteria and which met the sex ratios for each group. Successive files were examined until the cells were completed.

Each protocol was scored by an experienced scorer. A second, less experienced scorer evaluated a random sample consisting of 32 protocols selected from within each of the four clinical conditions: Young Adjustment Group, Young Chronic Group, Old Adjustment Group and Old Chronic Group. Inter-scorer coefficients of correlation were determined for each of the three dependent measures.

CHAPTER IV

RESULTS

Subject Characteristics: Intelligence

To evaluate the homogeneity of intelligence between the Adjustment Reaction Group and the Chronic Disorder Group a 2(group) x 2(age) analysis of variance was performed. The results are presented in Table 2. There was no significant difference in intelligence between age levels. A significant difference was found across levels of psychopathology, $F(1,76) = 6.14, p < .01$. To understand this significant interaction the diagnostic group means were compared using t -tests. The results are presented in Table 3. Results indicated that among the older subjects, there was no significant difference between the intelligence quotients of the Chronic and Adjustment Reaction subjects. However, among the younger subjects, children in the Chronic Disorder Group demonstrated significantly lower intelligence quotients than the Adjustment Reaction subjects, $t(38) = 3.60, p = .001$.

The results indicate that the younger adjustment reaction subjects had significantly higher measured intelligence than either of the remaining groups; the children in the Younger Chronic Disorder Group, the Older Adjustment Reaction Group and the Older Chronic

TABLE 2
Analysis of Variance on the Intelligence Quotients
of the Two Clinical Groups

<u>Source of Variation</u>	<u>Mean Square</u>	<u>DF</u>	<u>F</u>	<u>Significance</u>
Age Level	90.31	1	.70	.40
Group	787.51	1	6.14	.01
Interaction	738.11	1	5.76	.01
Residual	128.09	76		

TABLE 3

Means and Standard Deviations of the Intelligence Quotient Data

	<u>Adjustment Group</u>	<u>Chronic Group</u>
Younger Subjects	110.6a (11.49)	98.3a (11.40)
Older Subjects	102.4 (10.80)	102.2 (12.60)

Note: Means with identical subscripts are significantly different across rows, a: $p = .01$.

Disorder Group did not differ significantly in intellectual functioning. The differences in intelligence noted above reflect variation within the normal range of intellectual functioning.

Inter-Scorer Reliability

The inter-scorer reliabilities of each scoring system used in this study were evaluated. Inter-scorer coefficients of correlation between an experienced and an inexperienced scorer were significant for the Koppitz Developmental System, $r = .96$, $p = .001$, the Koppitz Emotional Indices Sum score, $r = .92$, $p = .001$, and the Psychopathology Scale score, $r = .97$, $p = .001$. As hypothesized, each scoring system appeared to be adequately objective and reliably scored by persons with minimal training.

Koppitz Developmental Scoring System

To examine the differences among the three levels of psychopathology and the two age levels, a 3(group) x 2(age level) analysis of variance was performed on data from the Koppitz Developmental Scoring System. Results of the analysis are presented in Table 4. There was a significant main effect of age, $F(1,114) = 13.16$, $p < .001$, indicating that younger children scored higher than older children. This effect was expected since the scale is developmentally normed with older children expected to demonstrate lower, developmentally more advanced, scores than younger children.

Unexpectedly, a significant main effect was found for levels of psychopathology, $F(2,114) = 6.20$, $p < .003$, indicating differences in

TABLE 4

Analysis of Variance on the Koppitz Developmental Score

<u>Source of Variation</u>	<u>Mean Square</u>	<u>DF</u>	<u>F</u>	<u>Significance</u>
Age Level	70.53	1	13.16	.001-
Group	33.25	2	6.20	.003
Interaction	29.15	2	5.44	.006
Residual	5.36	114		

the Developmental score across the three levels of psychopathology. An analysis of the group means presented in Table 5 indicated that the Chronic Disorder Group scored significantly higher (poorer) than either the Control Group, $t(38) = 2.89$, $p < .05$, or children in the Adjustment Reaction Group, $t(38) = 2.17$, $p < .05$.

A significant interaction was found, $F(2,114) = 5.44$, $p < .006$, between the three levels of psychopathology and the two age levels. Analyses of means presented in Table 5 indicated that among older children, no significant difference was found among the three diagnostic groups. Among younger children, however, the Control Group scored significantly lower (better) than the Chronic Disorder Group, $t(38) = 3.39$, $p < .001$, and slightly, but not significantly lower than the Adjustment Reaction Group. Similarly, the Adjustment Reaction Group scored significantly lower than the Chronic Disorder children, $t(38) = 2.62$, $p < .01$.

In summary, two trends appeared in the analyses of Koppitz' Developmental Scoring System data that were unexpected. First, among data pooled across age levels, the children in the Chronic Disorder Group were scored exceptionally poorer than either the Control or Adjustment Reaction children. Closer examination of that relationship revealed that the significant difference was accounted for by the performance of the younger disturbed children. The performance of the younger Adjustment and Chronic children was poorer than all older children and the younger Control children. The Developmental

TABLE 5

Descriptive Data on the Koppitz Developmental System

	Control Group	Adjustment Group	Chronic Group
Younger Subjects	$\underline{M} = 3.05c$ $\underline{SD} = 1.35$	$\underline{M} = 3.95b$ $\underline{SD} = 2.25$	$\underline{M} = 6.45c$ $\underline{SD} = 3.62$
Older Subjects	$\underline{M} = 3.00$ $\underline{SD} = 2.00$	$\underline{M} = 2.80$ $\underline{SD} = 1.96$	$\underline{M} = 3.05$ $\underline{SD} = 2.06$
Pooled Data	$\underline{M} = 3.02c$ $\underline{SD} = 1.68$	$\underline{M} = 3.37a$ $\underline{SD} = 2.16$	$\underline{M} = 4.75ac$ $\underline{SD} = 3.38$

Note: Means with identical subscripts are significantly different
across rows.

$$^a_p = .05$$

$$^b_p = .01$$

$$^c_p = .001$$

Scoring System appeared responsive to psychopathology in younger children. However, among normal children at any age and in older disturbed children the Developmental scores appeared relatively age-appropriate in terms of visual-motor integration ability.

Koppitz Emotional Indices Sum Score

To examine the effects of age level and level of psychopathology on the Emotional Indices Sum score, a 3(group) x 2(age level) analysis of variance was performed. The results of the analysis are presented in Table 6. A significant main effect was found on the level of psychopathology, $F(2,114) = 13.43$, $p < .001$, which indicated that level of psychological disturbance was related to the incidence of Emotional Indices. Analysis of the means presented in Table 7 indicated that children in the Control Group demonstrated significantly fewer Emotional Indices than children in either the Adjustment Reaction Group, $t(38) = 3.05$, $p < .01$, or the Chronic Disorder Group, $t(38) = 4.46$, $p < .001$. The Adjustment Reaction Group and the Chronic Disorder Group did not differ significantly on their incidence of Emotional Indices, thus confirming hypothesis 2b which predicted no difference in the incidence of Emotional Indices between the two clinical groups of children. Similarly, hypothesis 2a, positing the lowest incidence of Emotional Indices in the Control Group, was supported.

There was no significant age level effect or interaction between age level and level of psychopathology. This wholly confirms hypothesis 3 which posited no relationship between the Indices and age level or interactive effects of age and diagnostic group.

TABLE 6

Analysis of Variance on the Koppitz Emotional Indices Sum

<u>Source of Variation</u>	<u>Mean Square</u>	<u>DF</u>	<u>F</u>	<u>Significance</u>
Age Level	0.40	1	0.32	.56
Group	13.44	2	10.81	.001-
Interaction	0.93	2	0.75	.47
Residual	1.24	114		

TABLE 7

Means and Standard Deviations of the Sum of Emotional Indices

	Control Group	Adjustment Group	Chronic Group
Younger Subjects	$\bar{M} = 1.20$ $\underline{SD} = 1.35$	$\bar{M} = 2.55$ $\underline{SD} = 1.00$	$\bar{M} = 2.55$ $\underline{SD} = 1.09$
Older Subjects	$\bar{M} = 1.65$ $\underline{SD} = 1.08$	$\bar{M} = 2.05$ $\underline{SD} = 1.05$	$\bar{M} = 2.60$ $\underline{SD} = 1.42$
Pooled Data	$\bar{M} = 1.42ab$ $\underline{SD} = 1.03$	$\bar{M} = 2.12a$ $\underline{SD} = 1.01$	$\bar{M} = 2.57b$ $\underline{SD} = 1.25$

Note: Pooled means with identical subscripts are significantly different.

$$a_p = .01$$

$$b_p = .001$$

The Koppitz Emotional Indices Sum effectively differentiated the Control children from each level of the disturbed children. Between the two disturbed groups of children, the Emotional Indices did not differentiate between the Adjustment Reaction Group and the Chronic Disorder Group although there was a trend in the means with the less disturbed children scoring fewer Indices than the more disturbed children in the Chronic Disorder Group.

Psychopathology Scale

To examine the hypothesized ability of the Psychopathology Scale score to differentiate among levels of psychopathology at all age levels, a 3(group) x 2(age level) analysis of variance was performed. The summary table for this analysis is presented in Table 8. A significant effect for level of psychopathology was revealed, $F(2,114) = 5.25$, $p < .007$, and a significant interaction between age level and level of psychopathology, $F(2,114) = 2.99$, $p < .05$. The age level effect did not reach significance.

To understand the nature of these significant relationships, t-test analyses were performed on the three diagnostic groups' means pooled across age levels and within each age level to isolate the source of interactive variation. Table 9 reports the results of these analyses.

Following up on the significant effect of level of psychopathology, Table 9 revealed that Control subjects scored significantly lower, thereby better, than children in the Chronic Disorder Group,

TABLE 8

Analysis of Variance on the Psychopathology Scale Scores

<u>Source of Variation</u>	<u>Mean Square</u>	<u>DF</u>	<u>F</u>	<u>Significance</u>
Age Level	283.82	1	1.53	.218
Group	972.15	2	5.25	.007
Interaction	554.00	2	2.99	.054
Residual	184.91	114		

TABLE 9

Psychopathology Scale Means and Standard Deviations

	<u>Control</u> <u>Group</u>	<u>Adjustment</u> <u>Group</u>	<u>Chronic</u> <u>Group</u>
Older Subjects	$\underline{M} = 48.4$ $\underline{SD} = 14.4$ (N = 20)	$\underline{M} = 44.3b$ $\underline{SD} = 9.0$ (N = 20)	$\underline{M} = 54.3b$ $\underline{SD} = 16.0$ (N = 20)
<hr/>			
Younger Subjects	$\underline{M} = 44.0ab$ $\underline{SD} = 9.0$ (N = 20)	$\underline{M} = 54.7b$ $\underline{SD} = 12.8$ (N = 20)	$\underline{M} = 57.5a$ $\underline{SD} = 17.8$ (N = 20)
<hr/>			
Pooled Data	$\underline{M} = 46.2a$ $\underline{SD} = 12.1$ (N = 40)	$\underline{M} = 49.5b$ $\underline{SD} = 12.0$ (N = 40)	$\underline{M} = 55.9ab$ $\underline{SD} = 16.8$ (N = 40)

Note: Means with identical subscripts are significantly different
in rows.

$a_p = .001$

$b_p = .01$

$t(38) = 2.96, p < .01$. The Adjustment Reaction Group also scored lower than the Chronic Disorder Group, $t(38) = 1.96, p < .05$. However, Control subjects scored lower, although not significantly so, than Adjustment Reaction subjects.

To appreciate the significance of the interaction in the absence of an age level effect, t -tests of mean scores were compared within each age level. Results are presented in Table 9. Among the younger subjects, Control children scored significantly lower (and thus better) than either the Adjustment Reaction Group, $t(38) = 3.08, p < .01$, or the Chronic Disorder Group, $t(38) = 3.01, p < .01$. The hypothesized trend of scores with the Control Group revealing the lowest score and the Chronic Disorder Group scoring the highest on the Psychopathology Scale was evident; however, the mean difference between the Adjustment Reaction Group and the Chronic Disorder Group did not reach significance.

Paradoxically, among older subjects this hypothesized trend did not hold true. Although the highest, thereby most psychopathological, score was obtained by the Chronic Disorder Group, the Adjustment Reaction Group scored the lowest. The Control Group scored between the two. The only significant difference appeared between the Chronic Disorder and the Adjustment Reaction Groups.

In sum, hypothesis 4, postulating significant differences among the levels of psychopathology was not totally confirmed, although this predicted trend was evident in the pooled data. As hypothesized,

there was no age effect in the Psychopathology Scale score data. It was hypothesized that there would be no interaction between age level and level of psychopathology. This was not confirmed. The significant interaction appeared to be the result of an anomalous reversal of predicted scores among older children; the normal children appeared more disturbed than the Adjustment Reaction subjects, although this difference did not reach significance.

Correlational Analysis and Analyses of Covariance

To evaluate the possible relationship among the three Bender Gestalt scoring systems given their apparent similarity in scorable factors and factor definitions, coefficients of correlation were calculated between each system for each diagnostic group and for the entire sample. Pearson Product-Moment Correlation Tables are presented in Tables 10 and 11. From these matrices, one pair of coefficients of correlation appeared consistently significant across levels of psychopathology, and accounted for the highest proportion of variation between variables. The relationship between the Koppitz Developmental Score and the Psychopathology Scale score was significant in the Control Group, $r = .506$, $p = .001$, in the Adjustment Reaction Group, $r = .651$, $p = .001$, and in the Chronic Disorder Group, $r = .635$, $p = .001$.

A second set of correlations demonstrated significance in two of the three levels of psychopathology; the relationship between the Emotional Indices Sum score and the Developmental Score. This

TABLE 10
Pearson Product-Moment Correlation Matrices

TOTAL SAMPLE

(N = 120)

	(1)	(2)	(3)
(1) Koppitz Developmental Score	-	.30*	.63*
(2) Emotional Indices Sum		-	.33*
(3) Psychopathology Scale Score			-

Note: * $p = .001$

CONTROL GROUP

(N = 40)

	(1)	(2)	(3)
(1) Koppitz Developmental Score	-	.44*	.50*
(2) Emotional Indices Sum		-	.43*
(3) Psychopathology Scale Score			-

Note: * $p = .001$

TABLE 11
Pearson Product-Moment Correlation Matrices

ADJUSTMENT REACTION GROUP

(N = 40)

	(1)	(2)	(3)	(4)
(1) Koppitz Developmental Score	-	.43**	.65***	-.24
(2) Emotional Indices Sum		-	.27*	-.38*
(3) Psychopathology Scale Score			-	-.43**
(4) Full Scale Intelligence				-

Note: * $p = .05$

** $p = .01$

*** $p = .001$

CHRONIC DISORDER GROUP

(N = 40)

	(1)	(2)	(3)	(4)
(1) Koppitz Developmental Score	-	.03	.63*	-.23
(2) Emotional Indices Sum		-	.15	-.25
(3) Psychopathology Scale Score			-	-.27
(4) Full Scale Intelligence				-

Note: * $p = .001$

relationship was significant in the Control Group, $r = .449$, $p = .001$, and in the Adjustment Reaction Group, $r = .431$, $p = .001$. The relationship between systems in the Chronic Disorder Group was not significant.

The unexpected and robust relationship between each system for assessing psychological disturbance and the standard Koppitz Developmental Score mandated a re-analysis of the data with the Developmental Score as a covariate. This provided statistical control for its effects as an extraneous variable on each diagnostic system. A 3(group) x 2(age level) analysis of covariance was performed on the Koppitz Emotional Indices data, with the Developmental Score serving as a covariate. Results are presented in Table 12. The Koppitz Developmental Score was a significant covariate, $F(1,113) = 13.10$, $p < .001$. The original significant effect, level of psychopathology, remained a significant effect in the re-analysis, $F(2,113) = 7.49$, $p < .001$, indicating the efficacy of the Emotional Indices to differentiate among levels of psychopathology independent of the child's level of visual-motor integration.

The re-analysis of the Psychopathology Scale data with the Developmental Score added as a covariate is presented in Table 13. The Koppitz Developmental Score was a significant covariate, $F(1,113) = 88.15$, $p < .001$. In the original analysis of Psychopathology Scale data (Table 8), the level of psychopathology main effect was significant. In this re-analysis, however, that main effect did not

TABLE 12

Analysis of Covariance on the Emotional Indices Sum With
the Koppitz Developmental Score a Covariate

<u>Source of Variation</u>	<u>Mean Square</u>	<u>DF</u>	<u>F</u>	<u>Significance</u>
Covariate: Developmental Score	15.39	1	13.10	.001-
Age Level	2.39	1	2.04	.156
Group	8.80	2	7.49	.001
Interaction	0.73	2	0.62	.538
Residual	1.17	113		

TABLE 13

Analysis of Covariance on the Psychopathology Scale Score
 With the Koppitz Developmental Score a Covariate

<u>Source of Variation</u>	<u>Mean Square</u>	<u>DF</u>	<u>F</u>	<u>Significance</u>
Covariate:				
Developmental Score	9966.25	1	88.15	.001-
Age Level	141.00	1	1.24	.266
Group	124.84	2	1.10	.335
Interaction	624.35	2	5.52	.005
Residual	113.05	113		

reach significance indicating that the Developmental Score accounted for a significant portion of the variation within the Psychopathology Scale. With the Developmental covariate removed, the Psychopathology Scale could not differentiate among levels of psychopathology.

In summary, the results of the analyses of covariance indicated that the Emotional Indices Sum appeared independent of the Developmental Score in its ability to differentiate among levels of disturbed children. Its validity was confirmed under stringent criteria. The Psychopathology Scale score was revealed to be significantly correlated to the Developmental Score. With the effects of the Developmental score statistically removed, the Psychopathology Scale score could not differentiate among levels of psychopathology, indicating inadequate validity for children younger than eleven years of age.

To examine the relationship between intelligence and scores on the Psychopathology Scale and the Emotional Indices, coefficients of correlation were computed for each of the two clinical groups and are reported in Table 11. Intelligence was significantly correlated with the Psychopathology Scale score, $r = -.43$, $p < .01$, and with the sum of Emotional Indices, $r = -.38$, $p < .05$, only in the Adjusted Reaction Group. Among the children in the Chronic Disorder Group, the relationship between the scores on either system of assessing psychopathology and intelligence was nonsignificant. Thus, hypothesis 6 which posited nonsignificant correlations between intelligence and the scores on the Psychopathology Scale and the Emotional Indices for

each clinical group was partially confirmed.

To examine the relationship between the scores on the Psychopathology Scale and the sum of Emotional Indices, coefficients of correlation were computed for each diagnostic group. The results are reported in Tables 10 and 11. The Psychopathology Scale score and the Emotional Indices sum demonstrated inconsistent correlations at the three levels of psychopathology. A significant relationship was found in the Control Group, $r = .43$, $p < .01$, and in the Adjustment Reaction Group, $r = .27$, $p < .01$. The two systems were unrelated in the Chronic Disorder Group. Thus, hypothesis 7, which posited a significant relationship between systems, was partially confirmed.

In sum, the results of this study offered limited support for the efficacy of the Emotional Indices sum to differentiate normal children from their disturbed peers. While the Emotional Indices sum could not differentiate between levels of psychopathology in the two established clinical groups, the Adjustment Reaction Group did produce fewer, but not significantly fewer, Emotional Indices than the Chronic Disorder Group. A clear trend in the hypothesized direction was evident. Similarly, the validity of the Psychopathology Scale was assessed in this study. Results indicated that the Psychopathology Scale score was significantly correlated with the Koppitz Developmental Score. When the variation attributed to the Developmental Score was statistically removed, the Psychopathology Scale score was unable to differentiate among the three levels of psychopathology. Its validity with school-aged children was judged inadequate.

CHAPTER V

DISCUSSION

This study had two purposes. First, it was designed as a psychometric study attempting to establish the criterion validity of two systems for assessing psychological disturbance on the Bender Gestalt performance of school-aged children. The results addressed that issue directly. A second, more heuristic purpose was to offer suggestions concerning the effective use of the Bender Gestalt Test in childhood personality assessment. This goal can be reached indirectly. As validity is inferred from logical measurement so too clinical suggestions involve an extrapolation from group data to individual decisions. This discussion will evaluate the results within the theme of comparable literature with a focus on effective clinical practice. Suggestions for subsequent research will be offered.

The overall results of this study indicated both the Psychopathology Scale and the Emotional Indices Sum score were significantly related to the Koppitz Developmental Score which assesses a child's visual-motor integration ability. Analyses of both diagnostic systems with the effects of the Developmental score statistically removed indicated that significant differences among clinical groups could be detected using the Indices but not with the Psychopathology Scale score. Each system was unrelated to the age level of the subjects;

children between seven and ten could be considered as one homogeneous sample for each of these diagnostic systems.

The Emotional Indices Sum significantly differentiated normal children from each level of psychopathology. Even when an attempt was made to statistically control for differences in Developmental level, the ability of this score to differentiate normal from disturbed children persisted. This suggested prima facia evidence for the validity of these Indices in the identification of disturbed children. Chronically disturbed children produced more Emotional Indices than Adjustment Reaction children, although this difference did not quite reach significance. What was evident was that the incidence of Emotional Indices was linearly related to the severity of the psychological disturbance; more disturbed children produced more Emotional Indices than their less disturbed peers, and significantly more than normal children.

This finding contrasts with the theme of previous research culminating with the conclusion of Buckley (1978) disavowing the Emotional Indices as valid indicators of emotional disturbance in school psychological assessment.

The present results appeared tempered and their clinical implications muted when these data were compared with normative data from other studies. Compared to large-scale samples from other studies it appeared that the range of scores in the present Control and Adjustment Reaction Groups would roughly fall into a "normal" range of

performance. The Chronic Disorder Group score was significantly higher, and clearly atypical. Yet within each of the three diagnostic groups there was a comparable range of scores of approximately zero to four Emotional Indices present. While the range of scores for the three groups were equivalent, the modal number of Emotional Indices had clinical implications. The modal number of Indices in the Control Group was 1, the Adjustment Reaction Group was bi-modal at 2/3 while the Chronic Disorder Group had a mode of 3 Emotional Indices. As with the mean scores, the mode scores reflect a positive and linear trend in the relationship between severity of psychological disturbance and the frequency of Emotional Indices.

To evaluate the present results in light of previous research it is necessary to compare each group with comparable groups from other studies. The sampling distribution of Control Group means appeared to be quite variable, with Koppitz' (1963) standardization sample the most conservative. Control Group children in the present study had a mean of 1.42 Emotional Indices. This appeared significantly larger, and thereby worse, than the Koppitz Control Group mean of 0.9 Indices. Paradoxically, the present sample mean was significantly lower than the Control Group mean score reported by Dibner and Korn (1966) in the largest and best identified normative sample. It appeared that Koppitz' normative data represent an underestimation of the incidence of Emotional Indices in normal children. This seems especially likely since Koppitz' normative sample had few minority or disadvantaged children included.

Similar comparisons among the clinical samples of other studies also suggested that the statistical significance of the Emotional Indices has little clinical significance. The mean number of Indices for the Adjustment Reaction Group and the Chronic Disorder Group was 2.12 and 2.57 respectively. Each was greater than the normative Koppitz mean of 1.9 for a sample of "emotionally disturbed" children. However, each group scored less than the group means reported by McConnell (1967) for analogously labeled clinical groups; Adjustment Reaction Group ($\bar{M} = 2.8$) and the Neurotic/Characterological Group ($\bar{M} = 2.7$). Similarly, the Adjustment Reaction Group in the present study was comparable in performance to the normal children in Dibner and Korn (1966). What can be concluded from these comparisons is that the Emotional Indices appear to vary considerably depending on the sample of disturbed children. The presence of two or more Emotional Indices in an individual record appears to reflect the possibility of psychological disturbance that warrants additional study.

Koppitz (1963) suggested a clinical rule of thumb that three or more Emotional Indices was strong evidence of psychopathology. Most children in the present study (62%) had Bender Gestalt protocols with fewer than three Indices present and the overall sample distribution of Emotional Indices was bi-modal at 2/3. Given this finding, a post hoc hypothesis was tested assessing Koppitz' contention ". . . that three or more EIs are necessary before one can say with some degree of confidence that a child has serious emotional problems. . . ." (p. 89). Koppitz' contention was for the identification of serious

emotional disturbance, operationally defined in the present study as Chronic Disordered subjects. Using that more stringent criterion, that of rendering a probability of Chronic Group membership given three or more Emotional Indices can be considered an application of Bayes' Theorem. Post hoc analysis indicated that the conditional probability that a protocol was of the Chronic Disorder Group given the presence of three or more Emotional Indices was .51. The implications of this conditional probability coefficient suggest that three or more Emotional Indices in a given record is merely suggestive of some type of psychopathology. The coincidence of serious, i.e., Chronic, psychopathology given three or more Indices is only that of chance expectation when a dichotomous clinical decision is required--whether a protocol represents a "seriously" disturbed child or not. This proportion exceeds the thirty-five per cent rate of "false positive" classifications reported in Koppitz' normative data. A three Emotional Indices cut-off point appears inadequate for making clinical decisions.

In sum, the Emotional Indices have demonstrated adequate validity in the identification of psychologically disturbed children. While they cannot differentiate among levels of psychopathology, the pattern of the data suggested that children evaluated as more disturbed tend to generate Bender Gestalt protocols with more Emotional Indices than peers evaluated as disturbed. Clinically, the presence of two or more Emotional Indices can be considered atypical and warrants additional consideration. However, two or more Indices

cannot be considered pathognomonic, nor pejorative without corroborating data.

The results of the present study cannot add support for the clinical use of the Psychopathology Scale among school-aged children. In this study the Psychopathology Scale failed to demonstrate adequate validity in the identification of disturbed children, or to differentiate among levels of psychopathology as hypothesized. Similarly, this scale appeared significantly confounded by developmental changes in children's visual-motor integration.

Analyses of the Psychopathology Scale data indicated that the hypothesized trend, with the Control Group scoring the lowest and the Chronic Disorder Group scoring the highest, was evident among younger children. Among the older children, however, there was an anomalous reversal of mean scores; the Control Group scored higher than the Adjustment Reaction Group. These trends were obviated upon re-analysis of the data with the Developmental Score added as a covariate. The diagnostic efficacy of the Psychopathology was eliminated with the inclusion of the Developmental Score into the analysis of covariance.

These data represent the only available data on the Psychopathology Scale for children between seven and ten years of age. The only comparable data would be the tentative data reported by Hutt (1977) for children ten to twelve years of age. These comparisons offer marginal cross-validation of the reliability of the

Psychopathology Scale to assess the performance of roughly comparable children. Hutt (1977) suggested that the Psychopathology Scale was valid in the assessment of school-aged children with normal intelligence. The present study addressed that question more systematically and could not confirm Hutt's anecdotal recommendation.

In comparison to the preliminary normative data (Hutt, 1977), the present results generally confirm the hypothesized pattern of scores. Control Group children in both studies performed equivalently. Children in the present Chronic Disorder Group performed most similarly to Hutt's undefined "disturbed" sample of children. The present Adjustment Reaction Group was anomalous. Its mean score was not significantly different for the Control Group mean and in comparison to Hutt's norms, it was equivalent to his Control Group.

In sum, the Psychopathology Scale appeared to lack the validity necessary for clinical use among young school-aged children. While scores among the three levels of psychopathology were in the hypothesized pattern, the instrument lacked the statistical efficacy to differentiate among groups. At this time, further research is needed to evaluate the merit of the Psychopathology Scale among children.

The results of any study must be evaluated within a methodological context. The logic of a criterion-related validation study assumes that the criterion itself possesses validity. For the present study, and all of its genre, the validity of psychiatric diagnoses was tacitly accepted. This problem was the major

methodological criticism of the early research on the personality applications of the Bender Gestalt with children. The seminal studies (Byrd, 1956; Clawson, 1959) were criticized for inexact subject identification. Heterogeneous groups of "emotionally disturbed" children cannot be considered adequate representatives of the clinical population. The present study offered the most refined inclusion criteria to date and can be considered a better example of homogeneous subject selection in accord with Achenbach's (1978) recommendations.

In a twenty-five year retrospective review of the related literature on the application of the Bender Gestalt Test to personality assessment, Laretta Bender concluded that ". . . the application of this test to various categories of personality deviation is a futile task until such time as these categories have some definition" (Tolor & Schulberg, 1963, p. xvi). Bender doubted that that condition would ever be adequately met, implying that psychiatric diagnosis would remain a speculative endeavor with considerable variation of individual differences among members of an identical diagnostic category.

The empirical results of the present study are tempered when interpreted in the alternative paradigm that the criterion of classification was itself marginally valid. Future research needs to assess the validity of the criterion as well as the validity of the dependent measures. This can be effected through more accurate

diagnosis as with the multi-axial model of the Diagnostic and Statistical Manual-III. By reducing intra-group variation with more precise diagnosis with the two required axes of DSM-III, and with studies designed to focus on specific age levels, the efficacy of the instrument can be better assessed.

Recommendations for subsequent research with the Emotional Indices should include an investigation of the Major versus Minor Indices for their respective efficacy. Among children eight to ten, four Indices, Confused Order, Small Size, Second Attempt and Expansion can be considered Major Indices. These demonstrated unequivocal power in the differentiation between the Control Group and the Clinical Group in the Koppitz (1963) standardization sample. The remaining Indices did not reach statistical significance in their ability to differentiate between groups, but were evident in the hypothesized direction.

The Emotional Indices might be standardized similar to the Developmental Scale in half-year intervals for greater consistency between Koppitz' systems. However, there is some doubt many developmental changes would occur. The list of Emotional Indices could also be restandardized similar to the Developmental Scoring System in 1974 to better represent minority children in proportion to their frequency in the population of children.

This study offered the initial data for normative patterns of children younger than ten years of age on the Psychopathology Scale.

Subsequent research must address larger-scale normative studies, validation studies with age specific samples and corroborating evidence of reliability over time. The present results documented the significant correlation between the Psychopathology Scale and the Koppitz Developmental Score. This relationship needs further elucidation with perhaps the Psychopathology Scale revised to include developmental consideration for children.

The Bender Visual Motor Gestalt Test remains an extremely popular instrument with extensive research and clinical applications. Its brevity, simplicity of administration and scoring and its interpretative flexibility account for its wide use especially with children. Psychologists like to give it and children like to take it. Such ease and enthusiasm have resulted in rather uncritical acceptance of the test, especially its contributions to personality assessment. Evidence indicates that groups of atypical children deviate from normal children in some aspects of their Bender Gestalt Test performance, notably their incidence of Emotional Indices. Bender reproductions of markedly atypical children are more different from those of normal children than from each other. Children diagnosed as seriously disturbed are the most atypical in their Bender Gestalt performance. Less disturbed children appear somewhat similar to normal children, although consistently performing at the hypothesized upper range of normal performance.

This study adduced evidence for the continued use of the

Koppitz Emotional Indices in the assessment of psychopathology in preadolescent children. In the initial comparison of the Emotional Indices with the Psychopathology Scale, the Emotional Indices demonstrated better statistical and clinical validity. Each system warrants further research to identify more specific populations of children who could benefit from an evaluation with the Bender Visual Motor Gestalt Test.

SUMMARY

The purpose of this study was to investigate the validity of two systems for assessing psychopathology in the Bender Gestalt Test reproductions of school-aged children between seven and ten years of age. The system of eleven discrete Emotional Indices, proposed by Koppitz was compared with the Psychopathology Scale standardized by Hutt for use with adolescents and adults. This study proposed the initial comparison of these scales and the original validation study of the Psychopathology Scale with preadolescent children. The specific aims of the study were to investigate the relationship among two age levels of children and three levels of psychological disturbance and the degree of psychopathology manifested on each scoring system.

The Bender Gestalt performance of three groups of children were compared on three measures: two systems of psychological disturbance and the standard Koppitz Developmental Score. Three groups of normally intelligent children were compared, a Control Group (N = 40), an Adjustment Reaction Group (N = 40), and a Chronic Disorder Group (N = 40). Membership in the latter two groups was determined by psychiatric diagnosis. Each group was divided equally into younger (ages 7 to 8) and older (ages 9 and 10) children. The total sample size was 120 children.

Results indicated that both the Psychopathology Scale and the Sum of Emotional Indices were significantly correlated with the Koppitz Developmental Score. Analyses of covariance indicated that the Emotional Indices significantly differentiated the normal children from each group of disturbed children; however, it could not differentiate between the two disturbed groups of children although there was a trend in that direction evident. The Psychopathology Scale score, once the Developmental covariate was statistically removed, was unable to differentiate among the three groups.

The results generally supported the validity of the Emotional Indices as a clinical scale and could not support use of the Psychopathology Scale with preadolescent children. The implications of these results for clinical practice were a cautious endorsement of the Emotional Indices Sum score of two or more as an alerting sign of potential psychopathology in children. Records with two or more Emotional Indices were considered typical of disturbed children though not pathognomonic in normal children. Recommendations for subsequent research on the personality assessment aspects of the Bender Gestalt Test were suggested.

REFERENCES

- Achenbach, T. M. Psychopathology of childhood: Research problems and issues. Journal of Consulting and Clinical Psychology, 1978, 46, 759-776.
- Ackerman, P. T., Peters, J. E., & Dykman, R. A. Children with specific learning disabilities: Bender Gestalt findings and other signs. Journal of Learning Disability, 1971, 35, 437-446.
- Bender, L. A. A visual motor gestalt test and its clinical use. New York: American Orthopsychiatric Association, 1938.
- Bender, L. A. Manual for instructions and test cards for visual motor gestalt test. New York: American Orthopsychiatric Association, 1946.
- Blau, T. H. Diagnosis of disturbed children. American Psychologist, 1979, 34, 969-972.
- Brannigan, G. G., Barone, R. J., & Margolis, H. Bender Gestalt signs as indicants of conceptual impulsivity. Journal of Personality Assessment, 1978, 42, 233-236.
- Brown, W. R., & McGuire, J. M. Current psychological assessment practice. Professional Psychology, 1976, 42, 475-484.
- Buckley, P. D. The Bender Gestalt Test: A review of the reported literature with school-aged subjects, 1966-1977. Psychology in the Schools, 1978, 15, 327-338.
- Byrd, E. The clinical validity of the Bender Gestalt Test with children: A developmental comparison of children in need of psychotherapy and children judged well-adjusted. Journal of Projective Techniques, 1956, 20, 127-136.
- Cass, L. K. The training of clinical child psychologists. In G. J. Williams & S. Gordon (Eds.), Clinical child psychology. New York: Behavioral Publications, 1974.
- Clawson, A. The Bender Visual Motor Gestalt Test as an index of emotional disturbance in children. Journal of Projective Techniques, 1959, 23, 198-206.

- Dibner, A. S., & Korn, E. J. Group administration of the Bender Gestalt Test to predict early school performance. Journal of Clinical Psychology, 1966, 25, 265-268.
- Eme, R. F. Sex differences in childhood psychopathology: A review. Psychological Bulletin, 1979, 86, 574-595.
- Flint, F. S. A validation and developmental study of some interpretations of the Bender Gestalt Test. (Doctoral Dissertation, New York University). Dissertation Abstracts International, 1966, 27, 606B. (Abstract)
- Greenbaum, R. S. A note on the use of the word association test as an aid to interpreting the Bender Gestalt. Journal of Projective Techniques, 1955, 19, 27-29.
- Gregory, M. L. Emotional indicators on the Bender Gestalt and the Deverux Child Rating Scale. Psychology in the Schools, 1977, 14, 433-437.
- Gries, L. T. The role of the psychologist in a developmental center. Professional Psychology, 1978, 9, 685-691.
- Handler, L., & McIntosh, J. Predicting aggression and withdrawal in children with the D-A-P and the Bender Gestalt. Journal of Personality Assessment, 1971, 35, 331-337.
- Hutt, M. L. Revised Bender Visual Motor Gestalt Test. In A. Weider (Ed.), Contributions toward medical psychology. New York: Ronald Press, 1950.
- Hutt, M. L. The Hutt adaptation of the Bender Gestalt Test. (2nd ed.). New York: Grune & Stratton, 1969.
- Hutt, M. L. The Hutt adaptation of the Bender Gestalt Test. (3rd ed.). New York: Grune & Stratton, 1977.
- Hutt, M. L., & Briskin, G. J. The Hutt adaptation of the Bender Gestalt Test. New York: Grune & Stratton, 1960.
- Hutt, M. L., & Dates, B. G. Reliabilities and interrelationships of two HABGT scales for a male delinquent population. Journal of Personality Assessment, 1977, 41, 353-357.
- Hutt, M. L., Dates, B. G., & Reid, D. M. The predictive ability of HABGT scales for male delinquent population. Journal of Personality Assessment, 1977, 41, 492-496.

- Hutt, M. L., & Miller, L. J. Interrelationships of psychopathology and adience-abience. Journal of Personality Assessment, 1976, 40, 135-139.
- Keogh, B. K. The Bender Gestalt with children: Research implications. Journal of Special Education, 1969, 3, 15-22.
- Koss, M. P. Descriptive characteristics and length of psychotherapy of children and adolescents seen in private psychological practice. Psychotherapy: Theory, Research and Practice, 1980, 17, 268-271.
- Koppitz, E. M. The Bender Gestalt Test for young children. New York: Grune & Stratton, 1963.
- Koppitz, E. M. The Bender Gestalt Test for young children. Volume II: Research and applications, 1963-1973. New York: Grune & Stratton, 1975.
- McConnell, O. L. Koppitz' Bender Gestalt scores in relation to organic and emotional problems in children. Journal of Clinical Psychology, 1967, 23, 370-374.
- Miller, L. J., & Hutt, M. L. Psychopathology scale of the Hutt Adaptation of the Bender Gestalt Test: Reliability. Journal of Personality Assessment, 1975, 39, 129-131.
- Mogin, L. S. Administration and objective scoring of the Bender Gestalt Test in group screening of primary grade children for emotional maladjustment. (Doctoral Dissertation, Rutgers--The State University). Dissertation Abstracts International, 1966, 27, 1665A. (Abstract)
- Naches, A. M. The Bender Gestalt Test and acting out behavior in children. (Doctoral Dissertation, Colorado State University). Dissertation Abstracts International, 1967, 27, 2146B. (Abstract)
- Nathan, P. E. Diagnostic and treatment services for children. American Psychologist, 1979, 34, 967-968.
- Oliver, R. A., & Kronenberger, E. J. Testing the applicability of Koppitz's Bender-Gestalt scores to brain damaged, emotionally disturbed and normal adolescents. Psychology in the Schools, 1971, 8, 250-253.
- Pfeffer, C. R., & Plutchik, R. Psychopathology of latency age children. Journal of Nervous and Mental Disease, 1982, 170, 193-197.

President's Commission on Mental Health. Volume I. Washington, D.S.: U. S. Government Printing Office, 1978.

Simpson, W. H. A study of some factors in the Bender Gestalt reproductions of normal and disturbed children. (Doctoral Dissertation, The University of Oklahoma, 1958). Dissertation Abstracts International, 1958, 26A, 1120. (Abstract)

Trahan, D., & Stricklin, A. Bender Gestalt emotional indicators and an acting out behavior in young children. Journal of Personality Assessment, 1979, 43, 365-375.

Watkins, J. M., & Watkins, D. A. Comparison of normal and emotionally disturbed children by the Plenk system for the Bender Gestalt. Journal of Clinical Psychology, 1975, 31, 71-74.

Zuelzer, M. B., Stedman, J. M., & Adams, R. Koppitz Bender Gestalt scores in first grade children as related to ethnocultural, socioeconomic, class and sex factors. Journal of Clinical Psychology, 1976, 44, 875.

APPENDIX A

HUTT PSYCHOPATHOLOGY SCALE FACTORS

1. Sequence

DEFINITION

Sequence refers to the relative degree of regularity in the successive placement of the drawings on the page(s) used in making the reproductions. The expected order of successive placements is either from left to right or from top to bottom. Sequence is scored in terms of the number of shifts which occur as a deviation from the order in succession which the examinee has already manifested. When the examinee places his successive drawings in order from right to left or from bottom to top, this is scored as one shift in sequence. Otherwise, each time the sequence is altered it is counted as one shift. Thus, when an examinee has been arranging his drawings from left to right and then places the next figure above or below or to the left, this is counted as a shift in sequence. The score on Sequence is the total number of shifts. If an examinee places each drawing on a separate page, this is considered an irregular sequence (see below) and is counted as 2 shifts. (Note: Allowance is made, and no score for shift in sequence is counted, when the examinee reaches the edge of the paper and places the next figure in an appropriate space below.)

Sequence is characterized as: normal or methodical; overly methodical; irregular; confused or symbolic. The definitions and Psychopathology Scale values (PSV) for each characteristic are given below.

Confused or symbolic sequence (PSV = 10.0): Drawings are placed in a jumble on the page--that is, without any apparent plan--or some symbolic arrangement is evident. A symbolic placement may be a figure-eight pattern, spiral pattern, etc.

Irregular sequence (PSV = 7.0): More than one shift in sequence, but no obvious confusion or jumble can be detected.

Overly methodical sequence (PSV = 4.0): No shift occurs; the figures are placed in rigid sequence, without deviation.

Normal or methodical sequence (PSV = 1.00): Only one shift in sequence occurs.

2. Position of the First Drawing

DEFINITION

This factor refers to the placement of figure A on the page, whether the examinee uses a single page for all of the drawings or uses more than one page.

Abnormal placement (PSV = 3.25): Any portion of the drawing is within one inch of any edge of the page.

Normal placement (PSV = 1.0): The entire figure lies within the upper one-third of the page, and no portion of the figure is less than one inch from any edge of the page.

Other placements: These are not scored but will be considered in terms of possible interpretative significance.

3. Use of Space, 1

DEFINITION

This refers exclusively to the spacing between successive or adjacent drawings. The criterion is always the preceding or adjacent figure, and space is judged in terms of the relevant axis of that figure. Thus, when space is judged in relation to two figures in the same vertical plane, the vertical axis of the drawing above is considered. When the drawing is placed between two figures (either to the left or above), the relevant axis of the nearer figure is considered.

This factor can be either normal or abnormal. It is abnormal if the space between two successive drawings is either more than half the size of the relevant axis of the preceding or adjacent figure, or is less than one-fourth the size of that axis.

Abnormal use of space (PSV = 10.0): It can be abnormal and excessively constricted, or it can be abnormal and excessively expansive. In either case, abnormal space is scored when the phenomenon, as defined above, occurs on two or more occasions.

4. Collision

DEFINITION

Collision refers to the actual running together or overlapping of one figure by another. The perimeter of one figure must either touch or overlap the perimeter of another figure. If the line of dots or circles of one figure intrudes into the open space of an adjacent figure, but there is no actual touching or overlapping of the perimeter, this is not counted as collision; rather it is considered a collision tendency, and is so scored and interpreted.

There are seven subcategories of collision and collision tendency.

Collision, extreme (PSV = 10.0): Collision occurs more than 2 times.

Collision, moderate (PSV = 8.5): Collision occurs 2 times.

Collision, present (PSV = 7.0): Collision occurs 1 time.

Collision tendency, extreme (PSV = 5.5): Occurs more than 2 times.

Collision tendency, moderate (PSV = 4.0): Occurs 2 times.

Collision tendency, present (PSV = 2.5): Occurs 1 time.

No collision or collision tendency (PSV = 1.0).

When the phenomena of both collision and collision tendency are present in the same record, the scores for both are summated, but in no case can the total score for this factor be more than 10.0.

5. Shift in the Position of the Paper

DEFINITION

This factor refers to the actual rotation of the test paper from the vertical position, in which it is presented, to the horizontal position (or a shift approximating 90 degrees). (This factor should not be confused with rotation or perceptual rotation, as defined in Factor 6, below.) Three degrees of paper rotation are considered.

Rotation of paper for all figures (PSV = 10.0).

Rotation of paper for 3 to 8 figures (PSV = 7.0).

Rotation of paper for 1 or 2 test figures (PSV = 5.5).

No rotation of paper (PSV = 1.0).

6. Closure Difficulty

DEFINITION

This factor refers to difficulty in "joining" parts within a figure or two adjacent figures which touch each other. The figures on which closure difficulty may occur are A, 2, 4, 7, and 8. Problems in "joining" may occur in connection with completing the circle or the diamond in figure A, in completing the circles in figure 2, in joining the sides of figure 4 or the junction of the curve and the open square in the same figure, and in joining the sides or connecting the adjacent parts in figures 7 and 8. The difficulty may be manifested in a number of ways: gaps at the point of joining; "overshooting the joining" (overlap at this point); erasures or corrections at the point of joining; and, noticeably, increased line pressure or redrawing at such points.

The raw score on this factor is the total number of closure difficulties which occur, but there are two special provisions: (1) no more than 2 closure difficulties may be counted on any one figure; and (2) on figure 2, count 2 closure difficulties as a raw score of 1, and 3 or more closure difficulties as a raw score of 2.

There are five subcategories of this factor: very severe, severe, moderate, mild, and absent.

Very severe (PSV = 10.0): Raw score is 9 or above.

Severe (PSV = 7.75): Raw score is 6-8.

Moderate (PSV = 5.5): Raw score is 3-5.

Mild (PSV = 3.25): Raw score is 1-2.

Absent (PSV = 1.0): Raw score is 0.

7. Curvature Difficulty

DEFINITION

This difficulty refers to any obvious change in the nature of the curves in figures 4, 5, and 6. Such changes occur when: there is an increase or decrease in the amplitude of the curve; the curve is replaced by straight or spiked lines; the curve is flattened; the curve is made unevenly or irregularly; the number of loops on figures 4 and 6 is either increased or decreased. The raw score is the total number of figures on which such changes occur.

Severe (PSV = 10.0): Raw score is 3.
 Moderate (PSV = 7.0): Raw score is 2.
 Mild (PSV = 4.0): Raw score is 1.
 Absent (PSV = 1.0): Raw score is 0.

8. Change in Angulation

DEFINITION

This change refers to an increase or decrease, by 15 degrees or more, of the angulation in stimulus figures 2, 3, 4, 5, 6, and 7. In figure 2, a change of 15 degrees or more in the angle the columns make with the horizontal rows is the criterion. In figure 3 it is a change in the angle made by the columns of dots. In figure 4 it is the angle of the curve in relation to the open square. In figure 5 it is the angle of the external dots in relation to a presumed horizontal line. In figure 6 it is the angle at which the two curves cross. In figure 7 it is the angle of intersection of the two figures. The score is the number of figures on which the defined degree of change is present.

Five figures (PSV = 10.0).
 Four figures (PSV = 8.0).
 Three figures (PSV = 6.0).
 Two figures (PSV = 4.0).
 One figure (PSV = 2.0).
 No figures (PSV = 1.0).

9. Perceptual Rotation

DEFINITION

This factor refers to the reproduction of the test figure with a rotation of its major axis while the stimulus card and the test paper are in their normal, standard positions. This factor is called "perceptual" rotation to distinguish it from factor 7, Shift in the Position of the Paper. This factor is not scored if only a segment of the figure is rotated (as when there is a shift in the position of the external line in figure 5). Similarly, change in only some of the columns of figure 2 is not scored as rotation. The score is the highest score obtained.

Severe rotation (PSV = 10.0): Rotation of 80-180 degrees on any figure.

Moderate rotation (PSV = 7.0): Rotation of 15-79 degrees on any figure.

Mild rotation (PSV = 4.0): Rotation of 5-14 degrees on any figure.

No rotation (PSV = 1.0): No rotation on any figure greater than 4 degrees.

10. Retrogression

DEFINITION

Retrogression refers to the substitution of a more primitive Gestalt form for the more mature Gestalt of the stimulus figure. The most common forms of retrogression are as follows: substitution of impulsive loops for reasonably well-formed circles (as in figure 2); substitution of dashes for dots (as in figures 1, 3, and 5); and substitution of dots for circles (as in figure 2). The criterion for counting such phenomena as retrogression is its occurrence at least twice on any given figure. The psychopathology scale value then depends upon the number of times this criterion of retrogression is met.

Severe retrogression (PSV = 10.0): Present in more than 2 figures.

Moderate retrogression (PSV = 7.0): Present in 2 figures.

Mild retrogression (PSV = 4.0): Present in 1 figure.

No retrogression (PSV = 1.0): No retrogression, as defined, occurs in any figure.

11. Simplification

DEFINITION

Simplification refers to the reproduction of the stimulus as a simpler figure or one which is much simpler to draw. The following are examples of simplification: drawing the two parts of figure A as noncontiguous; reducing the number of elements in figures 1, 2, 3, and 5, by at least 3 less than in the stimulus; reducing the number of curves in figure 6; and reproducing the parts of figures 7 and 8 as rectangles or crude ellipses. Simplification is not scored when either fragmentation or retrogression is involved; i.e., only

one phenomenon is scored per drawing.

Severe (PSV = 10.0): Present on more than 2 figures.
 Moderate (PSV = 7.0): Present on 2 figures.
 Mild (PSV = 4.0): Present on 1 figure only.
 None (PSV = 1.0).

12. Fragmentation

DEFINITION

Fragmentation refers to the essential destruction of the Gestalt. It can be manifested in a number of ways. The most common are as follows: the reproduction is obviously unfinished or incomplete; or the Gestalt is drawn as if composed of separate parts, i.e., the parts are grossly separated so that the Gestalt is lost.

Severe (PSV = 10.0): Present on more than 2 figures.
 Moderate (PSV = 7.0): Present on 2 figures.
 Mild (PSV = 4.0): Present on only 1 figure.
 None (PSV = 1.0).

13. Redrawing of the Total Figure

DEFINITION

Redrawing refers to a second attempt to reproduce a figure when the first attempt (which may not necessarily include reproduction of the total figure) is left without complete erasure or is simply crossed out. On rare occasions more than two attempts may be made to reproduce a figure.

Very severe (PSV = 10.0): Present 4 or more times.
 Severe (PSV = 7.75): Present 3 times.
 Moderate (PSV = 5.5): Present 2 times.
 Mild (PSV = 3.25): Present 1 time.
 None (PSV = 1.0).

14. Elaboration of Doodling

DEFINITION

This factor refers to doodling or elaboration on the reproduction in which the form is markedly changed. The elaboration can include the addition of loops or curlicues, or the addition of lines or curves which change the Gestalt. This factor should be differentiated from Factor 22, Perseveration.

Severe (PSV = 10.0): Present on more than 2 figures.

Moderate (PSV = 7.0): Present on 2 figures.

Mild (PSV = 4.0): Present on 1 figure only.

None (PSV = 1.0).

15. Perseveration

DEFINITION

Perseveration refers to either of two types of perseverative phenomena: (a) perseveration in which elements of a previous design are utilized in a succeeding design when they are not present in the stimulus figure; and (b) perseveration of the elements present in a given figure beyond the limits called for in the stimulus. An example of perseveration of type (a) is the use of dots in figure 2 (instead of circles) perseverated from the dots used in figure 1. An example of perseveration of type (b) is the presence in figure 1 of 14 or more dots instead of the 12 dots present in the stimulus design. Another example is the presence of 12 or more columns of circles in figure 2 instead of the 10 columns present in the stimulus design. The criterion for scoring the presence of perseveration is the occurrence of 2 or more of the perseverated elements in the case of type (a), or the addition of 2 or more elements (perseverated) within the same figure in the case of type (b).

Severe (PSV = 10.0): Present in more than 2 figures.

Moderate (PSV = 7.0): Present in 2 figures.

Mild (PSV = 4.0): Present in only 1 figure.

None (PSV = 1.0).

APPENDIX B

Scoring Manual for the Emotional Indices

I. CONFUSED ORDER

Figures are scattered arbitrarily on the paper without logical sequence or order. Credit is given for any sort of order or logical sequence. This may include arrangements from top of page down and up again, from left to right or from right to left. Subject is not penalized for placing Figure 8 at top of page if there is no room left at the bottom or side of page.

II. WAVY LINE

Two or more abrupt changes in the direction of the line of dots or circles in Figure 1 and Figure 2 respectively. A continuous, gradual curve or rotation of line is not scored for this category. A change of direction must involve at least two consecutive dots or circles. A single dot or column of circles out of line is not scored. Wavy Line is only scored once regardless of whether one or both Figures show this deviation.

III. DASHES SUBSTITUTED FOR CIRCLES (FIGURE 2)

At least half of all circles on Figure 2 are replaced with dashes 1/16 inch or longer. Substitution of dots for circles is not scored as this does not seem to differentiate between children with and without serious emotional problems.

IV. INCREASING SIZE (FIGURES 1, 2 and 3)

Dots and circles increase progressively in size until the last ones are at least three times as large as the first ones. This item is scored only once regardless of whether it occurs on one or three figures..

V. LARGE SIZE

One or more designs are drawn one third larger in both directions than design on stimulus card. When a design consists of two parts, then both parts have to be enlarged in order to qualify for this category. Large Size is only scored once regardless of whether one or all nine designs are enlarged.

VI. SMALL SIZE

One or more designs are drawn half as large as the designs on the stimulus card. The size of each figure is measured in each direction. When a drawing consists of two parts, then both parts have to be reduced in size in order to score in this category. Small Size

is scored only once regardless of whether one or all nine designs are reduced in size.

VII. FINE LINE

Pencil line is so thin that it requires effort to see the completed design.

VIII. OVERWORK OR REINFORCED LINES

Total design or part of it is redrawn or reinforced with heavy, impulsive lines. The design may be first erased and then redrawn or it may be corrected without any erasures. The item is scored only once regardless of whether overwork occurs once or several times on the record.

IX. SECOND ATTEMPT AT DRAWING

Drawing of design or part of it is spontaneously abandoned before or after it has been completed and a new drawing of the design is made. This item is only scored when two or more distinct drawings are made of one design on two different locations on the paper. It is not scored when a drawing on the same spot after the first was erased. When a drawing is erased and a second drawing is then made on a different location on the paper it is scored.

X. EXPANSION

Two or more drawings are used to complete the drawing of all nine Bender designs. The reverse of the paper counts as another sheet.

XI. BOXED IN

A box is drawn around one or more designs after they have been copied. This is an extremely rare occurrence.

XII. SPONTANEOUS ELABORATION

Spontaneous changes are made on one or more Bender Test figures that turn them into objects or combine them into bizarre designs. These kinds of drawings are rare.

APPENDIX C

Dear Parent:

Your child has been invited to take part in a brief research study being conducted as a thesis/dissertation within the Psychology Department of Loyola University of Chicago, if you consent. This project has been fully approved by the Institutional Review Board of the university. This research will attempt to explain how different groups of children perform on a test of visual-motor integration, essentially how groups of children learn to accurately draw geometric figures.

Your child has been invited because of his/her age and placement in a normal classroom. He/She will be just one member of the Normal Control Group. There will be 40 children in this group. Your child's drawings will be mixed in with the other 39 so that it will not be possible to identify any child's drawings. This study will be totally confidential and anonymous. No names will be recorded.

The task is called the Bender Visual Motor Gestalt Test, a common drawing test that has been used with children since 1938. It takes about five minutes and will be administered at school by a doctoral candidate in clinical psychology with great experience with children. He will make the task as easy and interesting as possible. Kids like this test.

Please understand that your consent or refusal will in no way affect your child's educational evaluation. You may withdraw your consent at any time without question. Similarly, your child's informed consent is required. If he/she decides not to participate there will be no questions asked. Because this is a group test, and a confidential study, individual test results will not be available to parents or teachers. Most children in this age range do quite well and earn a nearly perfect score.

Thank you for your interest in this project.

Sincerely,

Edward D. Rossini

Parent Consent Form

I give permission for my child _____ to participate in a brief drawing task. I understand that this test will take approximately five minutes and that the results will be totally confidential; my child's name will not appear anywhere nor will it be possible to identify my child's drawings from all of those collected. I have been informed that this research is part of a Master's Thesis conducted within the Psychology Department of Loyola University of Chicago. I understand that my child can discontinue his/her participation at any time without question. This project will in no way have any negative effect on my child's academic progress or placement.

This release will be destroyed after the project is completed.

Signature of Parent

Date

APPROVAL SHEET

The thesis submitted by Edward D. Rossini has been read and approved by the following Committee:

Dr. J. Clifford Kaspar, Director
Associate Professor, Psychology, Loyola

Dr. Patricia A. Rupert
Assistant 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.

12/6/82
Date

J. Clifford Kaspar
Director's Signature