1984

KINETIC FAMILY DRAWINGS AND CLINICAL JUDGMENT: AN EVALUATION OF JUDGES' ABILITY TO DIFFERENTIATE BETWEEN THE K-F-D'S OF ABUSING, CONTROL, AND CONCERNED MOTHERS.

PAMELA SUSAN. HOWITT
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation
https://scholar.uwindsor.ca/etd/2556

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.
NOTICE
The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.
KINETIC FAMILY DRAWINGS AND CLINICAL JUDGMENT: AN EVALUATION OF JUDGES' ABILITY TO DIFFERENTIATE BETWEEN THE K-F-D'S OF ABUSING, CONTROL, AND CONCERNED MOTHERS

by

PAMELA SUSAN HOWITT
Hon. B.A., Brock University, 1975
M.A., University of Windsor, 1977

A Dissertation Submitted to the Faculty of Graduate Studies through the Department of Psychology in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the University of Windsor
Windsor, Ontario, Canada

1984
ABSTRACT.

The purposes of the present study were: (1) to investigate the validity of the Kinetic Family Drawing (Burns & Kaufman, 1970) for use in child-abuse evaluations; (2) to explore variables which may affect the accuracy of clinical judgments based on the drawings. Drawings and social history data were obtained from three groups of mothers, 21 in each group, defined as Abusing, Control, and Concerned mothers. Twelve judges who had varied professional experience and assessment expertise rendered judgments about the mothers' group membership. Judgments were made using: (1) drawings alone; (2) social histories alone; (3) drawings plus social histories combined. Additionally, the artistic quality of the drawings was rated by three artists. Results indicated that:

1. judges could not differentiate between the mothers' drawings;
2. naive judges were more accurate than clinicians;
3. training in drawing interpretation did not improve judges' accuracy;
4. judgments based on drawings plus social histories were more accurate than judgments based on drawings alone;
5. artistic quality of drawings did not influence judgments of mothers' group membership.

Implications of these results were discussed, and suggestions for future research were offered.
ACKNOWLEDGEMENTS

I am especially grateful to Dr. C. Ron Frisch for his guidance in the preparation of this dissertation. His support and advice were greatly appreciated. In addition, I wish to thank Dr. Barry Taub and Dr. Stewart Page, who served as committee members. Dr. Taub's expressed belief in the worthiness of this project, from its inception, was a vital source of encouragement, and Dr. Page's suggestions regarding the methodology were most helpful. I am thankful for the contributions of Dr. Leonard Handler, who acted as the outside reader. His time and interest were much appreciated.

I am grateful to the mothers who volunteered to participate in this experiment, and to the individuals who served as judges. Special thanks go to the three clinicians, who arranged their schedules to accommodate the needs of the study, and offered time "above and beyond the call." Appreciation is also extended to the three artists who participated, and to Ms. Linda Scharf, who served as the second rater of the drawings. Additionally, I wish to thank Mr. Rex Kline, who assisted with the data analysis, and Ms. Eleanor Bultuck, who provided excellent typing services.

I gratefully acknowledge the assistance of the following agencies and individuals:

Children's Hospital of Michigan: Dr. Flossie Cohen, Chairperson, Research Grants and Investigations Committee

Ms. Alice Belden, Director of Social Services
Parents and Children Together
(P.A.C.T.):

Ms. Guadalupe Lara, social worker
Ms. Oretha Seay, social worker
Ms. Shirley Gray, social worker

Ms. Dorothy Kispert, Director
Ms. Nanette Gottfried, Training Coordinator

Wayne County Probate Court – Juvenile Division
Wayne County Clinic for Child Study

Finally, I wish to express my sincere appreciation to Dr. Joséph Avore. "Boss," mentor, and friend, his contributions were both real and spiritual. He offered invaluable commentary on the design, ran interference with various bureaucracies, and, most importantly, never stopped believing. Truly, he shares in this accomplishment.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Assessment and Child Abuse - An Overview</td>
<td>1</td>
</tr>
<tr>
<td>The Kinetic Family Drawing - Theory and Research</td>
<td>5</td>
</tr>
<tr>
<td>Issues in Projective Test Validation</td>
<td>11</td>
</tr>
<tr>
<td>Clinical Judgment and Figure Drawings</td>
<td>16</td>
</tr>
<tr>
<td>Additional Considerations in Clinical Judgment Research</td>
<td>19</td>
</tr>
<tr>
<td>Experience and Training as Variables</td>
<td>20</td>
</tr>
<tr>
<td>Elements of the Judgment Process</td>
<td>24</td>
</tr>
<tr>
<td>Relative Contributions of Data</td>
<td>28</td>
</tr>
<tr>
<td>The Confounding Influence of Aesthetics</td>
<td>31</td>
</tr>
<tr>
<td>To Summarize</td>
<td>34</td>
</tr>
<tr>
<td>Statement of the Problem and Hypotheses</td>
<td>35</td>
</tr>
<tr>
<td>II. METHOD</td>
<td>38</td>
</tr>
<tr>
<td>Definitions of the Mothers' Groups</td>
<td>38</td>
</tr>
<tr>
<td>Case Selection Criteria for the Mothers' Groups</td>
<td>39</td>
</tr>
<tr>
<td>Definitions of the Judges' Groups</td>
<td>42</td>
</tr>
<tr>
<td>Materials</td>
<td>43</td>
</tr>
<tr>
<td>Procedures</td>
<td>44</td>
</tr>
<tr>
<td>Statistical Analyses</td>
<td>51</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>III. RESULTS</td>
<td>53</td>
</tr>
<tr>
<td>Sociodemographic Data</td>
<td>54</td>
</tr>
<tr>
<td>K-F-D Scoring</td>
<td>59</td>
</tr>
<tr>
<td>Judges' Ability to Differentiate the Drawings</td>
<td>67</td>
</tr>
<tr>
<td>Factors Which Influenced Judges' Performance</td>
<td>69</td>
</tr>
<tr>
<td>Judges' Race and Sex</td>
<td>69</td>
</tr>
<tr>
<td>Abuse Experience/Assessment Expertise</td>
<td>71</td>
</tr>
<tr>
<td>Training</td>
<td>75</td>
</tr>
<tr>
<td>Types of Data</td>
<td>75</td>
</tr>
<tr>
<td>Artistic Quality</td>
<td>80</td>
</tr>
<tr>
<td>IV. DISCUSSION</td>
<td>86</td>
</tr>
</tbody>
</table>

APPENDICES

A. Social History Questionnaire Completed by Mothers | 105 |
B. Social History Data Sheet Provided to Judges | 108 |
C. Kinetic Family Drawing Scoring Guide | 111 |
D. Explanation of Research Provided to Mothers | 114 |
E. Informed Consent Form for Mothers | 116 |
F. Explanation of Study Provided to Judges | 118 |
G. Explanation of the Kinetic Family Drawing Given to Drawing-Naive Judges | 121 |
H. Judges' Background Information Form | 123 |
I. Judges' Informed Consent Form | 125 |
J. Sample Rating Form for K-F-D's | 127 |
K. Sample Rating Form for Social Histories | 129 |
L. Instructions to the Judges | 131 |
<table>
<thead>
<tr>
<th>APPENDICES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. Instructions and Rating Materials for Artists' Evaluation of Aesthetic Quality</td>
<td>134</td>
</tr>
<tr>
<td>N. Summary Letter Sent to Interested Mothers</td>
<td>137</td>
</tr>
<tr>
<td>O. Summary Letter Sent to Interested Judges</td>
<td>140</td>
</tr>
<tr>
<td>P. Base-Rate Scores on the 27 K-F-D Scoring Guide Variables for 63 Mothers</td>
<td>142</td>
</tr>
<tr>
<td>Q. Percentages of Drawings Elicited by Four Examiners Which Contained K-F-D Features</td>
<td>145</td>
</tr>
<tr>
<td>R. Means for K-F-D Features of Drawings Elicited by Four Examiners</td>
<td>147</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>149</td>
</tr>
<tr>
<td>VITA AUCTORIS</td>
<td>157</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Summary of Analyses for Mean Age of Mothers in Abusing (A),</td>
<td>57</td>
</tr>
<tr>
<td>Control (C₁), and Concerned (C₂) Groups</td>
<td></td>
</tr>
<tr>
<td>2. Summary of Analyses of Selected Social History</td>
<td>58</td>
</tr>
<tr>
<td>Variables for Three Groups of Mothers</td>
<td></td>
</tr>
<tr>
<td>3. Summary of Chi-Square Analysis for Presence of Barriers in</td>
<td>61</td>
</tr>
<tr>
<td>the K-F-D's of Abusing, Control, and Concerned Mothers</td>
<td></td>
</tr>
<tr>
<td>4. Summary of Chi-Square Analysis for Number of Household Members</td>
<td>62</td>
</tr>
<tr>
<td>in the K-F-D's of Abusing, Control, and</td>
<td></td>
</tr>
<tr>
<td>Concerned Mothers</td>
<td></td>
</tr>
<tr>
<td>5. Summary of Analyses of Additional K-F-D Variables for</td>
<td>64</td>
</tr>
<tr>
<td>the Three Groups of Mothers</td>
<td></td>
</tr>
<tr>
<td>6. Summary of Chi-Square Analyses for Actions are Strange,</td>
<td>66</td>
</tr>
<tr>
<td>Rotation, and Evasion in the K-F-D's of Mothers in SES Groups 1, 2,</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>7. Judges' Hit-Rates and False Positive/False Negative Errors</td>
<td>68</td>
</tr>
<tr>
<td>in Differentiating Between Abusing and Nonabusing Mothers</td>
<td></td>
</tr>
<tr>
<td>8. Judges' Hit-Rates in Differentiating Between Abusing,</td>
<td>70</td>
</tr>
<tr>
<td>Control, and Concerned K-F-D's</td>
<td></td>
</tr>
<tr>
<td>9. Mean Hit-Rates of the Four Judge Groups in Differentiating</td>
<td>72</td>
</tr>
<tr>
<td>Between Abusing versus Control versus Concerned K-F-D's</td>
<td></td>
</tr>
<tr>
<td>10. Two Measures of Inter-Judge Agreement for the Four</td>
<td>74</td>
</tr>
<tr>
<td>Groups of Judges Using K-F-D's</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF TABLES CONTINUED

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Pre- and Post-Training Hit-Rates for Judge Group Two Using K-F-D's to Differentiate Three Mothers' Groups</td>
<td>76</td>
</tr>
<tr>
<td>12. Mean Hit-Rates for the Differentiation of Abusing versus Control versus Concerned Mothers under Three Data Conditions</td>
<td>78</td>
</tr>
<tr>
<td>13. Results of Wilcoxon Test: Differences Between Rank Order of Hit-Rates for K-F-D Alone versus Social History Alone versus K-F-D + Social History</td>
<td>79</td>
</tr>
<tr>
<td>14. Mean Hit-Rates of Four Groups of Judges Under Three Data Conditions for the Differentiation of Abusing versus Control versus Concerned Mothers</td>
<td>81</td>
</tr>
<tr>
<td>15. Summary of Chi-Square Analysis for Judgments of Abusing versus Control versus Concerned Group Membership Based on Poor, Average, and Good Quality K-F-D's - Judge Number 11</td>
<td>84</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

FIGURE

1. Mean hit-rates as a function of judge group ($J_1$-$J_4$) and data condition (K-F-D versus SH versus K-F-D + SH) .................................................. 82
CHAPTER I
INTRODUCTION

In recent years there has been growing public and professional concern about the number of children who are subjected to physical abuse by their caretakers. While it is difficult to ascertain precisely the extent of the problem, estimates suggest that the annual nation-wide count in the United States is on the order of 375 cases for every one million children in the population (Cohen & Sussman, 1975). This means that roughly 1.5% of children in the U.S. are victims of child abuse each year.

A major thrust of the efforts to deal with this tragic problem has been the search for an effective assessment technique with which at-risk families can be identified. The Kinetic Family Drawing (K-F-D), introduced by Burns and Kaufman (1970), is one instrument which has been proposed for use in this fashion (Schornstein & Derr, 1977). These authors are convinced not only of the K-F-D's clinical validity in the hands of skilled professionals, but also of its potential utility as a screening device for use by nonclinical field workers. To date, these premises have not been exposed to experimental evaluation. The empirical validity of the Kinetic Family Drawing as an assessment tool for child-abuse evaluations, and the parameters which affect the accuracy of clinical judgments based on the drawings, were the subjects of this study.

Assessment and Child Abuse - An Overview

Scarcely 20 years have passed since Kempe, Silverman, Steele, Droegemueller, and Silver's (1962) milestone study of the phenomenon
which those authors dramatically labeled "the battered child syndrome."
In the ensuing two decades, considerable professional attention has
been focused on the problem of child abuse, with research efforts
directed toward a variety of issues. The published literature includes
studies of the following: the characteristics of abusing parents
(Blumberg, 1974; Paulson, Afifi, Thomas, & Chaleff, 1974; Spinetta
& Rigler, 1972; Steele, in Kempe & Helfer, 1980); the sociological and
situational factors that often contribute to abuse (Gelles, 1979; Gil,
1979; Straus, in Kempe & Helfer, 1980); and, to a lesser extent, the
elements of maladaptive family interactions which play a part (Gil,
1979; Polansky, Chalmers, Buttonweiser, & Williams, 1978).

It is possible to state with some confidence that the family most
at risk to be involved in child abuse is one characterized by some or
all of the following problems: (1) "deviance in the family structure,"
i.e., a single parent situation, usually mother only (Gil, 1979);
(2) if two parent figures are in the home, there is marital strife or
instability, such that the child may become a target for displaced
aggression (Bennie & Sclare, 1969; Gil, 1979); (3) the interactions
between caretaker and child are disturbed or negatively-toned, even for
"neutral" activities such as meal-time, play, and bed-time routines
(Polansky et al., 1978); (4) one or both of the parents experienced
childhood abuse or neglect (Schneider, Helfer, & Hoffmeister, in Kempe
& Helfer, 1980; Steele, in Kempe & Helfer, 1980); (5) the targeted
child is under the age of four (Bennie & Sclare, 1969; Gelles, in Gil,
1979). With respect to the factor of socioeconomic class, the research
findings are somewhat equivocal. Jacobson and Straker (1979) as well
as Korbin (in Kempe & Helfer, 1980) concluded, on the basis of their
reviews of the literature, that child abuse is a classless phenomenon. On the other hand, Pelton (1978) strongly disputed this position, and cited another series of studies which found a higher incidence of abuse among poorer, less educated parents. Nonetheless, whether the data are classified as "known," suspected, or still being hotly debated, there is considerable information being accumulated on the subject of child abuse.

Despite this amassing of scholarly effort, the development of a solid assessment instrument has remained an elusive goal, according to many authors. Blain, Berger, Lewis, and Goldstein (1981) noted that "... at present, we have developed little in the way of an assessment technology that would help us in this regard" (p. 667). Milner and Wimberley (1979) claimed that "... there is a paucity of carefully devised measures for assessing these characteristics [of abusing parents]" (p. 95). A complicating factor is, of course, the fact that abusing and neglectful parents are often uncooperative and inclined to resist intensive clinical or experimental scrutiny. Moreover, even when these parents do submit to testing, traditional kinds of assessment devices may not be useful or valid. Schneider et al. (in Kemper & Helfer, 1980) argued, for example, that the validity of self-report inventories might be jeopardized by an over-learned response tendency on the part of this population to answer questions in an inconsistent and evasive manner. Additionally, these authors pointed to factors such as low educational level, which could affect the parents' ability to comprehend and respond appropriately to verbal materials. Thus, Schneider et al. expressed reservations about, and counseled caution in the use and interpretation of the various questionnaire-type screening devices
which have been proposed (Milner & Wimberley, 1979; Paulson, Schwemer, Afifi, & Bendel, 1977; Paulson, Schwemer, & Bendel, 1976). Wright (1976) contended that the veracity of parents' responses may be influenced by the parents' desire to present themselves in the best possible light. The comments of these authors suggest, then, that the assessment devices which are employed ought, perhaps, to possess certain specific characteristics.

Following Wright's (1976) line of argument, Blain et al. (1981) suggested that a disguised, unobtrusive measure should be considered. Milner and Wimberley (1979) expressed the opinion that any technique for assessing abusing parents ought to be simply and easily administered. Their position is based on the notion that field caseworkers who lack extensive formal training in psychological assessment could, nevertheless, be trained to do basic screening. Shorkey (1978) agreed that complicated instruments which require lengthy testing sessions have limited utility. He added that the measure used ought to relate to basic characteristics of abusers and aspects of the abuse situation which have been identified in descriptive studies. Schneider et al. (in Kempe & Helfer, 1980) suggested that the ideal instrument would be one which could simultaneously tap multiple factors, such as the child's characteristics, parents' characteristics, interactional features, and stress factors.

An instrument which has the various qualities delineated by these experts, and which is demonstrably useful in the clinical assessment of abusing parents, would obviously be invaluable. It has been contended by Schornstein and Derr (1977) that the K-F-D (Burns & Kaufman, 1970) is one instrument which "fits the bill." A review of this test's
development, previous research findings, and the rationale for its proposed use in child-abuse evaluations follows.

**The Kinetic Family Drawing – Theory and Research**

The K-F-D was first introduced in 1970 (Burns & Kaufman), as an assessment technique to be used primarily with children. The technique involves presenting the child with a drawing task, with these instructions:

- Draw a picture of everyone in your family, including you, doing something. Try to draw whole people, not cartoons or stick people. Remember, make everyone doing something—some kind of action. (pp. 19-20)

Human figure-drawing tasks of one form or another had, of course, enjoyed widespread clinical use for many years at the point of the K-F-D’s entry into the field (Hammer, 1958; Machover, 1949). In fact, there was already a family drawing technique available (Hulse, 1951, 1952). Hulse had argued that children’s emotional problems stem, in part, from disturbed family relations. He postulated that the child’s drawing of his family could reveal important information about the child’s perceptions of the conflicts. Burns and Kaufman (1970) agreed with this view. They devised their instrument to serve as a means of evaluating the child’s sense of self within the family constellation, as well as his/her perceptions of other family members and interactions between them. Thus, the personal and interpersonal dynamics of the child could be examined. The unique feature of the K-F-D, and one which, according to Burns and Kaufman, contributes to its greater effectiveness, is the element of movement. They argued:

The approach of using kinetic (action) instructions, i.e., asking the child to produce a drawing where figures were moving or doing something, has been found to produce much more valid and dynamic material in the attempt to understand the psychopathology of children in the family setting. (pp. 17-18)
Burns and Kaufman believed that the addition of movement would help mobilize the child's feelings because the family is considered as a functioning, active unit.

After studying some ten thousand drawings, Burns and Kaufman (1972) reported the clinical significance of certain characteristics, actions, styles, and symbols found in the drawings. They described "characteristics" as static, structural qualities, such as pencil-erasures and hyperextended arms on figures (reflecting the need to control the environment). "Actions" in K-F-D's are movements of energy between people and objects. According to Burns and Kaufman (1970, 1972), actions reflect "fields of force" illustrating such things as competitive-strivings (ball-playing), anger (hammering), need for affection (eating), etc. "Styles" refer to certain features of the drawing that are considered indicative of defensiveness and emotional disturbance. Compartmentalization (intentionally separating family members by use of lining) and edge placement (placing all family members on the perimenter of the page in rectangular fashion) are examples of this drawing feature. Finally, "symbols" denote elements in the drawing which are attributed traditional psychoanalytic interpretations (e.g., beds are associated with sexual themes).

Professional response to Burns and Kaufman's (1970, 1972) work has been somewhat mixed. Gerstein (in Buros, 1978) was sharply critical, complains that "... the authors provide sparse data, no empirical relationships, and little theoretical framework and explanation to guide the interpretive diagnostic task" (p. 883). Other authors (Myers, 1978; O'Brien & Patton, 1974) have bemoaned the absence of normative data in Burns and Kaufman's publications. Additionally, there
is the complaint that the authors of the test established the "validity" of the instrument by the case presentation method only. O'Brien and Patton (1974) opined that "this method . . . often shows the confounding effect that prior knowledge about the family has upon the interpretation" (p. 156).

In contrast to the negative reactions reflected in the above-cited comments, there has been a very positive response to the introduction of the K-F-D in some professional circles. Ames (1973) indicated that she finds the K-F-D "extremely helpful" in clinical practice, when utilized as a part of a comprehensive battery. In her introductory remarks to Burns and Kaufman's (1972) interpretive text, Ames wrote, "Seldom has a test shown itself to be able to tell us so much about a subject so quickly and so simply" (p. v). Sobel and Sobel (1976) observed that the K-F-D has achieved widespread use among child and adolescent psychologists. Still, these authors expressed the very pointed concern that ". . . there is practically no existing research on its [the K-F-D's] validity, reliability, predictability, or applicability to varying populations or subcultures" (p. 91).

A review of the literature reveals, in fact, only 14 published studies and another half dozen unpublished doctoral studies using the K-F-D. Some attempts have been made to develop an objective scoring procedure for the drawings (Myers, 1978; O'Brien & Patton, 1974). While some promising preliminary findings are reported by Myers (1978), the author concluded that ". . . the scoring procedure employed offers relatively poor clinical discrimination for the individual case and should be used only as a research tool from which more sensitive scoring approaches can be developed" (p. 364).
With respect to the issue of empirical validity, the research results are varied. McPhee and Wegner (1976) found, for example, that stylistic aspects of the K-F-D productions of disturbed children did not differ from those of an adjusted control group in the predicted fashion. Sobel and Sobel (1976) also had disappointing results when they compared the drawings of normal male adolescents with the K-F-D's of delinquent male adolescents. The drawings were significantly different on only three of 16 possible variables. Rhine's (1978) research revealed no significant differences between the K-F-D's of children who were rated high and low in adjustment, based on their scores from the California Test of Personality. Rhine concluded that Burns and Kaufman's assumptions regarding a relationship between kinetic family drawings and adjustment were not supported.

On the other hand, Levenberg (1975) had earlier examined the validity of the Kinetic Family Drawing as an indicator of psychopathology in children, with results indicating that the drawings of disturbed children could be reliably discriminated from those of normal children. According to Levenberg, "The K-F-D as a clinical tool appears to offer the clinician at least as much useful data on which to base diagnostic judgments as the DAP or Bender-Gestalt Tests" (p. 392). Myers (1978) also found that the K-F-D's of boys diagnosed as emotionally disturbed on the basis of a behavior checklist could be differentiated from the drawings of emotionally well-adjusted boys. Myers concluded that his results did "add support to the concurrent validity of the K-F-D" (p. 364). In a series of studies (Raskin & Baker, 1975; Raskin & Baker, 1977; Raskin & Bloom, 1979), it was demonstrated that the K-F-D is a useful instrument for providing information about the
emotional status of learning-disabled children. Reportedly, these children display feelings of being different and isolated within their family, body concerns, and sibling rivalry, all of which are indicated in their K-F-D's. Finally, Sims (1974) administered the K-F-D and the Family Relations Indicator (a standardized picture projection technique) to emotionally disturbed children. A scoring procedure was devised, in which each figure depicted on the K-F-D was scored as positive, negative, or neutral. A similar system was applied to the responses on the Family Relations Indicator. Sims reported a highly significant correlation between the two tests on the scores which pertained to the quality of the subjects' relations with parent figures. He concluded that the K-F-D is a valid technique for assessing disturbed family interaction. Sims's conclusions are both encouraging and pertinent to the objectives of the present study, since abusive parent-child interaction might be considered the epitome of disturbed family relations. This brings us to the rationale for proposing the K-F-D as a potential tool in child-abuse assessment.

As indicated earlier, Schornstein and Derr (1977) have asserted that the Kinetic Family Drawing appears to have significant value in child-abuse evaluations. Although the K-F-D was originally devised as a measure of a child's perception of family members and family interactions, it is assumed that, when administered to adults, the K-F-D taps the same relevant variables. Since, in Schornstein and Derr's (1977) view, child abuse is "intimately related to parents' perceptions of their children and each other, and the manner in which they interact" (p. 297), the K-F-D may be the assessment instrument of choice. Moreover, Schornstein and Derr conceived of the K-F-D as a means of
evaluating individual and family psychodynamics, and uncovering areas of family stress which may be related to the occurrence of child abuse. This interpretive framework is consistent with one of the more credible conceptual models of child abuse. As our understanding of the problem has grown more sophisticated, the notion that child abuse is a multidimensional phenomenon has received increasing acceptance. A model which considers a multiplicity of factors, including individual psychopathology, intrafamilial tensions and conflict, and external environmental stresses has been proposed by Gelles (in Gil, 1979), Gil (1979), and others. It is precisely this interplay of variables which, according to Schornstein and Derr (1977), the K-F-D can reveal. The assumption is that individuals who have the particular "mix" of these factors which leads them to abuse their children, produce K-F-D's which are qualitatively different from the drawings of their nonabusive neighbors.

With respect to the characteristics of a "good" child-abuse assessment device, as delineated in discussion above, the K-F-D has many qualities to recommend its use. Schornstein and Derr (1977) noted, for example, that it is a nonverbal task, and hence, clients may be more at ease with it. Further, these authors indicated that parents are usually interested in the test, and may even offer spontaneous interpretations of their own or their spouses' drawings. The K-F-D has the added attractions of being a brief, easily administered test which is, according to Schornstein and Derr, difficult to "fake."

No empirical evidence is currently available regarding the utility of Kinetic Family Drawings in child-abuse assessment. Indeed, there is little in the way of information regarding the use of any projective drawing technique for this purpose. Blain et al. (1981) did report on
a study involving the assessment of abused children using the House-
Tree-Person technique. Their results indicated that individual test
indices, as well as clusters of indices, could be used effectively to
discriminate the drawings of abused children from those of nonabused-
disturbed children and well-adjusted children. This writer is aware of
no other relevant published research on this subject. In the absence
of a body of empirically derived data, we are left with what amount to
interesting, but unsubstantiated assumptions about the "clinical valid-
ity" of the K-F-D in child-abuse assessment. An experimental investi-
gation of the K-F-D's concurrent validity was the first purpose of the
present study.

Issues in Projective Test Validation

There can be little disagreement with the notion that responsible
psychodiagnostic practice requires attention to the validity of the
instruments used. As Meehl (1954) observed:

Regardless of one's theory about personality and regard-
less of one's choice of data, whether Rorschach, MMPI,
intuition, table, equation, or rational hypotheses devel-
oped in a case conference—the honest clinician cannot
avoid the question "Am I doing better than I could do by
flipping pennies?" (p. 136)

While the "clinical utility" of projective tests is often touted, many
authors have stressed the need for objective research evidence. MacFar-
lane and Tuddenham (in Anderson & Anderson, 1951) argued that users of
the tests have a professional responsibility to evaluate the "private
norms" they may have established. Interpretive skills can thereby be
sharpened and erroneous constructs discarded. Moreover, assumptions
which are empirically verified contribute to the body of knowledge
about personality functioning, and ought to be shared with colleagues
and students. Heidgerd (in Hammer, 1958) envisioned validation
research as a necessary step in preventing inappropriate use of the instruments. According to Heidgerd, this is especially critical as regards projective drawings:

Research plays an even more necessary role in drawing techniques than in any other method of personality evaluation since the use of drawing techniques superficially requires very little training. Complex scoring systems can be avoided. Popular magazines have capitalized on the apparent simplicity of drawing interpretation. Sidewalk psychologists freely interpret drawings. For these reasons, hypotheses and different interpretive theories are cheap. To differentiate the good from the bad is often not easy for even the serious and sincere psychologist.

The need for active and positive research integrated with the growing clinical use of drawings is therefore obvious. (p. 483)

Finally, Meehl (1954) offered this pointed comment:

The introduction of some special "clinical utility" as a surrogate for validation is inadmissible. If the clinical utility is really established and not merely proclaimed, it will have been established by procedures which have all the earmarks of an acceptable validation study. If not, it is a weasel phrase and we ought not to get by with it. (p. 138)

It has been proposed (Wanderer, 1969) that the clinical judgment method is one acceptable route for exploring the validity of projective tests. Wanderer argued that the method of correct matchings of judgments to criterion groups constitutes a measure of concurrent validity. The methodology requires that a criterion be established and held constant, as contrasted with the usual method of systematically manipulating some independent variable. The judges are the subjects and their judgments, which are based on test data produced by individuals who differ on the criterion, are the dependent variable. The extent to which the judges' judgments "match" or agree with the actual
distribution of the criterion reflects the efficacy of the instrument for assessing the criterion.

The literature is replete with studies of clinical judgment applied to various projective techniques. Contradictory results abound, and the general state of confusion in this research appears to be related to a number of important methodological problems. One major issue concerns the defining characteristics of the judges. Meehl (1954) complained that, in many cases, the level of experience and training of judges is not explicated. A wide range of skill may be represented in the catch-all category of "expert." Indeed, Harrower (1954) contended that much of the early failure to validate projective tests was related to the inexperience of the judges. Zimmer (1956) and Hamlin (1954) both viewed the judges' experience with the specific test as a critical factor. Moreover, according to Hamlin, familiarity with the population in question is an important consideration. He wrote:

We might assume that the "expertness" of a competent clinician consists of a high degree of skill in descriptive prediction for subjects and variables with which he is extensively and recently acquainted in a somewhat limited setting. A good state hospital psychologist can probably say much that is valid and useful about state hospital patients. Put him in a correctional institution or a clinic for children and his predictive validity may be limited for a time. Use him as a judge in one of our studies and his predictive validity may be limited indeed. (p. 237)

On a related point, Harrower (1954) observed that practitioners, exposed as they are to clinical populations most of the time, may become oversensitive to pathology. Thus, there may be little sense of what constitutes a "normal" protocol, or to what extent persons with "poor" protocols may, nonetheless, be entirely capable of adequate functioning.
There is a virtual chorus of critical commentary regarding the selection of criteria in clinical judgment studies. MacFarlane and Tuddenham (in Anderson & Anderson, 1951), Hamlin (1954), Harrower (1954), and others have stressed the necessity of choosing the criterion carefully. To begin with, there must be some reasonable basis for assuming that a relationship exists between test behavior and the criterion. Hamlin (1954) cautioned against expecting clinicians to perform "magic." Harrower (1954) rejected a methodology which requires clinicians to identify "good" pilots or salesmen on the basis of "good Rorschachs," for example, in the absence of any rationale. Oskamp (1962) and others have echoed this concern, and have pointed to the vast array of judgment studies which "failed" because of this procedural error.

According to several authors (Bialick & Hamlin, 1954; Holt, 1958; Oskamp, 1962), the variable to be judged ought to be one with which the judges are familiar and have had practical, everyday experience. Also important is the clarity with which the criterion is defined. Oskamp (1962) argued that the more objective the criterion is, the better. MacFarlane and Tuddenham (in Anderson & Anderson, 1951), Heidgerd (in Hammer, 1958), and Hamlin (1954) all called for the use of clear terms and constructs with adequate operational definitions. Hamlin suggested, for example, that if a criterion such as "adjustment" were the subject of study, the manner in which the external measure of adjustment was established should be described to judges in "precise and repetitive detail" (p. 236). Reliability of the criterion is another essential factor. Thus, studies which use as a criterion the pre-established judgment of another group of judges (typically psychiatrists who have generated a psychiatric diagnosis) often yield disappointing results.
The use of psychiatric diagnosis as a criterion has been strongly criticized by a number of authors (Goldberg, 1959; Hamlin, 1954; MacFarlane & Tuddenham, in Anderson & Anderson, 1951). The danger of contaminating criterion groups by including misdiagnosed subjects has been delineated by Heidgerd (in Hammer, 1958), Exner (1962), and Hammer (1969). Finally extraneous variables which impact on the criterion of test performance, or both, must be considered, since failure to do so may operate to lower the valid correlation between test and criterion (Zimmer, 1956). In this vein, Harrower (1954) was mindful of the need to explore the potential influence of sociocultural factors on projective test performance.

The final methodological issue concerns the elements of the judgment task itself. The most often repeated criticism of clinical judgment experiments is that clinicians are required to operate in ways which are atypical of their usual diagnostic practice. Specifically, it is noted that asking clinicians to perform "blind interpretations" of test protocols in the absence of any supportive test or case history data, places them at a major disadvantage (Goldberg, 1959; Hammer, 1958; Harrower, 1954). The degree of complexity of data on which the judgments are to be made is a related, important factor. In a series of studies (Bialick & Hamlin, 1954; Cummings, 1954; Hamlin, 1954), Hamlin and his associates demonstrated that the unit of material can be neither too complex (e.g., whole Rorschach protocols) nor too simplistic (e.g., single "signs"). Finally, the point has been made that judges participating in clinical judgment studies must not be asked to do too much. Cummings (1954) described the phenomenon of "rater demoralization and fatigue." According to this author, the condition is primarily a
function of complexity of the judgment task as well as size of the task. Cummings claimed that judges "report feelings of dissatisfaction and of being overwhelmed by having to retain and manipulate so many complex variables simultaneously, and having to continue to do so for what is perceived as a mountainous mass of protocols" (p. 246).

Clinical Judgment and Figure Drawings

With the above-detailed cautions borne in mind, we turn to a brief review of the literature concerning clinical judgment and figure drawings. The research on human figure-drawing tasks in general is addressed, since there is very little research on clinical judgment and the K-F-D per se. Pursuant to the discussion of the limitations and criticisms of clinical judgment research, it is noted that, to a large extent, we may be dealing with less than ideal, state-of-the-art research findings.

Results of research efforts in this area generally show that clinicians are able to discriminate between diagnostic criterion groups based on drawings, with moderate success (Albee & Hamlin, 1949; Goldberg, 1959; Plaut & Crannell, 1955). The figures reported for accuracy of diagnostic judgment, expressed as "hit-rates," range from a low of 64% (Hiler & Nesvig, 1965) to a high of 76% (Plaut & Crannell, 1955). Using the K-F-D, clinicians in Levenberg's (1975) study correctly discriminated between the drawings of disturbed and nondisturbed children in 72% of the cases.

Aside from diagnostic category, other criteria which have been investigated include ethnic background and socioeconomic class. Frisch and Handler (1967) found that judges could detect differences in the figure drawings of black and white subjects. Frisch and Handler
reported that the drawings of black junior high school students were characterized by greater emphasis on the treatment of the hair area, compared with the drawings of white students of similar economic background. These authors argued that care must be taken in interpreting projective drawings obtained from black subjects, when the interpretations are based on normative data which exclude this group.

Adler (1971) investigated the effects of race and socioeconomic status on drawing performance. Subjects were black, white, and Hispanic adult psychiatric patients. Socioeconomic status was defined as follows: Group 1 - any college education or employment in white-collar, sales, small-business, or skilled labor; Group 2 - high school graduate or regular employment in semiskilled or unskilled occupation; Group 3 - generally unemployed, supported by welfare, etc. The subjects' drawings were scored on 32 variables. A factor analysis of the scoring categories yielded four factors which accounted for the variance in figure-drawing performances. The factors were as follows: formal accuracy of the drawing and degree to which the drawing was differentiated with regard to detail; size of figure and placement on the page; disturbed relationship between the parts of the figure, reflecting integration problems; and motor control exhibited in the execution of the drawing. A comparison of the drawings revealed no significant differences attributable to race on the first factor. However, it was found that subjects in the highest social-class group scored significantly higher than the other two social-class groups on accuracy and differentiation of the drawing. With regard to the second factor, results indicated that Hispanic subjects tended to produce more constricted drawings than white subjects. No other significant differences
were noted. On the third factor, dealing with problems in integrating body parts, the drawings of black and Hispanic subjects showed significantly more disturbed integration than white subjects. Moreover, both lower social-class groups scored significantly higher on this measure of deviance than the higher social-class group. There were no significant effects reported for race or socioeconomic class on the fourth factor.

Some data are available regarding the effects of situational stress and differential motivation on judged drawing performance (Exner, 1962; Graham, 1956; Handler & Reyher, 1964; Menks, 1973). The results do suggest that situationally-induced anxiety impacts on drawing productions. However, subjects were unable consciously to manipulate their performance to conceal certain traits, after being exposed to conditions designed to induce the desire to distort (Graham, 1956). The postcondition drawings of subjects were essentially unaltered by the interpolated experience of receiving several negative interpretations of their first drawings.

The research concerning judgment reliability indicates, for the most part, respectable degrees of inter- and intrajudge agreement. Beck and Bart (1970) reported an interjudge reliability coefficient of .91 when judges used an objective scale to rate proportionality of figure drawings. Lehner and Gunderson (1952) also found reliability coefficients in this range for judgments of various structural and content elements of drawings. These authors reported an average intrajudge consistency rating (judging the same drawings at one week intervals) of .90, while Oskamp (1962) also described "encouraging" intrajudge consistency results (no figures provided). Judges in Albee and
and Hamlin's (1949) study achieved an astonishing .977 interjudge agreement rate on a global adjustment rating of drawings. In contrast to these data, Watson (1967) reported an interjudge reliability coefficient of only .41 on judgments of diagnostic group membership. However, there is the possibility that faulty definition of criterion groups (based on psychiatric diagnosis) undermined the judges' efforts in Watson's study.

**Additional Considerations in Clinical Judgment Research**

Returning to the general problem of clinical judgment and test validation at this point, Goldberg's (1959) comments are noteworthy. This author contended that at least two phases are involved in the validation process. First, the kinds of judgments which can be accurately generated on the basis of test data must be established. The foregoing discussion speaks to this phase in relation to figure drawings. After this step, according to Goldberg, the judgment process should be analyzed to determine the "who and how" of judgment accuracy. That is, the degree of expertise necessary for optimum diagnostic acumen using the test should be examined experimentally, and the manner in which judges formulate their clinical inferences explored.

The issues raised by Goldberg are very relevant to the diagnostic problem focused upon in the present study. In the area of child-abuse assessment, the responsibility for identifying and making decisions about critical family situations is not the exclusive purvey of mental health clinicians. The process of formulating diagnostic impressions, however general, begins from the moment that an abusive situation is discovered. Every professional who has contact with the family has some hand in this process. Schornstein and Herr (1977) have proposed
that the K-F-D constitutes a potentially valuable tool to be utilized at the various stages of assessment. Thus, when Protective Service workers involved in the initial investigation experience doubt as to whether to proceed to legal action, the K-F-D could, perhaps, be utilized as a rapid screening device. Similarly, community agency personnel might use the test as they search for signs that a family has managed to heal itself, or perhaps, that rehabilitative efforts have broken down. The specific question then becomes this: If the K-F-D's are indeed valid indicators of risk situations, can they be utilized effectively by nonclinicians, or does accurate interpretation require special expertise? An examination of the research concerning the parameters of clinical judgment related to projective tests will follow.

**Experience and Training as Variables**

In Goldberg’s (1959) view, the comparison of skilled clinicians’ judging ability with the judgment performance of nonclinical judges is an important and appropriate subject for investigation. There is a body of literature dealing with this issue, and the results are, in some sense, rather surprising. Tolor (1955) is virtually alone in reporting that clinicians performed significantly better than nonclinicians. In this case, the task involved rendering judgments about children's popularity based on their human figure drawings. The vast majority of research has failed to show expected differences among the judges' rating abilities, as a function of professional status. In fact, a few studies have found that nonclinicians actually performed slightly, though not significantly, better than clinicians (Goldberg, 1959; Hiler & Nesvig, 1965). Other research has revealed a nonsignificant trend towards better performance by clinicians (Guinan & Hurley,
results using the K-F-D fall into this latter category.

When the variable being studied is degree of professional experience, with groups of more experienced clinicians compared against less experienced clinical judges, the results are again equivocal. Guinan and Hurley (1965) found no significant differences between the performance of Ph.D. clinicians and graduate student trainees on a judgment task using the Draw-A-Person Test. Schaeffer's (1964) results also revealed no significant differences between the ability of experienced psychologists and clinical trainees to judge adjustment from figure drawings. However, Schaeffer did report a positive, though insignificant, correlation between experience and accuracy. Like Schaeffer, Levenberg (1975) did not find a significant effect for experience in a judgment task using the K-F-D. In Levenberg's study, doctoral-level clinicians with a mean of 13 years of experience were compared with predoctoral interns. The experienced clinicians did have a higher hit-rate than the interns, but not significantly so. Furthermore, experience was positively correlated with ability to perform better than chance. Five out of five experienced clinicians achieved this distinction, while only two out of five interns were as successful.

Clear-cut evidence of superior performance by more experienced clinicians was reported by Hunt, Jones, and Hunt (1957) and Oskamp (1962), but neither of these studies involved projective test data. Hunt et al. found that more experienced judges displayed more inter-judge agreement with regard to judgments of Wechsler-Bellevue vocabulary responses. Oskamp (1962) reported that experienced clinicians were significantly more accurate than trainees in judging MMPI profiles.
Furthermore, the experienced clinicians obtained a higher rating on a measure of "appropriateness" of judgment. This dimension was determined by comparing the judges' estimates of confidence in their ratings, i.e., how probable they felt it was that their decisions were correct, with the actual percentage of correct decisions made. Ideally, the percentage of correct decisions would match the probability estimates made by the judges. For example, of all the decisions made with 60% confidence, about 60% should be correct. The appropriateness measure was thus defined as the mean deviation of the judge's accuracy scores from the ideal match. Oskamp noted that specific, recent experience with the test being utilized was more useful than general professional experience.

At least two studies have demonstrated superior judgment performance by less experienced clinical judges. Stricker (1967) presented three groups of judges with the task of differentiating between the figure drawings of well-adjusted and disturbed adults. The judge groups were Ph.D. clinicians with a mean of 14 years of clinical experience, senior clinical graduate students with minimal practicum experience, and first-year clinical graduate students with no practical experience. All of the judges were provided with information from previous research regarding the types of characteristics which have been found to discriminate successfully drawings of normals from those of disturbed individuals. Stricker found that both student groups were more accurate than the experienced clinicians because the students were more willing to incorporate research data into their formulations. Goldberg (1959) reported that, using the Bender-Gestalt test, the diagnostic judgments of clinical trainees were more accurate than those of Ph.D. clinicians with a mean of six years of clinical experience. The judge who
performed most accurately was a trainee who had received intensive training in the use of the test.

The influence of specific task training has been formally examined by some authors: Oskamp (1962) and Goldberg (1968) both reported that when judges received interpolated training and feedback regarding their pretraining performance judging MMPI profiles, significant improvements were demonstrated in post-training accuracy rates. Murray and Deabler (1958) had judges attempt to sort human figure drawings of psychiatric patients into one of five possible diagnostic categories. There were 20 drawings in all, and after each series of five drawings was rated, the judges were informed of the correct diagnoses and encouraged to analyze their errors. The results indicated that in general, learning did occur, as reflected in increased accuracy over the final series of drawings. Murray and Deabler concluded that "Apparently, diagnosis on the basis of drawings can be taught" (p. 419). The authors added, however, that some judges learned more than others, while some did not profit at all from the training. Finally, Arkell (1976) studied the ability of five different nonclinical judge groups to differentiate between the drawings of adjusted and disturbed children. One group of judges received specific training in figure-drawing interpretations. The results showed that the trained judges achieved the highest accuracy rating, though they were not significantly better than untrained judges.

Clearly, the "last word" has not been offered regarding the effects of experience and training on clinical judgment. Certainly, with respect to the K-F-D, the lone contributions of Levenberg (1975) do not suffice to answer all the questions. Thus, the second major
focus of the present study was to explore the influence of experience and training on the accuracy of K-F-D judgments.

Elements of the Judgment Process

As noted earlier, Goldberg (1959) suggested that research efforts should be directed toward exploring the manner in which judges formulate their decisions. Much has been written about the phenomenon of "clinical intuition" as a modus operandi in the judgment process (Holt, 1958; Meehl, 1954, 1960). It is generally argued that this type of impressionistic cognition actually involves a process in which subtle cues are combined and hypotheses generated in a systematic way. According to Mahrer and Young (1961), the various cues are assigned some weighted value, as well as a positive or negative valence with regard to their relevance to the variable being judged. The efficacy with which accurate and meaningful judgments can be made by means of a global, intuitive impression may be dependent on both the nature of the judgment task and the stimuli being used.

With regard to the first issue, the nature of judgments being rendered, it seems reasonable to consider the distinction of general personality features versus more specific kinds of traits. Judgments of "adjustment," or broad diagnostic category might be thought of as more general, while decisions about the presence/absence of particular conflicts could be described as more specific. A review of the research concerning clinical judgment and figure drawings indicates that global judgments of adjustment can usually be made successfully (Albee & Hamlin, 1949; Plaut & Crannell, 1955). Levenberg's (1975) judges were able to differentiate between the K-F-D's of adjusted and disturbed children. In this case, judges were instructed to formulate
their judgments globally, using any system of evaluation they preferred.

In contrast to these findings, Fox, Davidson, Lighthall, Waite, and Sarason (1959) reported that the judge in their study was unable to differentiate between the drawings of children who had scored high and low on a criterion measure of anxiety, when a global method of judgment was used. These authors noted that the judge appeared to be overlooking, or incorrectly emphasizing various features of the drawings in forming her overall impressions. However, when the drawings were scored objectively, and the judge then presented with a descriptive list of six drawing characteristics which empirically differentiated the groups, she was able to discriminate between the drawings accurately. This research suggests that an intuitive method of formulating judgments may be effective for some kinds of clinical decisions but not for others.

The second issue concerns the cues to which judges are responding. As earlier remarks indicated, clinical judgment is not a magical activity, but rather, in some sense, a mathematical activity. Regarding figure drawings, Albee and Hamlin (1949) noted:

In the interpretation of drawings as projections, the various specific factors are usually gathered together by someone expert in clinical synthesis . . . . Clinical interpretation . . . brings together many specific elements from the drawings, perhaps emphasizing some elements and passing over others. (p. 389)

According to some authors, it is in the determination of which stimulus elements are relevant for attention that problems can sometimes arise. Hiler and Nesvig (1965) opined that most clinical psychologists rely on criteria which have been developed over years of unsystematic observation, for evaluating figure drawings. These authors contended that many such criteria are based on invalid hypotheses, and thus, can lead
to errors in judgment. In the study conducted by Hiler and Nesvig, judges were given the task of differentiating between the drawings of adult psychiatric patients and nonhospitalized "normal" controls. The judges were able to make the discrimination, but the overall accuracy rate was not impressive (64%). In the second phase of the study, judges were asked to state the criteria they had used in assigning each drawing to its category. The number of correct and incorrect decisions for each criterion or "sign" used was then tabulated. A prediction formula was developed on the basis of criteria which were found to discriminate correctly beyond the .1% level of significance. This formula, which included only 6 of the 32 different signs which clinicians claimed to use, was then applied to the discrimination task, with an accuracy rate of 79%. Hiler and Nesvig concluded that the predictive formula's improved efficacy was based on the elimination of sources of error and utilization of only the most valid indicators. The authors argued that an evaluation of criteria used in making clinical inferences and elimination of ones which result in error should lead to a sharpening of clinical judgment.

Unfortunately, some research suggests that even if judges are assisted in directing their attention to valid cues, they may not make appropriate use of the information. Stricker (1967) found, for example, that clinicians improved very little when they were given empirical data to consider. In his study, experienced clinicians were asked to differentiate between the drawings of adult psychiatric patients and controls. The judges were given a description of the drawing characteristics which Hiler and Nesvig (1965) had found significant in discriminating the two groups of drawings. It was suggested that judges
might find these research findings helpful in making their decisions.

The judges achieved an accuracy rate of 66%, as compared with the rate of 64% achieved by Hiler and Nesvig's judges. Stricker concluded that:

... the clinicians in this study seemed quite reluctant to abandon a diagnostic approach which they had used for many years, so that the data seemed to have minimal impact on their judgments. (p. 494)

Similar findings have been reported by Chapman and Chapman (1967, 1969). These authors identified the phenomenon called "illusory correlation," as a major factor in judgment error. Illusory correlation is defined as the report by an observer of a correlation between two events which, in reality, are not correlated, are correlated to a lesser extent than believed, or are correlated in the opposite direction than the one reported. A judge who acts on the belief, for example, that emphasis on eyes in figure drawings is correlated with the symptom of paranoid ideation, when, in fact, the data presented to him would suggest that no such valid correlation exists, would be committing systematic error based on illusory correlation. Chapman and Chapman have found that erroneous association of symptom statements with features of projective test protocols is resistant to alteration by the "facts." Their research suggests that this is so, because certain test signs have strong verbal associative connection to symptoms. The results of a series of experiments involving the DAP and Rorschach led Chapman and Chapman to speculate that many clinical interpretations of these tests have their roots in illusory correlation. Moreover, the authors concluded that illusory correlates can blind the observer to the presence of valid correlates of the symptom.
Relative Contributions of Data

A number of studies have focused on the process by which judges might combine data from several different tests or sources. The amount and type of data necessary for maximum success are variables which have been explored (Fisher, 1952; Golden, 1964; Kostlan, 1954; Sines, 1959). In general, the results suggest that amount of information is not related to accuracy of inference in any kind of simple, linear fashion. Golden (1964) had experienced clinicians complete personality questionnaires on psychiatric and nonpsychiatric hospital patients. The clinicians based their evaluations on various combinations of data which included identifying data alone, each of three tests alone (Rorschach, MMPI, and TAT), pairs of all possible test combinations, and all three tests combined. The criterion measure was the personality descriptions of the patients which had been prepared by therapists on the basis of case-history material. The results indicated that the judges' accuracy did not increase as a function of amount of data. Furthermore, Golden noted that interjudge reliability was negatively influenced by increasing amount of data, because the judges evaluated the additional information in different ways. Oskamp (1965) reported similar findings when raters made judgments about the personality characteristics of a single subject, using increasing amounts of historical/social information. Oskamp observed that, about midway in the four-stage presentation of data, judges reached a ceiling in their accuracy and did not improve beyond this level.

In contrast to these findings, Fisher (1952) indicated that judges were more accurate in judging personality traits of emotionally disturbed children when they used the Draw-A-Person test in combination
with a comprehensive test battery than when they attempted to make
personality judgments on the basis of the DAP alone.

Soskin (1954) observed that certain kinds of judgments could be
made more effectively on the basis of social history, while others were
more accurately rendered using the TAT. Naive and experienced judges
were asked to rate a single subject on 24 multiple-choice items reflect-
ing behavior which could be characteristic of the subject. The criterion
was Soskin's own assessment of the subject (an adult female), based on
extensive interviewing of the subject and her friends and family.
Judges were provided with demographic information and projective test
protocols. Results of the study suggested that judges could post-dict
certain attitudes and behaviors more accurately with history informa-
tion alone, while other postdictions were performed more accurately
using projective data. A striking finding was the tendency for both
naive and clinical expert judges to ascribe more maladjustment to the
subject when projective data were utilized.

Kostlan (1954) and Sines' (1959) also found that the type of
information being added to the data base seemed to be an important
factor. In Sines's study, clinicians were asked to evaluate data
obtained from male outpatient psychiatric clients. The judges' task
was to perform a series of Q-sorts of 97 items describing personality
and behavioral characteristics of the subjects, using various combina-
tions of test and interview data. The criterion measure was the Q-sort
performed by the patients' therapists. The data were presented to the
judges sequentially, with all judges receiving biographical data alone
as the initial data source. Rorschach, MMPI, and interview data were
then added to the data base, and different judges received the data in
different order. Sines reported that the addition of clinical interview material improved the accuracy of judgments consistently, while Rorschach and MMPI data had little effect. Sines also found that clinicians' perceptions of patients tended to crystallize early in the sequence. The judges appeared unwilling to change their descriptions after they had considered two or three kinds of data about a patient.

Kostlan (1954) presented judges with a checklist of personality statements describing five adult psychiatric patients. The validity criterion was the progress reports submitted by the patients' therapists. Judges rated the patients on the checklist, using differing combinations of data. The information sources were the Rorschach, MMPI, Sentence Completion Test, and Case History. Kostlan's results suggested that social history information was a critical element. Inferences based on a battery that did not include social history data were no more accurate than those based on basic demographic data alone. Interestingly, both Sines (1959) and Kostlan (1954) found that judgment decisions based on minimal demographic information were accurate at a level significantly greater than chance.

The criticism cited earlier regarding "blind" test-interpretation research constitutes a compelling argument for providing judges with data in addition to test protocols. However, the above-cited research suggests that careful consideration should be given to the separate and combined impact of each data source on the judgment rendered. In the interest of exploring this issue further, the relative contribution of K-F-D's versus social history information was examined in the present study.
The Confounding Influence of Aesthetics

As a final consideration, it is important to examine factors which can undermine the accuracy of clinical judgment and cast doubt on the validity of the assessment instrument. Rather than ignoring or rationalizing away problems with the tests, Harrower (1954) urged clinicians to assume the following posture:

... we should ... record systematically and study carefully all those cases where there are discrepancies between two sets of findings. Let us ... without guilt and embarrassment begin to sort out those occasions and conditions where, for various reasons, the projective material is not adequate to the task that has been imposed on it. (p. 294)

With respect to figure-drawing tests, a few studies have been done which point to the possibility that aesthetic quality acts as a confounding variable, and leads to erroneous judgments. Whitmyre (1953) was the first to examine this hypothesis. He collected figure drawings from psychiatric patients and "normal" veterans, matched on the variables of age, race, education, and IQ. The drawings were ranked by Ph.D. clinicians for level of adjustment reflected and artistic merit. Advanced art students also ranked the drawings according to degree of artistic excellence. Whitmyre reported a significant correlation between the psychologists' adjustment ratings assigned the drawings, and the artists' artistic value ratings. A highly significant correlation was also found between psychologists' adjustment ratings and psychologists' artistic merit ratings. Unfortunately, there was little or no relationship between the psychologists' ratings of adjustment based on drawings and actual behavioral criteria of adjustment. Whitmyre concluded that clinicians are actually responding to the artistic merits of drawings when they attempt to assess adjustment.
Sherman (1958b) had commercial artists rate the drawings of psychiatric patients and nonpatient controls for artistic quality, and in terms of whether good or poor drawing ability was reflected. Experienced clinicians also judged the drawings and attempted to discriminate between the patient and nonpatient groups. Again, the results indicated a significant relationship between judgments of artistic merit and ratings of adjustment. The clinicians were unable to differentiate accurately between the normal and disturbed groups, and Sherman attributed this failure to "... the extraneous influence of their [the drawings'] artistic quality which seemingly obscured any projective personality characteristics they may have possessed" (p. 339). Sherman argued that "... the determining effect that the variable of drawing ability seemingly has on the judgments of drawings" (p. 340) calls into question the validity of figure drawings as a projective technique. In another study, Sherman (1958a) found that psychologists' ratings of drawings on a scale measuring sexual differentiation were actually highly influenced by the artistic quality of the drawings.

Levy, Lomax, and Minsky (1963) attempted to refine these earlier research efforts by offering a definition of artistic merit in terms of the "proportional accuracy" of the figure, i.e., the extent to which life-like ratios among body parts are reflected. In their study, drawings which had predetermined proportional accuracy scores were judged by clinicians on the dimensions of overall adjustment, artistic merit, intelligence, dependency, aggression, and sexual difficulties. The results indicated that all but aggression were strongly correlated with proportional accuracy ratings. The authors concluded that "Whether in or out of awareness, the clinician is responding to the proportional
accuracy of the human figures he interprets" (p. 511).

In opposition to the conclusions drawn by the authors cited above, Holzberg and Wexler (1950) claimed that "It is accepted that [figure] drawings are not influenced to any substantial degree by school instruction in drawing or manual skill" (p. 344). Levy (in Abt & Beller, 1950) also voiced the opinion that artistic skill was not a significant source of variance. Hammer (1958) contended that artistic training and issues of aesthetic quality are not relevant matters of concern. Hammer likened art training to handwriting instruction in school, noting that, in the case of the latter, the uniformity of cursive style which is taught early soon drops out. Thus, according to Hammer, "Art training, by its very nature, favors free expression more than does handwriting training, and as such it probably contaminates projective drawing interpretation even less than handwriting training does handwriting analysis" (p. 50).

"Finally, there are at least a couple of studies which suggest that everybody's wrong." The results of the studies by Whitmyre (1953) and Sherman (1958a, 1958b) indicate that, while psychologists' judgments of adjustment were not related to actual adjustment, neither were artistic merit ratings. However, there are some contrasting findings which imply that the artistic quality of the drawing is related to personality characteristics of the drawer. For example, Eisenman and Smith (1966) reported that subjects who scored higher on a measure of authoritarianism and cognitive rigidity produced more barren, unelaborated, less creative figure drawings and expended less effort on the task than subjects who were less authoritarian. McPhee and Wegner (1976) found that disturbed children spent less time and effort producing K-F-D's
than did well-adjusted children. Furthermore, the adjusted children's drawings were far more elaborate and superior in technical quality.

Once again, the literature review leads us to the conclusion that definitive answers are not yet available with respect to this issue, and more research seems indicated.

To Summarize

While considerable attention has been focused on the growing problem of child abuse, the development of an effective screening device for at-risk families remains an unattained goal. Research regarding psychological assessment techniques for evaluating abusive parents has helped delineate the necessary characteristics of a good assessment tool. The K-F-D has been proposed as one "instrument of choice." This contention has not been investigated experimentally, but the K-F-D has been researched for use in other ways. It was argued that a clinical judgment format appears to be a viable method for investigating the empirical validity of projective measures. Methodological considerations, such as the issues of judge group selection, criterion definition, and task composition, were noted. Literature was reviewed which suggested that, in the second phase of test validation, the judgment process should be analyzed to determine the "who and how" of judgment accuracy. The question of degree of expertise necessary to use the Kinetic Family Drawing effectively was posed. An examination of research involving clinical judgment and figure drawings suggested that variables such as experience and training have impact on the accuracy of clinical judgment. However, it was argued that clarity regarding the issues involved is lacking. The relative contribution of various types of data to the judgment process was raised as a factor to be
explored. Finally, the possible confounding effects of aesthetics were considered, and it was noted that this seems to be yet another area where resolution of contradictory views is lacking.

Statement of the Problem and Hypotheses

This study was concerned with the problem of evaluating a projective test which has been proposed for use in identifying child-abusing parents. Specifically, the study was designed to investigate the concurrent validity of the K-F-D and to explore the variables which may affect the accuracy of clinical judgments based on the drawings. The external criterion was the group membership of mothers who produced the drawings, i.e., Abusing versus Control versus Concerned mothers. The independent variables were (1) the experience and training of the judges, i.e., abuse experience-assessment expertise versus abuse experience-no assessment expertise versus no abuse experience-assessment expertise versus no abuse experience-no assessment expertise; (2) the provision of specific task training to one group of judges; and (3) the amount of data available, i.e., social history versus no social history. The aesthetic quality of the drawings, i.e., good versus average versus poor, was also examined as a variable. The dependent variable was the ability of judges to differentiate the Abusing, Control, and Concerned mothers' drawings, expressed in terms of judgment accuracy.

This research is significant because it dealt with issues in the clinical assessment of child-abusing parents. Blain et al. (1981), Milner and Wimberley (1979), and others have highlighted the need for development of an effective assessment technology in this area. The importance of establishing empirical support for clinical practice was noted by Meehl (1954), and Heidgerd (in Hammer, 1958) placed particular
emphasis on the need for validation research involving projective
drawing techniques. The K-F-D is being used, and it has been argued
that the instrument has clinical validity (Schornstein & Derr, 1977).
This study exposed this claim to experimental scrutiny. Goldberg (1959)
has argued that the comparison of skilled clinicians' judging ability
with the judgment performance of nonclinical judges is an important
step in the process of validating an assessment instrument. The inves-
tigation of the judges' accuracy as related to their experience and
expertise, which was conducted in this study, is of significance in
this regard. Finally, pursuant to Golden's (1964) suggestion that the
inferential process itself should be investigated, the present study
attempted to explore some of the mechanics of clinical judgment. The
relative contribution of different kinds of data was examined, and the
possible confounding influence of aesthetics was considered.

The hypotheses investigated in this study were as follows:

1. Based on the arguments proposed by Schornstein and Derr (1977),
and on the findings of Levenberg (1975), it was hypothesized that the
K-F-D's of child-abusing mothers could be accurately differentiated
from those of nonabusing mothers.

1a. As a corollary hypothesis of (1), it was hypothesized that a
further differentiation could be made between the K-F-D's of Abusing
mothers, Control mothers, and Concerned mothers.

2. Based on the findings of Tolor (1955), Guinan and Hurley (1965),
and Levenberg (1975), it was hypothesized that judges with both clinical
experience and expertise in assessment techniques would achieve the
highest judgment accuracy levels.

2a. As a corollary hypothesis of (2), it was hypothesized that
judges who had neither clinical experience nor expertise in assessment techniques would achieve the lowest judgment accuracy levels.

2b. As a further corollary hypothesis, it was hypothesized that judges in the two groups with experience but no assessment expertise and assessment expertise but no experience would achieve accuracy levels which fell between those of the expert group and those of the nonexperienced assessment-naïve group.

3. Based on the findings of Murray and Deabler (1958) and Goldberg (1968), it was hypothesized that judges who had field experience with abusing clients but no assessment expertise would improve their accuracy after receiving training in drawing interpretation.

4. Based on the findings of Fisher (1952) and Kostlan (1954), it was hypothesized that judgments based on drawings plus social history would be more accurate than judgments based on drawings alone.

5. Finally, pursuant to the research conducted by Whitmyre (1953), McPhee and Wegner (1976), and others, the relationship between the artistic quality of the drawings and judgments of mothers' group membership, and between artistic quality and actual group membership of mothers was investigated.
CHAPTER II

METHOD

This study was conducted in four phases. In the first phase, K-F-D's and background information were obtained for three groups of mothers. Next, the drawings and social history data were evaluated by four groups of judges. Third, one group of judges received training regarding the interpretation of the K-F-D's, following which they rejudged the data. Finally, the aesthetic quality of the drawings was rated by three professional artists, and the drawings were scored using a scoring guide.

Definitions of the Mothers' Groups

The K-F-D's and social history data utilized in this study were generated by three different groups of mothers. Each group consisted of 21 women, for a total of 63. All of the women had at least one child between the ages of one month and six years. The mothers ranged in age from 18 years to 41 years. There were 40 black women, 21 white women, and 2 Hispanic women. The three groups were defined as follows:

(1) Mothers' Group One — Mothers who had been adjudicated in the Wayne County Probate Court - Juvenile Division on a petition alleging that their children had been abused. The petitions were filed against these women as a result of a referral from a hospital where the children had been treated for suspected abuse-related injuries. These mothers constituted the Abusing group.

(2) Mothers' Group Two — Mothers who were waiting for their children to receive routine checkups or treatment for minor ailments at the Outpatient Clinic of Children's Hospital of Michigan. These
mothers constituted the Control group.

(3) Mothers' Group Three — Mothers who were spending extended periods of time (including overnight) with their children, while the children received inpatient treatment at Children's Hospital for non-injury conditions. These mothers constituted the Concerned group.

Case Selection Criteria for the Mothers' Groups

Abusing Group. As noted, the Abusing mothers in this study were mothers who had had contact with the Juvenile Court. The definition of child abuse which is used in adjudication proceedings at the Wayne County Juvenile Court is that prescribed by state law. The relevant statute (Michigan Senate Bill No. 1974, 1981) reads as follows:

"Abuse" means any of the following actions:
(A) Physical injury which impairs or endangers the physical health of the minor.
(B) Sexual mistreatment of the minor.
(C) Impairment of the physical health or emotional well-being of the minor caused by cruel or inhumane treatment. (pp. 2-3)

The adjudicating phase of the Court proceedings involves the presentation of evidence relative to the allegation that child abuse (as defined above) has been committed. If the Court fails to find that the allegations are true, the case is dismissed at this phase. If the allegations are substantiated, however, the Court must proceed to hold a dispositional hearing, the purpose of which is to determine what actions are to be taken to rectify the abusive situation.

It is Court policy that parents appearing before the Court on allegations of child abuse are referred for evaluation to the Wayne County Clinic for Child Study, the community mental health clinic affiliated with the Court. The assessment is usually conducted after the petition has been substantiated, and prior to disposition of the case.
The Clinic then provides the Court with a report summarizing the clinical findings and recommending a disposition plan which would be in the best interests of the child. In rare instances, clients may be seen prior to the adjudication.

The Abusing group data utilized in this study were culled from the case records of the Clinic for Child Study. For purposes of this research, only those cases seen post-adjudication were selected. Further, only cases which involved physical injury to the child were included. This selection decision was based on two considerations. First, it was felt that physical injury offered the "strongest" possible operational definition of the abuse criterion. Second, Gil (1979) and others have suggested that the family dynamics involved in cases of sexual abuse are different from those of physical injury. This could introduce a complicating factor in drawing interpretation, and therefore, such cases were excluded. The additional selection criteria for Abusing group cases were as follows:

(1) Dates when the evaluations at the Clinic for Child Study had been conducted - only those cases on which the evaluations had been completed a minimum of four years prior to the conducting of this study were considered. This criterion was intended to minimize the possibility that the three clinicians participating as judges in the study would recall a case, and thus be able to identify a mother's group membership on the basis of previous knowledge.

(2) Identity of the examining clinician - only cases which had been assessed by one or two specific clinicians were selected. The clinicians were a white female social worker and a black female social worker who had been employed at the Clinic for Child Study when the cases had been referred. The K-F-D's used in the study had been obtained by the social
workers as a part of their routine evaluation of the clients.

(3) Quality of the drawings - only those cases which contained good-quality drawings, that is, in which the drawings had not been torn or marred, were considered. Additionally, only drawings done in pencil and with appropriate labelling were selected.

From the cases reviewed which met all of the above criteria, 11 cases which had been assessed by the white social worker and 10 cases which had been assessed by the black social worker were randomly selected for final inclusion in this study.

**Control and Concerned Groups.** As described above, the Control and Concerned mothers in this study were mothers whose children were receiving either outpatient or inpatient treatment at Children's Hospital. These cases were screened in order to ensure purity of the abuse/nonabuse classification. The Control and Concerned mothers were asked to indicate whether they had ever had any involvement with the Juvenile Court and/or Department of Social Services in relation to neglectful or abusive treatment of their children. Official Court and D.S.S. records were also checked. These procedures were intended to minimize the possibility that mothers classified as Control or Concerned mothers were actually abusers. The records check revealed previous contact with the authorities in only one case (a Control mother), and this mother was excluded from the study.

In all, protocols were obtained from 24 Concerned mothers and 28 Control mothers. Two of the Concerned mothers did not complete all of the experimental procedures. Of the remaining 22 protocols, the first 21 cases which had been collected were utilized in the study. There were six Control mothers who did not complete all of the procedures. As described
above, an additional Control case was rejected for use due to prior D.S.S. contact. The remaining 21 control cases were used in the study.

**Definition of the Judges' Groups**

Four groups of judges evaluated the K-F-D's and social history data generated by the mothers. Each group consisted of three judges for a total of 12. There were six males and six females. Three of the judges were black and nine were white. The groups were constituted as follows:

1. **Judge Group One** - Clinicians who were employed at the Wayne County Clinic for Child Study. The judges in this group had had both formal educational training in psychological/psychiatric assessment, and clinical experience with child-abusing clients. All had had training in the projective use of the K-F-D. Two of the three judges were Limited Licensed clinical psychologists (one Ph.D. and one M.A. level), and one was a Board Certified psychiatrist in the State of Michigan. Their years of experience working with abusing clients ranged from 2 to 20 years (mean = 13 years). Their years of experience in the diagnostic use of the K-F-D ranged from 4 to 16 years (mean = 11 years). These judges constituted the Abuse Experience - Assessment Expertise Group.

2. **Judge Group Two** - Family counselors who were employed by Parents-and-Children-Together (P.A.C.T.), an agency which provides ongoing social-work services to families involved in child abuse. The judges in this group had not had formal training in psychological assessment and, in particular, were naive as to the diagnostic use of the K-F-D. Two of the three judges had B.A. degrees in a social science field, and one had completed an M.A. in sociology. Each judge in this
group had had one year of field experience with child-abusing clientele. These judges constituted the Abuse Experience - No Assessment Expertise group.

(3) Judge Group Three - Clinical psychology graduate students who had received formal educational training in psychological assessment, with specific instruction in the diagnostic use of figure drawings and the K-F-D. These judges had not had clinical experience with child-abusing clients. One judge in this group had completed an M.A. in clinical psychology, and was in the sixth year of doctoral studies. The second and third judges had bachelors' degrees, and were in their third and fourth years of Ph.D. studies. These judges constituted the No Abuse Experience - Assessment Expertise group.

(4) Judge Group Four - Graduate students who were majoring in speech communications. These judges had neither relevant work experience with child-abusing clients nor training in psychological assessment. Two had B.A. degrees and were in their first years of graduate studies, and one was a third-year doctoral student who had already completed an M.A. These judges constituted the No Abuse Experience - No Assessment Expertise group.

Materials

The mothers' protocols in this study consisted of the K-F-D. (Burns & Kaufman, 1970) and a Social History Questionnaire (see Appendix A). The questionnaire was developed by the author, based on a number of considerations. First, the clinicians who routinely use the K-F-D report that knowledge about the number and ages of children in the home impacts on the interpretation of the drawing (Schornstein, Personal Communication, 1982). Further, a review of the literature
on child abuse suggests that abusing parents may differ from non-abusing parents on such variables as marital status and degree of marital satisfaction (Bennie & Scare, 1969; Gil, 1979); socioeconomic factors (Pelton; 1978); and history of being abused as children themselves (Steele, in Kempe & Helfer, 1980). Thus, items which tap these areas were included in the questionnaire. Finally, a number of "filler" items were added.

The K-F-D's constituted the test data which were evaluated by the judges. The judges also received a modified Social History Data Sheet (see Appendix B). This form was an amended version of the questionnaire filled out by the mothers. It excluded the mothers' names and their responses to the question about prior Court/D.S.S. contact.

The K-F-D Scoring Guide was used to score the mothers' drawings (see Appendix C). This guide was developed by the author, based on the system proposed by Myers (1978). The system rates 27 features of the drawing in all. Five of the features involve quantitative measures. These are the physical proximity of the mother to target child and mother to other figures (measured in inches), the heights of the mother and target child (measured in inches), and the ratio of mother to target child. The remaining 22 drawing characteristics are scored for their presence/absence in the drawing.

Procedures

The first phase of this study involved obtaining the K-F-D's and social history data on the three mothers' groups. As described above, the Abusing group data were selected from the case records of the Clinic for Child Study in accordance with the criteria cited. Case history information contained in the files was used to complete the
Social History Questionnaire for this group.

The K-F-D's and social history data for the Control and Concerned mothers were collected during a five-month period, from October, 1982, through February, 1983. The mothers were volunteers, and were not compensated for their participation. The examiners were two female graduate students, one black and one white. Half of the mothers were approached with the invitation to participate by the black examiner, and half by the white examiner, with the order randomly determined. The mothers were approached individually, and presented with a brief introduction into the nature of the study (see Appendix D). Informed consent was obtained from those mothers who agreed to participate (see Appendix E). The mothers were provided with a plain sheet of 8½" x 11" paper, and a number 2 pencil. The following instructions were delivered verbally:

Draw a picture of your family, with everyone in it, including yourself, doing something. Try to draw whole people, not cartoon or stick people. Remember, make everyone doing something, engaged in an activity.

All questions asked by the mothers regarding the execution of the drawing were answered in a standardized, nondirective fashion, i.e., with the response, "It's up to you." After the mothers had completed their drawings, they were instructed to label the figures by first name and indicate the ages of the figures on the drawing. The Social History Questionnaire was then verbally explained to the mothers. Following this, the mothers were asked to fill out the questionnaire, with assistance from the examiner if necessary.

The purpose of the study was intentionally described in vague terms, in the preliminary explanation offered to the mothers. Occasionally, mothers asked for additional information. For example, some
mothers inquired about what the drawing could reveal. A standard response was given to this question, as follows:

We are not certain exactly what the drawing can reveal. One of the reasons we are conducting this study is to investigate this question. As we indicated, we think the drawing can show something about family interactions.

Some mothers also inquired about the purpose of obtaining social history information. In response to this type of inquiry, mothers were told:

We will be looking at the drawings of a number of different mothers. We are interested in knowing the ways in which mothers may differ from each other, as well as things about their backgrounds which may be similar.

All of the mothers who voiced such concerns appeared to be satisfied with these explanations. When the mothers had completed both the K-F-D's and questionnaires, they were debriefed regarding the true purpose of the study and the various hypotheses under investigation.

After all the data were collected, the drawings and questionnaires were individually number-coded (with numbers assigned randomly), and any identifying information was removed from the protocols. The Social History Data Sheets to be presented to the judges were then prepared and identified with the appropriate code numbers.

In the second phase of the experiment, the drawings and background data were evaluated by the judges. The judges examined all of the protocols and did their evaluations during one sitting. They met together in their groups of three, but were asked to formulate their judgments independently, without conferring with one another. This procedure was previously used by Oskamp (1962, 1965). The judges in Group One (the clinicians) were treated to dinner by the author, as compensation for
their time. Judges in Group Two (the P.A.C.T. workers) received course-credit toward their accreditation as counselors. The judges in Groups Three and Four (the graduate student groups) were each paid $25 for their participation.

The first step in the judging procedures involved presentation of a brief explanation of the study (see Appendix F). As a part of this introduction, the judges in Groups Two and Four (the K-F-D naive judges) were given a written description of the rationale for the K-F-D (see Appendix G). Next, the judges' work experience and presence/absence of training in the use of the K-F-D were ascertained by having judges complete a brief form describing their backgrounds (see Appendix H). Finally, informed consent was obtained from the judges (see Appendix I). After these introductory procedures were completed, the mothers' data were presented to the judges.

Each judge evaluated the mothers' data three times during this phase of the experiment. During the first "round," judges were given either the drawings alone or the Social History Questionnaire alone, and asked to make a judgment regarding the group membership of the mothers. In addition, judges were asked to rate the degree of confidence they had in their judgments. Each judge was provided with individual Rating Forms on which judgments were to be recorded (see Appendices J and K). For purposes of distributing the data, the judges were randomly assigned to one of two working conditions. In each group, one judge worked alone with an entire data set (e.g., all of the social histories), while two judges shared the other data set (e.g., the K-F-D's), which was divided in half randomly. The paired judges exchanged their stacks when their judgments were completed.
After all of the judges in the group had completed the judgments for all 63 mothers, they were given the material they had not received the first time (i.e., either social history or K-F-D's) and asked to repeat the judging procedure using these data alone. In order to control for the possibility of an order effect, the order in which the two sets of data were presented was counterbalanced. Thus, for example, in Judge Group One, Judges One and Three received the drawings first and then social histories, while Judge Two received the social histories first and then the drawings. In Judge Group Two, Judges One and Three received the social histories first and then drawings, while Judge Two examined drawings alone and then social histories, etc.

For their third set of judgments, the judges received both the K-F-D and the social history information together for each mother. In this step, the combined protocols were randomly divided into thirds and distributed among the judges. The stacks were exchanged, so that each judge looked at all 63 combined protocols. The above procedures were presented to judges in the form of written instructions (see Appendix L).

In the third phase of the study, the judges in Group Two (Abuse Experience - No Assessment Expertise) participated in a two-hour training session on the interpretation of the K-F-D. The training was provided by two of the clinicians who were also expert judges in this study. Topics covered in the presentation included a description of the K-F-D's development, with brief discussion of projective drawing theory and research in general. The rationale for using the K-F-D in child-abuse evaluations was explicated, and the clinical significance of various drawing characteristics was discussed. Slides were shown giving examples of actual drawings produced by child-abusing clients,
and the instructors pointed out various noteworthy features of the drawings. However, the "sign-approach" to test interpretation was specifically eschewed. The importance of considering K-F-D test data in the context of known social history and other psychiatric and psychological findings was repeatedly stressed. The audience was cautioned that administration of the K-F-D was not a substitute for thorough clinical evaluation.

For purposes of this study, the training session took place after the Group Two judges had completed the initial judging session, but before the expert judge group had been exposed to the experimental drawings. After receiving the training, the Group Two judges rejudged the drawings and background data. A period of one week elapsed between the first judging session and the training. The post-training judging session took place four days after the training.

The fourth phase of the study involved obtaining ratings of the artistic quality of each drawing. Three professional artists evaluated the drawings and were paid $20 each for their participation. Two of the artists were white females and the third was a white male. In terms of professional credentials, all of the artists had completed some postgraduate training in fine arts. Each had had his/her work exhibited for sale at a public gallery, and two of the three were engaged in teaching fine-arts courses at a local university.

The artists were naive as to the true purpose of the study. They were told only that they would be examining the drawings of 63 different adult females, with the task of rating the drawings for artistic quality. Additionally, they were informed of the instructions which had been given to the women for completing the drawings. The artists
were instructed to rate each drawing in terms of artistic quality using the following scale:

1 - poor quality
2 - average quality
3 - good quality

Each artist was provided with written instructions and a form on which the ratings were to be recorded (see Appendix M).

Before beginning the actual ratings, six sample cases were examined and rated by the artists. The sample drawings were drawings which had been collected during Phase One of the project, but not included for use in the rest of the study. The artists were encouraged to discuss their ratings of these six samples among themselves, with the aim of establishing some consensus regarding rating criteria. The operational definition of "artistic quality" which was established on the basis of this discussion included technical considerations, such as proportional accuracy of the figures and general drawing ability, as well as more aesthetic features, such as composition or use of space on the page.

After these introductory procedures were completed, the artists were presented with the 63 experimental drawings. The drawings were randomly divided into thirds, and each artist worked independently, passing his/her stack of drawings on to the next person when the ratings were completed. Thus, each artist rated all 63 drawings. At the conclusion of the rating session, the artists were debriefed. The actual purpose of the study was explained and the various hypotheses were described.

Finally, as a check on the base-rate occurrence of certain scorable drawing characteristics, the K-F-D's were scored by the author and one other rater, using the K-F-D Scoring Guide. The second rater
was a volunteer and was not compensated for her participation. Again, the six sample drawings were utilized for discussion purposes, in order to establish consensual understanding of the scoring system. Following this, the raters scored the experimental drawings independently.

After all the data had been collected and analyzed, letters were prepared and mailed to those mothers and judges who had indicated a desire to be advised of the results (see Appendices N and O).

**Statistical Analyses**

In order to evaluate group differences among the mothers with regard to demographic and social history data, Chi-Square Analyses were performed for the nominal variables (Hays, 1981). One-way Analyses of Variance were utilized for the quantitative variables, with minimum significance set at \( \alpha = .05 \) (Winer, 1962). Further analysis, using the Newman-Keuls range test, was undertaken (Winer, 1962).

To assess whether the drawings of the three groups of mothers could be differentiated, Chi-Square Analyses were performed on the judgments rendered by each judge. These analyses yielded data concerning the "hit-rate" or proportion of correct decisions made by the judges. The degree of consistency or agreement among judges was assessed using Chi-Square Analyses.

The relative ability of different judge groups to differentiate the drawings was analyzed by using a Kruskal-Wallis Analysis of Variance with Ranks to assess differences in mean rank-ordering (Hays, 1981).

In order to investigate the effects of training on Judge Group Two's accuracy level, a Wilcoxon Matched Pairs Signed Ranks test was performed on the mean ranking of pre- and post-training accuracy levels.
The relative efficacy of K-F-D's alone versus Social History Questionnaires alone versus K-F-D and Social Histories together was analyzed using the Wilcoxon Matched Pairs Signed Ranks test to assess differences in judges' mean accuracy with the three data conditions.

The relationship between judges' judgments and artistic quality ratings was evaluated using Chi-Square Analyses and Cramer's V Correlations (Hays, 1981) for each judge. Chi-Square Analyses were also utilized to determine the relationship between actual group membership of the mothers and artistic ratings.

Finally, the K-F-D Scoring Guide data were analyzed. Inter-rater agreement in the use of the guide was calculated, using Cohen's Kappa tests of agreement (Cohen, 1960) for the nominal variables and Pearson Product-Moment Correlation Analyses for the quantitative variables. Differences in objective scores between the mothers' groups were analyzed, using a series of Chi-Square Analyses for the qualitative variables and one-way ANOVA's for the quantitative variables.
CHAPTER III

RESULTS

The presentation of results, to follow, is organized into three sections. In the first section, the sociodemographic data for the three mothers' criterion groups are considered.

In the second section, data which pertained in general to the question of the validity of the K-F-D are presented. First, the data regarding the K-F-D scoring system are detailed. Following this is a presentation of the data which dealt with judges' ability to differentiate the three groups of drawings (Hypotheses 1 and 1 (a)). These analyses involved collapsing across judge groups for such measures as accuracy and inter-judge reliability. Possible trends in the types of errors made are also considered for the judges as a whole.

The third set of results presented relates to the examination of factors which may have influenced the judges' ability to differentiate the drawings accurately. In this section, initial consideration is given to the possible influence of extraneous variables such as the judges' sex and race. Next, data regarding the judges' performance as a function of experience and expertise are presented (Hypotheses 2, 2 (a), and 2 (b)). Included here are analyses which compared the judge groups in terms of judgment accuracy, as well as analyses which dealt with group differences in inter-judge reliability. Following this, results which dealt with the effect of specific task training on judgment accuracy (Hypothesis 3) are reviewed. Data concerning the influence of presence/absence of social history information on judges' performance are then presented (Hypothesis 4). Analyses are presented
for the judges as a whole, as well as in terms of group differences with regard to this variable. Finally, the analyses which pertained to the relationship between ratings of artistic quality and judgments of mothers' group membership, and between artistic quality and actual group membership are reported.

Sociodemographic Data

A total of 22 variables relating to the demographic and background information collected from the mothers were analyzed, to determine whether differences existed between the three groups. The data were derived from the mothers' responses on the Social History Questionnaire. With regard to Occupation of the Head of the Household, 10 categories were utilized, as follows:

(1) housewife
(2) service worker (e.g., hairdresser, waitress)
(3) laborer
(4) transportation operator (e.g., cab and bus driver)
(5) operative (e.g., factory machine operator)
(6) skilled trade
(7) clerical
(8) sales
(9) management-administration
(10) technical and professional

The categories are those used by the U.S. Government to record census information. The mothers in the study were assigned to categories according to the definitions provided in a recent census publication (Statistical Abstract of the U.S., 1982-83).

A composite measure of the mothers' socioeconomic status was
developed, following the format suggested by Adler (1971). For purposes of the present study, three SES groups were generated, defined as follows:

SES Group 1 - mothers whose family income was derived from employment and who had either some college education, or head of household occupation was in categories 6-10.

SES Group 2 - mothers whose family income was derived from employment and who were either high school graduates, or head of household occupation was in categories 2-5.

SES Group 3 - mothers whose family income was derived from social security or public assistance.

Of the 22 variables analyzed, only four yielded significant differences between the Abusing, Control, and Concerned mothers. These variables were as follows: number of children residing in the home; number of children residing outside the home; contact with Court or Department of Social Services; and age of the mother. The Abusing mothers had significantly fewer children residing in the home ($\bar{X} = 1.24$) than either the Control mothers ($\bar{X} = 2.24$) or Concerned mothers ($\bar{X} = 2.71$). Concomitantly, the Abusing mothers had significantly more children residing outside the home ($\bar{X} = 1.00$) than either the Control mothers ($\bar{X} = .19$) or Concerned mother ($\bar{X} = 0.0$). In addition, 100% of the Abusing mothers had had contact with the Juvenile Court or D.S.S. regarding mistreatment of their children, while 0% of the mothers in the other two groups had had such contact. The latter finding is a direct result of data collection procedures. The results of the analyses on the two variables pertaining to number of children in and out of the home are also considered artifacts of the collection
procedures. The Juvenile Court routinely removes children from the custody of their parents temporarily while investigation of child abuse allegations is pending. It is suggested that this is the reason for the finding that Abusing mothers had fewer children in their custody, and more children residing outside the home (i.e., in foster care).

The results of the analyses on age of the mother are presented in Table 1. As Table 1 indicates, the one-way ANOVA for mean age of mothers yielded a highly significant difference among the three groups of mothers ($F(2, 60) = 10.35, p < .01$). Further analysis, using a Newman-Keuls test for differences between means, was undertaken. The results of this procedure, summarized in Table 1, indicate that the Abusing mothers were significantly younger than both Control and Concerned mothers ($p < .05$), and Control mothers were significantly younger than Concerned mothers ($p < .05$).

The analyses of selected social history variables considered important in the study of child abuse are summarized in Table 2. As noted earlier, there were no significant differences among the three groups of mothers on any of these variables. However, the general direction of the results is worthy of note. Inspection of Table 2 reveals, for example, that more than twice as many Concerned mothers as Abusing mothers were married. When the categories of Married and Living with Partner are combined, the total percentage of mothers in the Abusing group who reported having a significant cohabitation relationship is still only 42.8%. Thus, in contrast to the Concerned mothers, the majority of Abusing mothers were raising their children alone. Moreover, of the mothers who were married, more Abusing mothers
Table 1

Summary of Analyses for Mean Age of Mothers in Abusing (A), Control (C₁), and Concerned (C₂) Groups

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C₁</th>
<th>C₂</th>
<th>F</th>
<th>Newman-Keuls</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{X} ) Age in Years</td>
<td>23.14</td>
<td>26.62</td>
<td>30.48</td>
<td>10.35*</td>
<td>Abuse</td>
</tr>
</tbody>
</table>

*\( p < .01 \)

Note. For the Newman-Keuls procedure \( p < .05 \)
Table 2

Summary of Analyses of Selected Social History Variables for Three Groups of Mothers

<table>
<thead>
<tr>
<th>Marital Status</th>
<th>Abusing</th>
<th>Control</th>
<th>Concerned</th>
<th>X²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single</td>
<td>6 (28.6%)</td>
<td>8 (38.1%)</td>
<td>5 (23.8%)</td>
<td>17.09</td>
</tr>
<tr>
<td>Married</td>
<td>5 (23.8%)</td>
<td>9 (42.9%)</td>
<td>13 (61.9%)</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>1 (4.8%)</td>
<td>1 (4.8%)</td>
<td>3 (14.3%)</td>
<td></td>
</tr>
<tr>
<td>Separated</td>
<td>4 (19.0%)</td>
<td>2 (9.5%)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>1 (4.8%)</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Living w/Partner</td>
<td>4 (19.0%)</td>
<td>1 (4.8%)</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

| Marital Problemsᵃ       | 7 (33.3%) | 4 (19.0%) | 1 (4.8%) | 6.73 |
| Mother Abused as a Child| 4 (19.0%) | 2 (9.5%) | 0 | 4.42 |
| Spouse Abused as a Child| 0 | 1 (4.8%) | 1 (4.8%) | 1.25 |

Note. Numbers indicate the frequency count and percentages of mothers within each group who fell in a given social history category.

ᵃMothers who answered "No" to the question, "If you are married, do you consider your marriage a good one?"
than Control or Concerned mothers indicated that they did not feel their marriages were satisfying.

With regard to history of abuse as a child, the numbers for the total sample are small. Only six of 63 mothers, or 9.5% of the total sample, reported having been abused themselves as children. However, it is noteworthy that four of the six mothers with positive histories were Abusing mothers, while no mothers in the Concerned group reported such histories. The reverse is true in regard to spouse's history of being abused, although again, the numbers are very small.

**K-F-D Scoring**

The 63 K-F-D's produced by the mothers were scored by two independent raters on 22 qualitative variables (scored for presence/absence in the drawing) and 5 quantitative variables. Inter-rater agreement on the scoring of the nominal variables was calculated by first determining the proportions of judgments in agreement, using Chi-Square Analyses. The Cohen's Kappa Coefficient (a test of agreement which corrects for chance) was then calculated for each qualitative variable, according to the following formula:

\[ K = \frac{P_o - P_c}{1 - P_c} \]

where \( P_o \) = proportion of judgments in agreement and \( P_c \) = proportion of agreement due to chance.

The mean K-corrected agreement was 86%, with a range of 59% agreement (on the variable Shading) to 100% agreement (on the variables Ball and Folding Compartmentalism). With the single exception of Shading, all K-coefficients were significant at \( \alpha = .05 \).

The mean scores assigned by each rater were calculated for the five quantitative variables. All Rater 1 - Rater 2 mean differences
were nonsignificant at \( \alpha = .05 \). Pearson Product-Moment Correlation Coefficients were determined, and all exceeded the .90 level of agreement (range \( r = .90 \) to \( r = .99 \)).

The analyses indicate that the two raters in this study were able to utilize the K-F-D Scoring Guide with a generally high degree of reliability. The remaining analyses on the K-F-D scoring data were therefore performed using the scores assigned by only one rater (Rater 1).

The base-rate scores for each of the 27 K-F-D Scoring Guide variables are presented in Appendix P. The mothers' scores were analyzed to assess whether differences existed between the Abusing, Control, and Concerned groups. Significant differences were found on only two of the variables. These were Barriers Between Figures and Number of Household Members. The results of the Analysis on Barriers are summarized in Table 3. As Table 3 indicates, the Chi-Square Analysis of Barriers Between Figures yielded a highly significant difference among the three groups of mothers in presence of barriers \( (X^2 (2) = 9.8, p < .01) \). Inspection of the frequency data presented in Table 3 reveals that the vast majority (90.5\%) of the Concerned mothers placed objects between themselves and other figures in their drawings. In contrast, 76.2\% of the Control mothers' drawings possessed this feature, and only 47.6\% of the Abusing mothers depicted barriers.

Table 4 provides a summary of the Chi-Square Analysis on Number of Household Members. It can be seen that a significant difference was found among the three groups of mothers on Number of Household Members \( (X^2 (4) = 11.49, p < .05) \). The results indicate that the Concerned mothers were more accurate in depicting their family composition than
Table 3
Summary of Chi-Square Analysis for Presence of Barriers in the K-F-D's of Abusing, Control, and Concerned Mothers

<table>
<thead>
<tr>
<th></th>
<th>Abusing</th>
<th>Control</th>
<th>Concerned</th>
<th>$x^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10 (47.6%)</td>
<td>16 (76.2%)</td>
<td>19 (90.5%)</td>
<td>9.8*</td>
</tr>
</tbody>
</table>

*P < .01

Note. Numbers indicate the frequency count and percentages of mothers within each group whose K-F-D's depicted barriers.
Table 4
Summary of Chi-Square Analysis for Number of Household Members in the K-F-D's of Abusing, Control, and Concerned Mothers

<table>
<thead>
<tr>
<th></th>
<th>Abusing</th>
<th>Control</th>
<th>Concerned</th>
<th>(X^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure Omitted</td>
<td>6 (28.6%)</td>
<td>1 (4.8%)</td>
<td>2 (9.5%)</td>
<td>11.49*</td>
</tr>
<tr>
<td>Figure Added</td>
<td>6 (28.6%)</td>
<td>6 (28.6%)</td>
<td>1 (4.8%)</td>
<td></td>
</tr>
<tr>
<td>Correct Number</td>
<td>9 (42.9%)</td>
<td>14 (66.7%)</td>
<td>18 (85.7%)</td>
<td></td>
</tr>
</tbody>
</table>

*\(p < .05\)

Note. Numbers indicate the frequency count and percentages of mothers within each group whose K-F-D's possessed a given feature.
either the Control mothers or the Abusing mothers. Moreover, the three
groups differed in the types of errors they made. Abusing mothers
tended to omit family members from their drawings more than Control
or Concerned mothers. Abusing mothers and Control mothers added
figures to their drawings with equal frequency, and both groups did
this more often than the Concerned group.

The analysis of certain additional variables which Schornstein
and Derr (1977) consider to be possible indicators of pathological
family interaction are summarized in Table 5. As reported earlier,
there were no significant differences between the groups of mothers
on any of these variables. The results are presented for purposes of
reviewing trends.

The results presented in Table 5 indicate that Compartmentalism
of Figures was scored more frequently in the drawings of Concerned
mothers than in the K-F-D's of Abusing or Control mothers. Addition-
ally, more Concerned mothers erased in their drawings than Abusing or
Control mothers. Only two mothers in the total sample depicted figures
in dangerous positions, and these were both Concerned mothers. Evasion
and Encapsulation were found more often in the Abusing mothers' draw-
ings than in Control or Concerned mothers' drawings. Shading was used
fairly equally by the mothers, with Abusing and Concerned mothers show-
ing a slightly higher frequency than Control mothers on this feature.
Finally, there was no difference at all among the three groups on the
Distortion variable.

The mothers' scores on the K-F-D variables were further analyzed
to determine whether differences existed as a function of mothers' race. For purposes of these analyses, the two Hispanic women in the
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abusing</th>
<th>Control</th>
<th>Concerned</th>
<th>$X^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compartamentalism of Figures</td>
<td>3 (14.3%)</td>
<td>1 (4.8%)</td>
<td>4 (19.0%)</td>
<td>2.00</td>
</tr>
<tr>
<td>Encapsulation</td>
<td>7 (33.3%)</td>
<td>2 (9.5%)</td>
<td>4 (19.0%)</td>
<td>3.68</td>
</tr>
<tr>
<td>Pencil Erasures</td>
<td>12 (57.1%)</td>
<td>12 (57.1%)</td>
<td>16 (76.2%)</td>
<td>2.19</td>
</tr>
<tr>
<td>Shading</td>
<td>12 (57.1%)</td>
<td>9 (42.9%)</td>
<td>12 (57.1%)</td>
<td>1.15</td>
</tr>
<tr>
<td>Safety of Figures&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>2 (9.5%)</td>
<td>4.13</td>
</tr>
<tr>
<td>Distortion</td>
<td>2 (9.5%)</td>
<td>2 (9.5%)</td>
<td>2 (9.5%)</td>
<td>0</td>
</tr>
<tr>
<td>Evasion</td>
<td>16 (76.2%)</td>
<td>9 (42.9%)</td>
<td>11 (52.4%)</td>
<td>5.05</td>
</tr>
</tbody>
</table>

Note. Numbers indicate the frequency count and percentages of mothers within each group whose K-F-D's possessed a given feature.

<sup>a</sup>Mothers whose K-F-D's depicted figure(s) in dangerous position(s).
study were reclassified as "nonwhite." The data were then analyzed using White and Nonwhite as the two racial categories. Significant differences were found between White and Nonwhite mothers on only one variable. Two White mothers (9.5%) depicted figures in dangerous positions in their drawings, as contrasted with 0 Nonwhite mothers. This difference was significant \( X^2 (1) = 4.13, p < .05 \).

In order to determine whether the mothers' socioeconomic status had any effects on the K-F-D scoring, scores on the K-F-D variables were analyzed for the three SES groups defined above. There were three variables on which significant differences were found among the SES groups. These were Actions are Strange, Rotation, and Evasion. The results of the Analyses on these variables are summarized in Table 6. As Table 6 indicates, a significantly higher percentage of the mothers in SES Group 1 than in Groups 2 or 3 depicted actions that were strange, unreal, or unexpected \( X^2 (2) = 8.94, p < .01 \). Further, a significantly greater proportion of the SES Group 1 mothers rotated one or more figures in their drawings, relative to the proportion of mothers in SES Groups 2 and 3 whose drawings possessed this feature \( X^2 (2) = 6.34, p < .05 \). However, a significantly higher percentage of mothers in the middle SES Group demonstrated Evasion in their drawings, compared with the percentages of mothers in the lowest and highest SES Groups \( X^2 (2) = 7.28, p < .05 \).

The final analyses of the mothers' K-F-D scores assessed whether differences in the examiners who elicited the drawings had any effect. There were no significant differences between the drawings elicited by white examiners and those elicited by black examiners, in terms of the K-F-D scoring variables. Further, the drawings obtained by each of the
Table 6

Summary of Chi-Square Analyses for Actions are Strange, Rotation, and Evasion in the K-F-D's of Mothers in SES Groups 1, 2, and 3

<table>
<thead>
<tr>
<th></th>
<th>SES - 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>SES - 2&lt;sup&gt;b&lt;/sup&gt;</th>
<th>SES - 3&lt;sup&gt;c&lt;/sup&gt;</th>
<th>X&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actions are Strange</td>
<td>2 (40%)</td>
<td>0</td>
<td>3 (8.3%)</td>
<td>8.94**</td>
</tr>
<tr>
<td>Rotation</td>
<td>3 (60%)</td>
<td>7 (31.8%)</td>
<td>5 (13.9%)</td>
<td>6.34*</td>
</tr>
<tr>
<td>Evasion</td>
<td>0</td>
<td>14 (63.6%)</td>
<td>22 (61.1%)</td>
<td>7.28*</td>
</tr>
</tbody>
</table>

<sup>a</sup>Total number of mothers in SES Group 1 = 5 (7.9% of sample).

<sup>b</sup>Total number of mothers in SES Group 2 = 22 (34.9% of sample).

<sup>c</sup>Total number of mothers in SES Group 3 = 36 (57.1% of sample).
four examiners were compared in terms of objective features, and no consistent pattern emerged. The differences which were evident were randomly distributed across the examiners (see Appendices Q and R).

Judges' Ability to Differentiate the Drawings

In order to determine whether the K-F-D's of Abusing mothers could be differentiated from those of Nonabusing mothers (i.e., Control and Concerned groups collapsed), Chi-Square Analyses were performed on the K-F-D judgments rendered by each judge. These analyses yielded "hit-rates," or percentages of overall correct decisions, for each judge. In addition, the pattern of false positives (Nonabusing mothers who were incorrectly identified as Abusing) and false negatives (Abusing mothers who were incorrectly identified as Nonabusing) was calculated for each judge. Results of these analyses are summarized in Table 7.

Chance level discrimination for this two-group judgment is 50%. Using a Chi-Square Goodness of Fit test, it was determined that hit-rates of 60% or greater would be better than chance at the .05 level of significance. As Table 7 indicates, only one of the 12 judges was able to differentiate between the K-F-D's of Abusing and Nonabusing mothers. The hit-rates ranged from 39.7% to 61.9%, and the mean hit-rate for the judges was 52.4% (p > .05). The first hypothesis of this study predicted that the K-F-D's of Abusing mothers could be differentiated from those of Nonabusing mothers. The results do not support this hypothesis.

Inspection of Table 7 reveals that the mean false positive rate of judges' decisions was 26.4%, while the mean false negative rate was 21.2%. Eight of the 12 judges had false positive rates which exceeded their false negative rates. Thus, the errors which judges made were generally in the direction of overidentifying pathology.
Table 7
Judges' Hit-Rates and False Positive/False Negative Errors in Differentiating Between Abusing and Nonabusing Mothers

<table>
<thead>
<tr>
<th>Judge Number</th>
<th>Hit-Rate</th>
<th>False +</th>
<th>False -</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44.4%</td>
<td>28.6%</td>
<td>27.0%</td>
</tr>
<tr>
<td>2</td>
<td>39.7%</td>
<td>39.7%</td>
<td>20.6%</td>
</tr>
<tr>
<td>3</td>
<td>52.4%</td>
<td>23.9%</td>
<td>23.8%</td>
</tr>
<tr>
<td>4</td>
<td>49.2%</td>
<td>30.2%</td>
<td>20.6%</td>
</tr>
<tr>
<td>5</td>
<td>59.3%</td>
<td>17.5%</td>
<td>22.2%</td>
</tr>
<tr>
<td>6</td>
<td>55.5%</td>
<td>17.5%</td>
<td>27.0%</td>
</tr>
<tr>
<td>7</td>
<td>50.7%</td>
<td>34.9%</td>
<td>14.3%</td>
</tr>
<tr>
<td>8</td>
<td>53.9%</td>
<td>22.2%</td>
<td>23.8%</td>
</tr>
<tr>
<td>9</td>
<td>52.3%</td>
<td>33.3%</td>
<td>14.3%</td>
</tr>
<tr>
<td>10</td>
<td>50.8%</td>
<td>25.4%</td>
<td>23.8%</td>
</tr>
<tr>
<td>11</td>
<td>58.7%</td>
<td>28.6%</td>
<td>12.7%</td>
</tr>
<tr>
<td>12</td>
<td>61.9% *</td>
<td>14.3%</td>
<td>23.8%</td>
</tr>
</tbody>
</table>

\[
\bar{X} = 52.4\% \quad \bar{X} = 26.4\% \quad \bar{X} = 21.2\%
\]

*p < .05
Hypothesis 1 (a) predicted that a differentiation could be made between the K-F-D's of Abusing, Control, and Concerned mothers. Table 8 presents a summary of the results derived from the Chi-Square Analyses performed on judges' K-F-D judgments. For this three-group judgment, a hit-rate of 33.3% could be achieved by chance. A hit-rate of 44% or greater would be significantly better than chance discrimination at the .05 level of significance. The results presented in Table 8 indicate that none of the judges was able to differentiate between the three groups of drawings at a level significantly above chance. The hit-rates ranged from 20.5% to 38.1%, with a mean of 32.1% (p > .05). Thus, Hypothesis 1 (a) is not confirmed.

The inter-judge agreement for judgments of Abusing versus Control versus Concerned group membership was determined, using a Chi-Square Analysis. The proportion of judgments on which all 12 judges agreed was calculated. A more lenient measure, defined as the proportion of judgments on which at least 8/12 of the judges agreed, was also determined. The more conservative measure yielded an inter-judge reliability coefficient of only 1.6%. There was only one case in which all 12 judges agreed in their judgments of the mother's group membership, based on the K-F-D. With the standard of 8/12 judges in agreement, the inter-judge reliability was 28.6% (18/63 cases).

Factors Which Influenced Judges' Performance

Judges' race and sex. Analyses were undertaken to examine various factors which may have had impact on judges' ability to differentiate the drawings. First, the mean accuracy of black judges was compared with that of white judges, to determine whether race of the judge had a significant effect on performance. The mean hit-rate of the three
<table>
<thead>
<tr>
<th>Judge Number</th>
<th>Hit-Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.5%</td>
</tr>
<tr>
<td>2</td>
<td>26.9%</td>
</tr>
<tr>
<td>3</td>
<td>34.9%</td>
</tr>
<tr>
<td>4</td>
<td>33.3%</td>
</tr>
<tr>
<td>5</td>
<td>36.5%</td>
</tr>
<tr>
<td>6</td>
<td>30.1%</td>
</tr>
<tr>
<td>7</td>
<td>34.8%</td>
</tr>
<tr>
<td>8</td>
<td>30.1%</td>
</tr>
<tr>
<td>9</td>
<td>34.8%</td>
</tr>
<tr>
<td>10</td>
<td>27.0%</td>
</tr>
<tr>
<td>11</td>
<td>38.1%</td>
</tr>
<tr>
<td>12</td>
<td>38.1%</td>
</tr>
</tbody>
</table>

\[ \bar{X} = 32.1\% \]
black judges was 30.1% for the discrimination between Abusing, Control, and Concerned mothers. The nine white judges achieved a mean hit-rate of 32.8% on the differentiation between Abusing, Control, and Concerned mothers. A Kruskal-Wallis One-Way ANOVA with Ranks was performed on the rank ordering of judges. The results indicated that there were no significant differences between black and white judges on the ability to differentiate the K-F-D's.

The mean accuracy levels of male and female judges were calculated. The male judges attained a mean hit-rate of 30.7% for the differentiation between Abusing, Control, and Concerned mothers. The mean hit-rate for female judges was 33.5%. A Kruskal-Wallis One-Way ANOVA with Ranks yielded no significant differences between the male and female judges.

Abuse experience/assessment expertise. With regard to the effects of abuse experience and assessment expertise on judges' performances, Hypothesis 2 predicted that judges with both experience and expertise (i.e., Judge Group One) would achieve the highest accuracy levels. Hypothesis 2 (a) predicted that judges with neither experience nor assessment expertise (Judge Group Four) would achieve the lowest accuracy levels. The prediction of Hypothesis 2 (b) was that judges in Groups Two and Three would attain accuracy levels which fell between those of the expert group and the nonexperienced assessment-naïve group. Table 9 presents the mean hit-rates for the four groups of judges; using K-F-D's to differentiate between Abusing versus Control versus Concerned mothers.

The data in Table 9 indicate that judges in Judge Group Four achieved the highest mean accuracy levels. The expert judges (Judge Group One) had the lowest mean hit-rates in differentiating between
Table 9
Mean Hit-Rates of the Four Judge Groups in Differentiating Between Abusing Versus Control Versus Concerned K-F-D's

<table>
<thead>
<tr>
<th>Judge Group</th>
<th>Hit-Rate</th>
<th>Kruskal-Wallis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge Group 1</td>
<td>27.4%</td>
<td>$X^2 = 2.58$</td>
</tr>
<tr>
<td>(Clinicians)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge Group 2</td>
<td>33.3%</td>
<td></td>
</tr>
<tr>
<td>(P.A.C.T.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge Group 3</td>
<td>33.2%</td>
<td></td>
</tr>
<tr>
<td>(Clinical Students)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge Group 4</td>
<td>34.4%</td>
<td></td>
</tr>
<tr>
<td>(Speech Students)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Abusing versus Control versus Concerned K-F-D's. Judge Groups Two and Three achieved mean hit-rates which fell between those of Groups One and Four. Judge Group Two judges (P.A.C.T. counselors who had abuse experience but no assessment expertise) were slightly more accurate than the judges in Group Three (clinical psychology graduate students with assessment expertise but no abuse experience). A Kruskal-Wallis One-Way ANOVA with Ranks was performed on the group mean ranks. The results of this analysis indicated that the differences between the judges were not significant ($X^2 (3) = 2.58, p > .05$).

While no significant differences were found among the judge groups, the results fail to confirm the rank ordering predicted by Hypotheses 2 and 2 (a). In fact, a reverse trend is evident, with the naive group achieving the highest mean hit-rate and the expert group achieving the lowest mean hit-rate. However, Hypothesis 2 (b) is confirmed.

The inter-judge agreement on judgments of Abusing versus Control versus Concerned mothers was calculated for each judge group. The proportion of judgments on which all three judges in the group agreed was calculated. Additionally, a measure of reliability defined as the proportion of judgments on which at least 2/3 of the judges in the group agreed was determined. These data are presented in Table 10. It can be seen that the proportions of unanimous agreements were generally low for all four groups of judges (range 15.9% to 27.0%). With the standard of 2/3 judges in agreement, the reliability was much higher (range 79.4% to 92.1%). Judge Group Three had the highest inter-judge agreement. Judge Group Two had the next highest inter-judge reliability, and Groups One and Four demonstrated equally low levels of inter-judge agreement.
Table 10

<table>
<thead>
<tr>
<th>Judge Group</th>
<th>2/3 Judges Agree</th>
<th>3/3 Judges Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Clinicians)</td>
<td>50 (79.4%)</td>
<td>10 (15.9%)</td>
</tr>
<tr>
<td>2 (P.A.C.T.)</td>
<td>51 (81%)</td>
<td>14 (22.2%)</td>
</tr>
<tr>
<td>3 (Clinical Students)</td>
<td>58 (92.1%)</td>
<td>17 (27.0%)</td>
</tr>
<tr>
<td>4 (Speech Students)</td>
<td>50 (79.4%)</td>
<td>10 (15.9%)</td>
</tr>
</tbody>
</table>

Note. Numbers indicate frequency count and percentages of judgments in agreement, with total n = 63 for each judge group.
Training. Judges in Group Two (P.A.C.T. counselors) completed all of the judging procedures twice. Between their first and second judging sessions, these judges received training in the interpretation of the K-F-D. In order to assess the effects of this training on judgment accuracy, hit-rates for the pre- and post-training judgments were calculated. Table 11 presents the hit-rates for Judge Group Two's pre- and post-training judgments, using K-F-D's to differentiate between Abusing versus Control versus Concerned mothers. Inspection of Table 11 reveals that only one of the three judges improved as a result of training. The hit-rates for the other two judges actually declined post training. The results for the two-group discrimination between Abusing and Nonabusing K-F-D's are consistent with this finding, in that all three judges performed worse on this differentiation after receiving training. Further analysis, using the Wilcoxon Matched Pairs Signed Ranks test, indicated that the differences between the pre- and post-training performances of Judge Group Two were not significant. No significant differences were found between the post-training performance of Group Two judges and the performance of any other judge group.

Hypothesis 3 of this study predicted that judges who had abuse experience but no assessment expertise (Judge Group Two) would improve their accuracy after receiving training in K-F-D interpretation. This hypothesis is not confirmed.

Types of Data

It will be recalled that judges evaluated the mothers' data three times, rendering judgments regarding the mothers' group membership based on K-F-D's alone, Social History alone, and K-F-D plus Social History combined. Hit-rates were determined for each judge, under
Table 11
Pre- and Post-Training Hit-Rates for Judge Group
Two Using K-F-D's to Differentiate Three Mothers' Groups

<table>
<thead>
<tr>
<th>Judge Number</th>
<th>Pre-Training Hit-Rate</th>
<th>Post-Training Hit-Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>33.3%</td>
<td>27.0%</td>
</tr>
<tr>
<td>5</td>
<td>36.5%</td>
<td>28.6%</td>
</tr>
<tr>
<td>6</td>
<td>30.1%</td>
<td>45.9% *</td>
</tr>
<tr>
<td></td>
<td>( \bar{X} = 33.3% )</td>
<td>( \bar{X} = 33.8% )</td>
</tr>
</tbody>
</table>

\*p < .05
each data condition. Judges' hit-rates using the K-F-D alone were presented above. The hit-rates using Social History alone to differentiate between Abusing versus Control versus Concerned mothers ranged from 30.1% to 65.1%. Nine of the 12 judges achieved hit-rates which were significantly better than chance at \( \alpha = .05 \). Judges' hit-rates for the three-group discrimination using the K-F-D plus Social History combined ranged from 23.8% to 58.8%. Six of the 12 judges performed significantly better than chance \( (\alpha = .05) \) using the combined data.

The mean hit-rates for the three data conditions, collapsed across judge groups, are presented in Table 12. Inspection of Table 12 reveals that judgments based on Social History alone were the most accurate. Judgments based on K-F-D plus Social History were more accurate than judgments based on K-F-D's alone. The accuracy rates for the three data conditions were further analyzed, using Wilcoxon Matched Pairs Signed Ranks tests for differences between mean rank orders. The results of these analyses, summarized in Table 13, indicate that judgments based on Social History alone were significantly better than judgments based on K-F-D's alone (\( p < .01 \)). Judgments based on K-F-D plus Social History combined were also significantly more accurate than judgments based on K-F-D's alone (\( p < .01 \)). However, the superiority of judgments based on Social History alone versus judgments based on K-F-D plus Social History did not reach significance.

Hypothesis 4 of the study predicted that judgments based on drawings plus social history data would be more accurate than judgments based on drawings alone. Hypothesis 4 is confirmed.

Further analyses of the effects of data condition were undertaken. The results reported above point to the superiority of the social
Table 12
Mean Hit-Rates for the Differentiation of Abusing versus Control versus Concerned Mothers under Three Data Conditions

<table>
<thead>
<tr>
<th></th>
<th>Hit-Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-F-D Alone</td>
<td>32.1%</td>
</tr>
<tr>
<td>Social History Alone</td>
<td>47.9% *</td>
</tr>
<tr>
<td>K-F-D + Social History</td>
<td>43.0%</td>
</tr>
</tbody>
</table>

*p < .05

Note. Numbers indicate the mean hit-rates of all 12 judges, collapsed across judge groups.
Table 13
Results of Wilcoxon Test: Differences Between Rank Order of Hit-Rates for K-F-D Alone Versus Social History Alone versus K-F-D + Social History

<table>
<thead>
<tr>
<th>Data Condition</th>
<th>More Accurate Judgment</th>
<th>Wilcoxon Rank Order Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-F-D Alone versus Social History Alone</td>
<td>Social History Alone</td>
<td>( Z = -2.98^* )</td>
</tr>
<tr>
<td>K-F-D Alone versus K-F-D + Social History</td>
<td>K-F-D + Social History</td>
<td>( Z = -2.98^* )</td>
</tr>
<tr>
<td>Social History Alone versus K-F-D + Social History</td>
<td>Social History Alone</td>
<td>( Z = -1.33 )</td>
</tr>
</tbody>
</table>

\(*p < .01\)
history data for judging mothers' group membership accurately. Since half of the judges received the Social History first and half received it second, analysis was done to determine whether the efficacy of the social history data was affected by order of presentation. The hit-rates of judgments based on Social History alone were compared for the two groups of judges (Social History First vs. Social History Second). A Kruskal-Wallis ANOVA with Ranks revealed no significant differences in the performance of judges due to order of presentation.

Data concerning the mean performance of the clinicians versus the P.A.C.T. counselors versus the clinical students versus the speech students, under the three data conditions, are presented in Table 14 and Figure 1. It can be seen that Judge Group Two (P.A.C.T.) had the highest mean hit-rate for judgments based on Social History alone. Judge Group One (clinicians) achieved the lowest mean hit-rate using Social History alone. Groups Three (clinical students) and Four (speech students) achieved mean accuracy levels which fell between those of Groups One and Two. Finally, under the condition of K-F-D plus Social History combined, the Group Three judges (clinical students) had the highest mean accuracy. Once again, the judges in Group One achieved the lowest mean hit-rate. Accuracy levels of Groups Two and Four fell between those of Groups One and Three.

**Artistic Quality**

The final analyses conducted concerned the influence of artistic quality. Three professional artists rated the mothers' K-F-D's in terms of Poor, Average, and Good artistic quality. Initial analyses of the artists' ratings indicated that the percentage of drawings on which all three artists agreed was rather low (68%). Therefore, a composite
Table 14
Mean Hit-Rates of Four Groups of Judges Under Three Data Conditions for the Differentiation of Abusing versus Control versus Concerned Mothers

<table>
<thead>
<tr>
<th></th>
<th>K-F-D Alone</th>
<th>Social History Alone</th>
<th>K-F-D + Social History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judge Group 1</td>
<td>27.4%</td>
<td>41.8%</td>
<td>29.1%</td>
</tr>
<tr>
<td>(Clinicians)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge Group 2</td>
<td>33.3%</td>
<td>55.0% *</td>
<td>46.5% *</td>
</tr>
<tr>
<td>(P.A.C.T.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge Group 3</td>
<td>33.2%</td>
<td>51.8% *</td>
<td>53.5% *</td>
</tr>
<tr>
<td>(Clinical Students)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge Group 4</td>
<td>34.4%</td>
<td>42.8%</td>
<td>43.0%</td>
</tr>
<tr>
<td>(Speech Students)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
Figure 1. Mean hit-rates as a function of judge group ($J_1$–$J_4$) and data condition (K-F-D vs. SH vs. K-F-D + SH).
artistic rating was generated for each drawing. This rating was defined as the rating which was assigned by two of the three artists. Thus, for example, if a drawing were rated as Poor by Artist One, but Artists Two and Three rated the drawing as Average, the composite score for the drawing would be Average. It was possible to assign a composite artistic rating to all 63 K-F-D's (i.e., there were no cases in which all three artists disagreed). Of the total sample of drawings, 31.7% were rated Poor, 42.97% were rated Average, and 25.4% were rated Good.

In order to assess whether a relationship existed between mothers' actual group membership and artistic quality of the K-F-D's, a Chi-Square Analysis was performed, using the composite artistic ratings. Results of this analysis yielded no significant differences in the artistic quality ratings assigned to the K-F-D's of Abusing versus Control versus Concerned mothers.

The relationship between judges' judgments of mothers' group membership and artistic quality of the K-F-D's was also examined. Chi-Square Analyses were performed for each judge, cross-tabulating his/her judgments of group membership based on K-F-D's with the artistic quality ratings assigned to the drawings. A significant relationship between artistic quality and judgment of group membership was found for only one judge ($X^2 (4) = 11.29, p < .05$). This judge was one of the speech students. A summary of the Chi-Square Analysis for this judge is presented in Table 15. As Table 15 indicates, this judge associated poor quality of the drawing with Abusing group membership. Nearly one half (48.4%) of the drawings which were judged as Abusing were poor quality. Only 9.7% of the drawings judged as Abusing were good quality. In
Table 15

Summary of Chi-Square Analysis for Judgments of Abusing versus Control versus Concerned Group Membership Based on Poor, Average, and Good Quality X-F-D's
- Judge Number 11 -

<table>
<thead>
<tr>
<th></th>
<th>Abusing&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Control&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Concerned&lt;sup&gt;c&lt;/sup&gt;</th>
<th>X&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor Quality</td>
<td>15 (48.4%)</td>
<td>4 (16.0%)</td>
<td>1 (14.3%)</td>
<td>11.29*</td>
</tr>
<tr>
<td>Average Quality</td>
<td>13 (41.9%)</td>
<td>11 (44.0%)</td>
<td>3 (42.9%)</td>
<td></td>
</tr>
<tr>
<td>Good Quality</td>
<td>3 (9.7%)</td>
<td>10 (40.0%)</td>
<td>3 (42.9%)</td>
<td></td>
</tr>
</tbody>
</table>

*<sup>p < .05</sup>*

Note. Numbers indicate the frequency count and percentage of judgments regarding group membership based on drawings with a given quality rating.

<sup>a</sup>Total number of drawings judged as Abusing = 31.

<sup>b</sup>Total number of drawings judged as Control = 25.

<sup>c</sup>Total number of drawings judged as Concerned = 7.
contrast, 42.9% of the Concerned group judgments were based on good quality K-F-D's, and only 14.3% of the drawings judged to be Concerned were poor quality. Further analysis, using a Cramer's $V$ Correlation test, indicated that, while the relationship existed, it was not very strong ($V = .30$).
CHAPTER IV
DISCUSSION

This study was concerned with the problem of evaluating a projective test which has been proposed for use in assessing child-abusing parents. The growing problem of child abuse has led many authors to underscore the urgent need for adequate assessment techniques (Blain et al., 1981; Milner & Wimberley, 1979). Ideally, practitioners need to be able to predict the potential for child abuse and thus, take measures to prevent its occurrence. According to Meehl (1954), one step in the process of establishing that a given instrument possesses predictive capabilities is to examine the efficacy of post-diction judgments based on the test. That is, in the present case, the instrument's utility for post-hoc identification of abusing parents should be investigated.

The impetus for the present research arose from a discussion of these assessment issues in relation to the K-F-D. The forum for the discussion was a meeting known as a "Life Conference." Such conferences are held by the Wayne County Juvenile Court when a tragic situation occurs where a child is returned to the custody of parents and subsequently dies as a result of re-abuse. All of the professionals involved in the case (i.e., the judge, prosecuting and defense attorneys, D.S.S. caseworkers, and clinicians) gather to explore possible errors in the decision-making process which led to the return of the child. Clinicians who work at the community mental health clinic affiliated with the Court have been using the K-F-D in child-abuse evaluations since 1970. They have contended (Schornstein & Derr, 1977) that
family drawings can be of value in terms of predicting further abuse. In the course of discussion about a case at a Life Conference, an attorney challenged this contention. More pointedly, the clinicians were asked to offer evidence that they could accurately discriminate between known abusers and "good" parents, using the K-F-D. This challenge provided the conceptual core for the present study.

The issue raised by the attorney is, fundamentally, a question of the K-F-D's empirical validity. Some previous research has cast doubt on the validity of the instrument, in terms of discriminating between various diagnostic groups of children (McPhee & Wegner, 1976; Sobel & Sobel, 1976). However, these studies dealt with specific "signs" on the test (as delineated by Burns & Kaufman, 1972), in relation to diagnostic category. Moreover, in the case of McPhee and Wegner's study, the subjects were adolescent delinquents. It was noted by the authors that many of the subjects resisted drawing their whole families, so that the base-rate occurrence of potentially significant features of the K-F-D was low. Positive results were reported by Levenberg (1975), when clinical judges rated the K-F-D's of younger children and were permitted to use whatever global, impressionistic system of judging they preferred.

To date, no research has been conducted to examine the validity of the K-F-D for use with child-abusing parents. The arguments proposed by Schornstein and Derr (1977) provide a theoretical rationale for the utility of the K-F-D, without supporting evidence. The first purpose of the present study was to investigate the concurrent validity of the K-F-D for use in the evaluation of child-abusing parents.

Judges were given the task of differentiating between K-F-D's of
three groups of mothers, defined as Abusing, Control, and Concerned mothers. Results were analyzed for the two-group discrimination between Abusing and Nonabusing mothers (i.e., Control and Concerned groups collapsed), as well as for the three-group discrimination. The results did not support the hypothesis that the K-F-D's of Abusing mothers could be differentiated from those of Nonabusing mothers. Further, the hypothesis that a differentiation could be made between the K-F-D's of Abusing versus Control versus Concerned mothers was not confirmed.

It is possible to argue that methodological flaws in the data collection and judging procedures of the study undermined the judges' performance. It is to be noted, for example, that the Abusing mothers' group data were obtained under circumstances which differed markedly from the procedures for the Control and Concerned groups. First, the examiners who elicited the K-F-D's were different for the three groups of mothers. However, the analyses conducted using the K-F-D Scoring Guide indicated that the mothers' drawings did not differ objectively, as a function of the race of the examiner. Further, no consistent pattern of differences between the drawings elicited by the four examiners was evident. It is felt, therefore, that differences in examiner did not impact on the drawings appreciably, so as either to blur or to heighten possible differences between the mothers' groups.

An additional consideration is the fact that Abusing mothers produced the K-F-D's as part of a comprehensive clinical evaluation. They had been adjudicated in Court and were aware that the results of the clinical assessment would have impact on the Court's disposition. In contrast, the Control and Concerned mothers were simply participating
in an experiment. Their contacts with the examiner were time-limited and inconsequential, in terms of possible impact on their lives. It could be argued that, as a group, the Abusing mothers were more guarded in their execution of the K-F-D's, relative to the Control and Concerned mothers. However, Schornstein and Derr (1977) have asserted that the K-F-D cannot be "faked." Research involving the Draw-A-Person test (Graham, 1956) supports this contention. Further, the clinicians who served as the expert judges in the present study deal routinely with abusing clientele. Thus, they have presumably established their diagnostic norms for evaluating the K-F-D, based on this very type of "guarded" drawing. The difference in experimental procedures across the mothers' groups is therefore considered to be an unlikely explanation for the negative findings.

Another possible problem with the methodology concerns the composition of the three mothers' criterion groups. Efforts were made to ensure, insofar as it was possible, that the Control and Concerned groups did not contain any mothers who were actually child abusers. While it is clear that such a determination could not be made with absolute certainty (because child abuse may occur undetected or unsuspected), there is no reason to believe that the procedure was faulty in this regard. However, it may be that the distinction between the Control and Concerned mothers was not sufficient to define two separate and unique groups. It could be argued that mothers who take their children to a clinic for routine checkups, perhaps to prevent the onset of serious illness, are as "concerned" as mothers who spend time with their hospitalized children. Some of the judges commented during the judging procedure that they were "just guessing" on the differentiation between
the Control and Concerned mothers. Thus, there is some anecdotal support for the possibility that judges' accuracy rates were lowered as a result of inadequate criterion definition. However, the fact remains that when the results were analyzed with the two control groups collapsed into a single Nonabusing category, judges were still unable to differentiate accurately between the K-F-D's.

Considerable criticism has been voiced against a methodology which requires judges to render "blind" judgments of test data (Goldberg, 1959; Harrower, 1954). One step in the judging procedures utilized in the present study involved judges' ratings of the K-F-D's in the absence of any background information. The hit-rates under this data condition were uniformly low, for both the two and three mothers' group differentiations. In part, the finding that judges were unable to differentiate between the K-F-D's of the mothers' groups may be attributable to this methodology. However, the judges in Levenberg's (1975) study also rendered their judgments on the basis of K-F-D's alone. The mean hit-rate reported by Levenberg for the dichotomous differentiation between "disturbed" versus "normal" children was 65% (p < .05). Thus, in contrast to the findings of the present study, Levenberg's judges were able to differentiate between the children's groups with an accuracy level that significantly exceeded chance, using K-F-D's alone. Furthermore, while the accuracy of the judges in the present study did improve when K-F-D's were combined with Social History data, the mean hit-rate did not exceed chance significantly. The data thus do not support a conclusion that the negative findings are largely attributable to the use of "blind judgment" methodology.

A more likely source of explanation for the judges' inability to
differentiate the K-F-D's lies in the examination of judges' judging behavior. Results of the analyses on inter-judge reliability indicated, for example, that there was very little agreement among the judges as a whole, regarding their decisions. The judges did tend to agree more among themselves, within their own judge groups, but not relative to judges in other groups. Of the four groups of judges, the clinical experts and the naive speech students shared the lowest inter-judge reliability.

These findings are in sharp contrast to those reported by Albee and Hamlin (1949), regarding judgments of adjustment based on the Draw-A-Person test (inter-judge agreement for 15 judges = 97%). In Albee and Hamlin's study, all of the 15 judges were clinicians. Thus, this reliability figure must be compared with the within-group agreement measure for Judge Group One in the present study. Albee and Hamlin's judges had much more consensus than the clinicians in the current study. The contrast may be mitigated somewhat by the fact that Albee and Hamlin's judges were making a two-choice decision, rather than a (perhaps more difficult) three-choice decision. Further, the total n in their study was 45, as compared with 63 decisions in the present study. In Watson's (1967) study, 24 clinicians made judgments about the diagnostic category of four different groups of patients, based on the Draw-A-Person test. The total number of judgments for each judge was 48. Watson's inter-judge reliability results (41%) were much lower than Albee and Hamlin's, but still higher than the results in the present study. Again, the judges made fewer decisions in Watson's study than in the present study. Watson's conclusions certainly appear applicable to the current research, however, given the even smaller degree of
consistency among the judges in the present study. In Watson's opinion, the apparent disagreement among clinicians regarding the diagnostic implications of various drawing characteristics significantly impedes clinicians' ability to assign accurate diagnoses, based on the drawings. Simply put, if clinicians cannot agree, it is likely that they cannot be right.

The second aspect of judges' judging behavior which bears examination is the type of errors which judges tended to make. Harrower (1954) postulated that clinicians might have a tendency to be overly sensitive to pathology, resulting from limited practical exposure to "normal" populations. Results of research conducted by Hiler and Nesvig (1965) and Soskin (1954) confirmed this impression. Moreover, Soskin found that judges who had had very limited training in the interpretation of projective test data were as vulnerable to this error as experienced clinicians. The findings of the present study concur with Soskin's results. In general, judges tended to err in the direction of overidentifying pathology. That is, a higher percentage of errors involved the misidentification of Nonabusing mothers as Abusing than vice versa. Of the four judges who erred in the opposite direction (i.e., misidentifying Abusing mothers as Nonabusing), three were assessment-naive judges. Based on the results of the current research, the author is inclined to endorse Harrower's (1954) observation that, "If one has been embedded in a community which netâ one a fairly consistent and uniform type of record, one tends to set one's standards or norms by this standard and to superimpose it everywhere" (p. 295).

In defence of the judges in the current study, it might be argued that, philosophically, a false positive error is a "better" error to
make. From the point of view of protecting the welfare of children, perhaps it is preferable to err on the side of overdiagnosing than to miss a correct diagnosis of child-abusing behavior. An informal poll of the clinicians who participated as expert judges in the study revealed that this was, indeed, their shared opinion. Nevertheless, the design of the current study favored a judgment pattern involving false negatives, in that the actual frequency of Nonabusing mothers outnumbered Abusing mothers by a ratio of 2:1. The judge who achieved the highest accuracy rates for both the two- and three-mothers' group discriminations did commit more false negative than false positive errors.

The final consideration regarding the judges' inability to differentiate among the Abusing versus Control versus Concerned K-F-D's concerns the assumption which judges may have made about drawing features. The results of the analyses regarding the objective scoring of the K-F-D's indicated that the mothers' groups differed significantly on only two variables. The differences were not consistent, however. The Concerned mothers' drawings more frequently showed one feature considered to be "pathological" (Barriers), while the Abusing mothers' drawings more frequently depicted the other "pathological" sign (Omission of family members). Trends in certain other features were also inconsistent. It would therefore be reasonable to conclude that the scorable features of the K-F-D's simply did not discriminate the mothers.

The judges in the study were instructed to formulate their judgments of the mothers' group membership using any system or method they preferred. The implication is that judges would develop some global impression of the drawings. However, as noted by Albee and Hamlin (1949), Hiler and Nesvig (1965) and others, the clinical interpretation
of projective drawings is likely based on some synthesis of drawing elements, at least for those judges who have some assessment expertise. It follows, then, that if the elements of the drawing which are being "added up" in this private formula do not bear any relationship to the criterion, neither will the final judgment. It has been argued that when judges act on the assumption that signs are associated with certain symptomatic categories, and no such association exists, their judgments are based on illusory correlations (Chapman & Chapman, 1965, 1967). As noted above, the judges in the current study were not given specific drawing signs with which they were expected to work. Additionally, they were not questioned regarding the manner in which they had formulated their judgments. Thus, it cannot be said with certainty which, if any, drawing features judges may have been utilizing. However, the fact that the judges who achieved the two lowest hit-rates were clinicians does suggest that these judges, at least, may have been systematically erring. The argument that illusory correlations interfered with judges' accuracy seems compelling.

The second purpose of the present study was to explore some variables which may affect the accuracy of judgments based on the K-F-D. A review of the literature concerning clinical judgment and figure drawings suggested that one important factor may be the experience and professional training of the judges (Goldberg, 1959; Guinan & Hurley, 1965; Hiler & Nesvig, 1965). In Levenberg's (1975) study using the K-F-D, doctoral level clinicians were more accurate than psychology interns and secretaries, although the differences in accuracy levels between groups were not significant.

In the present study, four groups of judges evaluated the K-F-D's
of the mothers' criterion groups. The judge groups varied in terms of their professional work experience with child-abusing clientele, and their assessment expertise. The expectation that the clinical expert group would achieve the highest accuracy levels was not confirmed. In fact, the most naive group (speech students who had neither abuse work experience nor assessment training) were the most accurate, while the clinicians were the least accurate. However, the differences between the groups were not significant.

The direction of these results is consistent with the findings reported by Goldberg (1959) and Hiler and Nesvig (1965). Hiler and Nesvig offered the explanation that well developed intuitive ability, rather than formal clinical training, was of primary importance in drawing interpretation. It should be noted that in Hiler and Nesvig's study, all of the judges (clinicians and nonclinicians alike) were able to differentiate the drawings of diagnostic groups at a level which significantly exceeded chance. Since the findings in the present study are virtually the opposite (i.e., none of the judges could discriminate the three mothers' groups better than chance), Hiler and Nesvig's explanation does not seem to apply.

Regarding the 'failure' of clinicians to perform significantly better than nonclinicians using the K-F-D, Levenberg (1975) suggested that circumstances of the judgment task impeded clinicians' performance. He pointed out that simple judgments of "disturbed" versus "normal" K-F-D's are not the kind of decisions which clinicians might ordinarily make using the drawing. If judges were asked to make a more finely tuned discrimination or diagnosis, the superiority of clinical judges over nonclinical judges might prove significant. Furthermore, Levenberg
contended that the limited nature of the clinical data on which the judgments were based (i.e., K-F-D's alone) may have attenuated differences between judge groups. Specifically, the relative accuracy of clinical judges may have been lowered because data they might ordinarily have available to them (other test data, interview material, social history) were not included.

The first explanation offered by Levenberg does not appear to be relevant to the current research. In this study, the clinicians were rendering judgments about a criterion (namely, child abuse) which they consider routinely in their professional work. Moreover, these experts have argued that this judgment is the precise purpose for which they do use the K-F-D (Schornstein & Deër, 1977). The second argument posited by Levenberg also appears to be inapplicable to the present study. When judges were given K-F-D's plus Social Histories with which to make their judgments, the clinicians still achieved the lowest accuracy levels, among the four judge groups.

The most parsimonious explanation for the findings regarding the rank-ordering of judge groups' accuracy ratings seems to relate to the discussion of judges' biases, noted above. The clinicians may have performed most poorly because they were most vulnerable to errors based on unfounded assumptions about the drawings. Additionally, they showed a greater tendency to overidentify pathology, as a group, than did the other judge groups. It is suggested that these factors may be responsible for the findings in the present study. The hypothesis concerning the middle ranking of the P.A.C.T. counselors and clinical students was confirmed, but the group differences are slight enough that particular discussion does not seem warranted.
The influence of specific task training on judges' accuracy has been explored experimentally by some authors (Goldberg, 1968; Oskamp, 1962). Murray and Deabler (1958) found that task training did improve judges' accuracy in making diagnostic decisions based on figure drawings. In the present study, the judges in Group Two (P.A.C.T. counselors) received training in the interpretation of the K-F-D with child-abusing clientele. The comparison of the judges' pre- and post-training accuracy rates did not reveal significant improvement as a result of training. In fact, the results were generally in the direction of a decline in performance after training.

One possible explanation for this finding lies in the elements of the training seminar. It may be that the training discussion was not in-depth enough or too broadly focused. Additionally, the timing of the training, relative to the post-training judging session, may not have been optimum. Goldberg (1968) noted that the effects of a training experience are likely greatest if the information is immediately applied to the judgment task. Finally, the training in the present study did not involve hands-on experience with protocols, nor feedback regarding the accuracy of the sample judgments. In Murray and Deabler's (1958) study, judges received feedback regarding their performance on successive series of drawings. The judges were encouraged to examine their own patterns of "hits" and errors, and incorporate the information into the next set of judgments. According to Goldberg (1968), such feedback is the critical element in efforts to improve judgment accuracy. The methodology utilized by Murray and Deabler would also provide a clear check on whether judges were actually learning anything about the task. The mere provision of training does not, of course,
guarantee that subjects will be able to absorb the relevant information.

It will be recalled that two of the clinical judges in the study also provided the training to the P.A.C.T. counselors. While this might have represented some contamination, in terms of the clinicians' own later performance, it is felt that such effects were minimal, if any. The clinicians judged the drawings themselves more than a week after the presentation of the training. Further, no one in the training audience voiced any specific questions about the experimental drawings. It is tempting to suggest, however, that the results concerning the effects of the training on the P.A.C.T. judges' performance are directly connected to the poor judgment performance of the clinicians. Perhaps the apparent decline in post-training accuracy occurred because the clinicians taught the P.A.C.T. counselors a set of illusory assumptions about the drawings.

Previous research has examined the effects of both quantity and type of data on judgment accuracy (Kostlan, 1954; Oskamp, 1965; Soskin, 1954). The relative superiority of batteries that included social history information over those that consisted of test data alone was established in Kostlan's (1954) study. Kostlan also reported that judgments based on demographic information alone were significantly more accurate than chance level prediction. In the present study, each judge rendered judgments about the mothers' group memberships under three data conditions: K-F-D alone; Social History alone; and K-F-D plus Social History combined.

Judgments based on Social History alone were found to be the most accurate for the judges as a whole. The P.A.C.T. counselors were especially adept at differentiating between the mothers' groups on the basis of social history data. Apparently the work experience which these judges had had with abusing clientele, along with the social work
orientation of their field experience, had positive impact on this aspect of their judgment abilities. Interestingly, analyses of the sociodemographic information obtained from the mothers in the study revealed very little actual difference between the groups. Certainly, there was no support for the argument that Abusing mothers differ significantly from Nonabusing mothers on various indices of socioeconomic status (Pelton, 1978). A number of trends were evident with regard to possible social/emotional stressors, however. Evidently the judges were able to detect and utilize this information in some fashion, though again, the mean accuracy was not particularly impressive.

The results of the present study indicated that judgments based on K-F-D plus Social History combined were significantly more accurate than judgments based on the K-F-D alone. This is consistent with the findings reported by Kostlan (1954), and is in consonance with the generally held opinion that drawings are best utilized in the context of a more comprehensive assessment (Fisher, 1957; Levenberg, 1975; Myers, 1978).

It is somewhat more difficult to explain the finding that judgments based on K-F-D plus Social History were generally less accurate than judgments based on Social History alone. Some authors have observed that the accumulation of additional information did not improve accuracy beyond a certain level (Golden, 1964; Oskamp, 1965). Golden postulated that judges did not utilize the additional data reliably. Sines (1959) and Oskamp (1965) both reported that judges' perceptions of the subjects appeared to crystallize at some point in the process. Judges seemed to become entrenched in their judgments and were unwilling to alter their decisions after initial formulation. However, none of these authors
was addressing a situation in which judgment accuracy was actually lower with additional data.

The comments of Cummings (1954) regarding "rater demoralization and fatigue" may be relevant to this discussion. According to Cummings, when judges are asked to deal with a large volume of protocols, or test data which are complex, they may become overwhelmed by the task. Some of the judges participating in the current study did remark that they were beginning to feel quite tired near the end of the judging session. A few complained that the task was becoming tedious. Thus, it may be that the number of protocols was too large to be comfortably handled in one sitting. The comments regarding fatigue were made by judges in Groups One and Two (clinicians and P.A.C.T. counselors). These two judge groups were the groups that exhibited lower performance in the combined data condition, relative to the Social History alone condition. In contrast, the two student groups expressed enthusiasm for and interest in the study. They were especially eager to hear the author's comments about the hypotheses of the study, and stayed on after the judging sessions to discuss the research. The clinical student group and the speech student group both showed improved accuracy (albeit slight) in the combined data condition.

A close examination of the judge groups' performances under the three data conditions suggests an alternative explanation for the results of the combined data situation. The clinicians were noticeably less accurate than the other three groups of judges, using K-F-D plus Social History. Further, the discrepancy between the Social History alone accuracy rate and the K-F-D plus Social History accuracy rate is larger for this group than for the other groups. Once again, it would
appear that the impact of the K-F-D served to lower the judgment accuracy of the clinicians in particular. This group's markedly lower performance on the combined data condition probably contributed a great deal to the lowered performance of the judges as a whole.

The potentially confounding influence of the artistic quality of drawings has been investigated by several authors (Levy et al., 1963; Sherman, 1958a, 1958b; Whitmyre, 1953). Strong correlations were reported between the artistic quality of drawings and judges' judgments of adjustment. It has generally been concluded that judges are actually responding to the aesthetic merits of drawings when they attempt to make diagnostic judgments based on drawings. In the present study, the K-F-D's obtained from the mothers were rated for artistic quality by three professional artists. The relationship between artistic quality of the K-F-D's and judges' judgments of mothers' group memberships based on the drawings was examined. The results suggested that the judges in this study were not influenced by the artistic quality of the drawings. Moreover, there was no evidence of a relationship between artistic quality of the drawings and actual group membership. Unlike Sherman (1958b), this author is unable to conclude that the judges' inability to differentiate between the groups of K-F-D's is attributable to the extraneous influence of the drawings' artistic quality.

The final aspect of this study concerned the exploratory use of an objective scoring guide for the K-F-D. The guide was developed by the author, based on the system proposed by Myers (1978). The inter-rater agreement in the application of the scoring system was encouragingly high. The results suggest that judgments about various structural and qualitative features of the drawings can be made reliably, when well
articulated methods of measurement are available.

The findings regarding differences among the objective scoring of Abusing versus Control versus Concerned mothers' K-F-D's were discussed above. Additional analyses using the K-F-D scoring system indicated little difference between the K-F-D's of black and white mothers. Adler's (1971) findings regarding racial differences on the Draw-A-Person test are not corroborated by the results of the present study. Analyses of differences in the K-F-D features of three socioeconomic groups were conducted in the present study. Significant differences were found on three K-F-D variables, but the results were inconsistent. No support was found for Adler's (1971) contention that lower-class subjects tend to exhibit more distortion and problems of integration in their drawings.

In conclusion, it has been argued that the K-F-D is clinically useful in the assessment of child-abusing parents (Schornstein and Derr, 1977). This study was undertaken to investigate the empirical validity of the K-F-D and to explore some parameters which may affect the accuracy of clinical judgments based on the drawings. It was contended that judges would be able to differentiate between the K-F-D's of three different groups of mothers. Further, it was expected that the professional experience and expertise of judges, as well as the provision of specific task training, would have impact on judges' performance. The effects of presence/absence of supporting sociodemographic data and the possible influence of drawings' artistic quality were explored. The major conclusion from this study is that judges were unable to differentiate between Abusing, Control, and Concerned mothers on the basis of the K-F-D, even when they had accompanying social history data to
consider in their decisions. Moreover, familiarity with the K-F-D and its possible clinical use was actually associated with lowered performance on the judgment task in this study.

The combined findings that judges could not differentiate between the mothers' drawings accurately or reliably, and that the mothers' drawings did not differ on numerous scorable patterns, cast serious doubt on the validity of the K-F-D for use in child-abuse evaluations. Meehl's (1954) comment seems particularly poignant:

In answer to a demand for validation, one sometimes hears it stated that, whereas a certain clinical device or method has not proved valid "in the usual sense," such formal validation is not required, since the instrument has been validated by its "clinical usefulness." When we hear this from a clinician, all we can say is that he thinks he is using it to advantage. (p. 136)

It should be noted that a single study cannot bear the burden of establishing/refuting an instrument's validity once and for all. The definition of the "Concerned" criterion group presented some problems in the present study, and additional research with other mothers' groups might prove more fruitful. Mothers who are actively involved in PTA or other children's interest groups, and mothers who are members of the La Leche League (an organization which strongly espouses the benefits of breast-feeding) are two groups which come readily to mind. It might also be worthwhile to replicate the study using larger n's (of both judges and mothers). Perhaps fathers could be utilized as criterion groups. Even more effective might be the exploration of K-F-D differences among different groups of children. Following Blain et al. (1981), the K-F-D's of abused versus control children could be compared. Since the K-F-D was originally developed for use with children (Burns & Kaufman, 1970), such a design would have rational basis. Finally, the
elements of the judgment task itself could be altered. The effectiveness of the K-F-D within a more standard clinical battery (i.e., interview, social history, additional test data) should be explored. Perhaps the use of serial K-F-D's, collected over a period of time (Harrower, 1954; Schornstein & Dorr, 1977) would reveal significant utility.
APPENDIX A

SOCIAL HISTORY QUESTIONNAIRE
COMPLETED BY MOTHERS
SOCIAL HISTORY QUESTIONNAIRE
COMPLETED BY MOTHERS

The following information is requested for research purposes only. This
information will be treated as confidential. No information which would
reveal the identity of the subjects will be included in the write-up of
the research results.

PLEASE ANSWER ALL QUESTIONS.

My First Name: ________________________ Age: _________

Name of Child who was treated at hospital: ___________________________

D.O.B.: __________ Marital Status: (check one) Single__ Married__
Divorced__ Separated__ Widowed__ Living with a
partner __

Race: Black ____ White ____ Other _________________________

Education: Highest level of education completed (check one)
Elementary school ____
Junior high school ____
High School ____
Community college/trade school ____
University Bachelor's program ____
University postgraduate program ____

Number of Children: Number residing in the home ________.
Number residing outside the home ________.

Age and Sex of each child living in the home:
M for Male, F for female.

____ M F _____ M F _____ M F _____ M F _____ M F _____ M F

Family Income for 1981 (check one)

Social Security ____ Amount per year ________
Public Assistance (ADC, General Welfare, etc.) ____ Amount per year ______
Less than 5,000 ______
5,000 to 7,999 ______
8,000 to 10,999 ______
11,000 to 14,999 ______
15,000 to 19,999 ______
20,000 to 29,999 ______
30,000 to 49,999 ______
50,000 or more ______

Occupation of Head of Household: ________________________________
Background Information:

How many sisters and brothers do you have? ________

Were you raised by your natural parents? ________ in foster care? ______
in an adoptive home? ________ (check one)

Did you suffer any serious illness or injuries as a child? Yes____ No____

Were you ever abused as a child? Yes____ No____

If you are married, do you consider your marriage a good one? Yes____ No____

Have you or any member of your family ever had emotional problems requiring professional treatment? Yes____ No____

If you have a spouse/living together partner, was he abused as a child? Yes____ No____

Have you or any member of your family ever had a drinking problem? Yes____ No____

Have you or any member of your family ever been convicted of a crime? Yes____ No____

Have you ever had contact with the court and/or the Department of Social Services on charges of abusing/neglecting your child(ren)? Yes____ No____

Do you wish to be contacted about the results of this research project? Yes____ No____

Address where you can be reached:

Number  Street  City  Zip Code

This space for use by researchers only.

Code Number: ____________________

Group Number: ____________________

Approval for group membership verified: Yes____ No____
APPENDIX B

SOCIAL HISTORY DATA SHEET
PROVIDED TO JUDGES
SOCIAL HISTORY DATA SHEET
PROVIDED TO JUDGES

The mother who completed the attached drawing provided the following information about herself:

Age: _____ Age of child seen at hospital: _____

Marital Status: Single _____ Married _____ Divorced _____
Separated _____ Widowed _____ Living with a partner _____

Race: Black _____ White _____ Other _______________________

Education: The highest grade I completed in school was _________

Number of Children: Number residing in the home _______
Number residing outside of the home _______

Age and Sex of each child living in the home:
M for male, F for female.

_____ M F _____ M F _____ M F _____ M F _____ M F

_____ M F _____ M F _____ M F _____ M F _____ M F

Family Income for 1981:

Social Security _____ Amount per year ______
Public Assistance (ADC, General Welfare, etc.) _____ Amount per year _____
Less than 5,000 ________
5,000 to 7,999 ________
8,000 to 10,999 ________
11,000 to 14,999 ________
15,000 to 19,999 ________
20,000 to 29,999 ________
30,000 to 39,999 ________
40,000 to 49,999 ________
50,000 or more ________

Occupation of Head of Household: ________________________________

Background Information:

How many sisters and brothers do you have? ________________

Were you raised by your natural parents? _____ in foster care? _____
in an adoptive home? _____

Did you suffer any serious illness or injuries as a child? Yes_____ No______
SOCIAL HISTORY DATA SHEET
PROVIDED TO JUDGES
CONTINUED

Were you ever abused as a child? Yes_____ No_____

If you are married, do you consider your marriage a good one? Yes_____ No_____

Have you or any member of your family ever had emotional problems requiring professional treatment? Yes_____ No_____

If you have a spouse/living together partner, was he abused as a child? Yes_____ No_____

Have you or any member of your family ever had a drinking problem? Yes_____ No_____

Have you or any member of your family ever been convicted of a crime? Yes_____ No_____
APPENDIX C

KINETIC FAMILY DRAWING
SCORING GUIDE
KINETIC FAMILY DRAWING SCORING GUIDE

Drawing Number ____________________

1. Physical Proximity of Figures (Measured in Inches):
   a) distance between mother and target child (if present) _____
   b) mean distance between mother and all figures _____

2. Height of Figures (Measured in Inches):
   a) height of mother figure ______
   b) height of target child (if present) _____
   c) ratio of mother to target child ______

3. Barriers between Figures (object(s) other than lines between the mother and another figure in the drawing):
   present_____ absent_____

4. Fields of Force:
   a) balls: present____ absent____
   b) fire: present____ absent____
   c) electrical appliances: present____ absent____
   d) x's: present____ absent____

5. Pencil Erasures: present____ absent____

6. Arm Extensions (arms extended away from the side of body by greater than a 45° angle):
   present_____ absent____

7. Description of Figures' Actions:
   a) visible actions are strange/unreal/unexpected present____ absent____
   b) mother or other figures highly distorted present____ absent____

8. Safety of Figures (figure(s) in dangerous position):
   present____ absent____

9. One or More Figures Missing Essential Body Parts:
   present____ absent____

10. Rotation of one or more figures by 45 degrees or more:
    present____ absent____

11. Shading or Crosshatching (except hair):
    present____ absent____

12. Compartmentalism of Figures (one or more straight lines used to separate one or more figures):
    present____ absent____
13. Folding Compartmentalism (folding the drawing paper into sections and drawing figures in more than one section): present ___ absent ___

14. Underlining of Individual Figures: present ___ absent ___

15. Lining at the Bottom of the Page: present ___ absent ___

16. Lining at the Top of the Page: present ___ absent ___

17. Encapsulation (complete enclosure of one or more figures, but not all, by lines which do not stretch the length of the page): present ___ absent ___

18. Edged Placement of Figures (the drawings of all figures on two or more edges of the paper): present ___ absent ___

19. Evasions (one or more drawings depicting stick figures or no action): present ___ absent ___

20. Number of Household Members:
   a) one or more figures omitted ___
   b) one or more figures added ___
   c) correct number depicted ___
APPENDIX D

EXPLANATION OF RESEARCH PROVIDED TO MOTHERS
EXPLANATION OF RESEARCH
PROVIDED TO MOTHERS'

We are university students doing a project for school. My name is (name) from (school) and this is (name) from (school). We are conducting a study designed to investigate the usefulness of a drawing test as an indicator of family relations. The name of the test is the Kinetic Family Drawing test. We would like you to complete a simple task that will take only a few minutes of your time. Your work will be kept confidential, and will be used for research purposes only. When you have finished your drawing, we will ask you a few questions about your family. Again, any information you do supply will be used in the research only, and will not be disclosed. All information which could identify you as a participant will be kept confidential. Although we ask you to give us your name, you will be assigned a code number and your name will then be removed from the information sheet you fill out. Would you like to participate?

(At this point the signed consent of each participant will be obtained)
APPENDIX E

INFORMED CONSENT FORM FOR MOTHERS
CONSENT FORM FOR MOTHERS

There is a study being done to see if drawings made by mothers will help people understand how families get along. I will take part in the study.

I will be asked to make a drawing of my family. I will also be asked to give some information about my background.

Information about what the doctors learn from this study may be published, or given to other people doing research, but my name will not be used.

This is a research study, and it may not be of any direct benefit to me. I may withdraw from this study at any time.

I understand that in the event of physical injury resulting from taking part in this study, no compensation and no medical treatment or reimbursement is offered.

Any questions I have asked about this study have been answered. If I have any questions later on, Mrs. Alice Belden will answer them for me. She may be reached at 494-5281.

By signing this paper I am saying that I have read and understood it and I agree to take part in the study.

DATE: ________________________________

NAME OF PARTICIPANT
APPENDIX F

EXPLANATION OF STUDY
PROVIDED TO JUDGES
EXPLANATION OF STUDY
PROVIDED TO JUDGES

The context of this study is, in the broadest sense, the tragic problem of child abuse. More specifically, the study's subject matter concerns the clinical assessment of parents who physically abuse their children. Considerable professional effort has been devoted to the search for an effective technique with which at-risk families can be evaluated. A psychological test called the Kinetic Family Drawing Technique is one instrument which has been proposed for use in this fashion.

[At this point, judges in Groups 2 and 4 review the "Explanation of the Kinetic Family Drawing"]

It has been argued that the K-F-D is not only a valuable tool in the hands of mental health professionals, but may also be an effective screening device for use by nonclinical field workers. To date, these premises have not been tested experimentally. They are the focus of this study.

The purpose of this experiment is twofold. The first purpose is to investigate the validity of the Kinetic Family Drawing as an assessment instrument in child-abuse evaluation. One measure of a test's validity is the extent to which people who are known to differ in a particular trait respond differently on the test. In this case, if the K-F-D is a valid measurement, known child-abusers should produce drawings which are detectably different from those done by nonabusing parents.

The second purpose of this experiment is to examine some factors which may influence the ability of judges to differentiate accurately the Kinetic Family Drawings of abusing and nonabusing parents. One aspect of this problem concerns the question of whether or not particular expertise is necessary to detect differences in the drawings. Factors such as clinical experience, education, and training in psychological testing may affect the accuracy with which the drawings can be evaluated. The amount of data which judges have available to them is also a consideration.

This study involves two phases. In the first phase, Kinetic Family Drawings and social history information were obtained from three different groups of mothers.

In the second phase, the part of the experiment in which you are participating, the K-F-D's and background information are going to be evaluated or judged by four different groups of judges. The groups of judges differ in regard to professional background and training. Besides your group, there is a group of ________ and ________. In order further to evaluate the effect of specific training on interpretation of the K-F-D, one group (your group) will
judge the data twice: once before they (you) receive training and then, after participating in a training seminar, they (you) will judge the data again.

As a final part of the study, three professional artists will assess the artistic quality of the drawings. Additionally, the drawings will be scored by myself and another rater, using an objective scoring system which I have developed.

Are there any questions?
APPENDIX G

EXPLANATION OF THE KINETIC FAMILY DRAWING
GIVEN TO DRAWING-NAIVE JUDGES
EXPLANATION OF THE KINETIC FAMILY DRAWING
GIVEN TO DRAWING-NAIVE JUDGES

Drawings of the human figure have for years been considered useful in the psychological evaluation of different clinical populations. Figure drawings are one example of a broad class of psychological tests called projective tests. Theory and research suggest that the subject depicts various aspects of his or her own self-concept and emotional concerns in the drawing, i.e., the subject "projects" a part of the self into the task. The authors of the Kinetic Family Drawing argued that by adding the element of imaginary movement to the drawing task, even more information about the subject's emotional functioning could be obtained. Thus, they devised the Kinetic Family Drawing, in which the subject is told to draw a picture of his or her family with everyone, including him/herself, doing something, some kind of activity. The subject then identifies the various figures by name and age. The authors of the test believe that the subject's drawing can indicate how he/she feels about self, self within the family context, perceptions of other family members, and possible conflicts in family interactions. Clinicians at the Clinic for Child Study have extended these ideas and applied them to the evaluation of clients suspected of abusing their children. It is postulated that the parents' individual problems as well as their perception of the child, conflicts within the family, and the stressors operating on the family members may all be revealed in the drawing. The study in which you are involved as a judge is concerned with exploring these ideas.
APPENDIX H

JUDGES' BACKGROUND INFORMATION FORM
JUDGES' BACKGROUND INFORMATION FORM

Name of Rater: __________________________

Race: ____________________  Sex: ____________________

Profession/Occupation: __________________________

Highest Educational Training Level Completed: ____________________

Years of Work Experience in the Field of Abuse/Neglect: ____________

Years of Work Experience using Figure Drawing Techniques: ____________

I have___ have not___ had training in the use of projective tests
(i.e., formal educational experience in this area).

I have___ have not___ attended training sessions and/or conferences
on the use of the Kinetic Family Drawing Technique or other figure-
drawing techniques.

Do you wish to be contacted about the results of this research project?

Yes____No____

Address where you can be reached:

Number and Street: ____________________  City: ____________________  Zip Code: ____________________

This space for use by researchers only.

Number assigned to rater: ____________

Rater Group Number: ____________
APPENDIX I

JUDGES' INFORMED CONSENT FORM
JUDGES' INFORMED CONSENT FORM

I hereby agree to participate as a subject in a research project designed to investigate (1) the validity of the Kinetic Family Drawing Technique as an assessment instrument in child abuse evaluations; (2) some factors which may influence the ability to differentiate accurately the Kinetic Family Drawings of abusing and nonabusing mothers.

I understand that my participation in this research project will be in the capacity of a judge. My task as a judge will be to make a decision regarding group membership of three groups of mothers, based on the drawings and background information provided. I understand that I will also be asked to rate my level of confidence in each decision made.

I understand that the results of the judgments made in this study will be used solely for the purpose of research, and that all information which could identify me as a subject will be kept confidential.

I understand that I am free to withdraw my consent and discontinue my participation at any point during the procedure.

The person whom I may contact if I wish to discuss this research is

Pamela Arnaut    Phone: 577-9312

_________________________________________  __________________
Signature of the Participant               Date Signed
APPENDIX J

SAMPLE RATING FORM
FOR K-F-D'S
JUDGES' RATING FORM

Judge Number ____________

Please refer to the attached sheet for group descriptions and rating scale.

<table>
<thead>
<tr>
<th>Drawing Number</th>
<th>Group Number</th>
<th>Confidence Rating</th>
<th>Drawing Number</th>
<th>Group Number</th>
<th>Confidence Rating</th>
<th>Drawing Number</th>
<th>Group Number</th>
<th>Confidence Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td></td>
<td></td>
<td>22</td>
<td></td>
<td></td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>02</td>
<td></td>
<td></td>
<td>23</td>
<td></td>
<td></td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>03</td>
<td></td>
<td></td>
<td>24</td>
<td></td>
<td></td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>04</td>
<td></td>
<td></td>
<td>25</td>
<td></td>
<td></td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>05</td>
<td></td>
<td></td>
<td>26</td>
<td></td>
<td></td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>06</td>
<td></td>
<td></td>
<td>27</td>
<td></td>
<td></td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>07</td>
<td></td>
<td></td>
<td>28</td>
<td></td>
<td></td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td></td>
<td></td>
<td>29</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>09</td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>31</td>
<td></td>
<td></td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>32</td>
<td></td>
<td></td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>33</td>
<td></td>
<td></td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td>34</td>
<td></td>
<td></td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td>35</td>
<td></td>
<td></td>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td>36</td>
<td></td>
<td></td>
<td>57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>37</td>
<td></td>
<td></td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td>38</td>
<td></td>
<td></td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td>39</td>
<td></td>
<td></td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>41</td>
<td></td>
<td></td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td>42</td>
<td></td>
<td></td>
<td>63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX K

SAMPLE RATING FORM
FOR SOCIAL HISTORIES
**JUDGES' RATING FORM**

Judge Number __________________

Please refer to the attached sheet for group descriptions and rating scale.

<table>
<thead>
<tr>
<th>Social History</th>
<th>Group Number</th>
<th>Confidence Rating</th>
<th>Social History</th>
<th>Group Number</th>
<th>Confidence Rating</th>
<th>Social History</th>
<th>Group Number</th>
<th>Confidence Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td></td>
<td></td>
<td>22</td>
<td></td>
<td></td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>02</td>
<td></td>
<td></td>
<td>23</td>
<td></td>
<td></td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>03</td>
<td></td>
<td></td>
<td>24</td>
<td></td>
<td></td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>04</td>
<td></td>
<td></td>
<td>25</td>
<td></td>
<td></td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>05</td>
<td></td>
<td></td>
<td>26</td>
<td></td>
<td></td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>06</td>
<td></td>
<td></td>
<td>27</td>
<td></td>
<td></td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>07</td>
<td></td>
<td></td>
<td>28</td>
<td></td>
<td></td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td></td>
<td></td>
<td>29</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>09</td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>31</td>
<td></td>
<td></td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>32</td>
<td></td>
<td></td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>33</td>
<td></td>
<td></td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td>34</td>
<td></td>
<td></td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td>35</td>
<td></td>
<td></td>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td>36</td>
<td></td>
<td></td>
<td>57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>37</td>
<td></td>
<td></td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td>38</td>
<td></td>
<td></td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td>39</td>
<td></td>
<td></td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>41</td>
<td></td>
<td></td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td>42</td>
<td></td>
<td></td>
<td>63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX L

INSTRUCTIONS TO THE JUDGES
INSTRUCTIONS TO THE JUDGES

You will be examining Kinetic Family Drawings and social history data provided by 63 different women. All of the women are mothers with at least one child between the ages of one month and six years. The drawings and social histories were obtained from three different groups of mothers, as described below:

(1) Mothers who had been adjudicated in Juvenile Court on a DSS-3200 petition alleging that their children had been abused. The petitions were filed against these women as a result of a referral from a hospital where the children had been treated for suspected abuse-related injuries. These mothers constitute the Abusing group.

(2) Mothers who were waiting for their children to receive routine checkups or treatment for minor ailments at the Outpatient Clinic of Children's Hospital. These mothers constitute the Control group.

(3) Mothers who were spending extended periods of time (including overnight) with their children while the children received inpatient treatment for noninjury conditions at Children's Hospital. These mothers constitute the Concerned group.

You will be examining data from the mothers three times. First, you will receive either the Kinetic Family Drawings alone or the Social Histories alone. Next, you will be given the material you did not receive the first time, and you will judge it alone. (Note that the Drawing #'s and Social History #'s are arbitrary, and do not necessarily correspond to each other. Thus, Drawing # 01 is not necessarily associated with Social History # 01.) Finally, you will receive both the K-F-D and the social history information together, correctly matched. During each phase, your task will be the same.

(1) Please come to a decision regarding the group membership (i.e., Abusing, Control, or Concerned) of each mother, based on the drawing and/or information provided. Make your judgment using any system or method you prefer, and note it beside the appropriate drawing/social history number on the rating sheet. Designate the Abusing group with the number "1," the Control group with the number "2," and the Concerned group with the number "3."

(2) Please indicate the degree of confidence you have in your judgment, using the five-point scale described below:
INSTRUCTIONS TO THE JUDGES
CONTINUED

Level of Confidence

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0%</td>
<td>25%</td>
<td>50%</td>
<td>75%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>absolutely uncertain</td>
<td>somewhat uncertain</td>
<td>neutral</td>
<td>somewhat certain</td>
<td>absolutely certain</td>
</tr>
</tbody>
</table>

Take all the time you need to do the very best you can, and do not leave any drawings out. Please, do not confer with the other judges during the judging session.

THANK YOU FOR YOUR PARTICIPATION
APPENDIX M

INSTRUCTIONS AND RATING MATERIALS
FOR ARTISTS' EVALUATIONS OF
AESTHETIC QUALITY
INSTRUCTIONS AND RATING MATERIALS FOR ARTISTS' EVALUATIONS OF AESTHETIC QUALITY

You will be examining 63 drawings done by adult females. Please examine each drawing carefully and come to a decision regarding the artistic quality of each. Rate the quality using the following scale:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor Quality</td>
<td>Average Quality</td>
<td>Good Quality</td>
<td></td>
</tr>
</tbody>
</table>

Please note your rating beside the appropriate drawing number on the attached form.
<table>
<thead>
<tr>
<th>Drawing Number</th>
<th>Quality Rating</th>
<th>Drawing Number</th>
<th>Quality Rating</th>
<th>Drawing Number</th>
<th>Quality Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td></td>
<td>26</td>
<td></td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>02</td>
<td></td>
<td>27</td>
<td></td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td></td>
<td>28</td>
<td></td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td></td>
<td>29</td>
<td></td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>05</td>
<td></td>
<td>30</td>
<td></td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td></td>
<td>31</td>
<td></td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>07</td>
<td></td>
<td>32</td>
<td></td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>08</td>
<td></td>
<td>33</td>
<td></td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>09</td>
<td></td>
<td>34</td>
<td></td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>35</td>
<td></td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>36</td>
<td></td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>37</td>
<td></td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>38</td>
<td></td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX N

SUMMARY LETTER SENT TO INTERESTED MOTHERS
Dear Research Participant,

As you may recall, you participated as a volunteer in a research study which was conducted in the fall and winter of 1982. You were asked to make a drawing of your family and to complete a questionnaire about your background. The purpose of the study was to investigate whether the drawings and background information of mothers who had abused their children were different from the drawings and social information produced by mothers who had not abused their children. Additionally, the study examined factors such as clinical experience and psychological test training, which might affect the ability of evaluators to differentiate the drawings accurately.

In the first phase of the project, drawings and social history data were collected from a total of 63 mothers. There were 42 mothers who, like you, served as "control mothers." That is, these were mothers who had not had contact with legal authorities regarding a charge of abusing/neglecting their children. In addition, data were collected from 21 mothers who were known child abusers. After the data were collected, names and identifying information were removed from all drawings and questionnaires. Next, the drawings and background information were evaluated by four different groups of "judges." The judges' task was to try to determine which mothers were abusing mothers and which were not, based on the drawings and questionnaires. The judges varied in terms of their clinical experience with child-abusing parents and their training in psychological testing. Some judges were "experts" who had worked professionally in this area for many years. In contrast, some judges were completely naive and had had neither relevant work experience in the area of child abuse nor training in psychological testing. In the final phase of the study, the drawings were rated by three professional artists, in terms of artistic quality, and the drawings were also scored using a formal scoring system.

The results of this research were as follows:

1. The judges were not able to differentiate between the family drawings done by mothers who had abused their children and those done by mothers who had not abused their children.

2. The judges were able to make accurate judgments about abusing versus nonabusing mothers on the basis of social history information.

3. The expert judges were not more accurate than the untrained judges.

4. Specific training in the interpretation of the drawings did not improve judges' accuracy.

5. Judges were able to make more accurate decisions when they judged social history information in combination with drawings, than when they judged drawings alone.
(6) The mothers' groups did not differ in terms of the artistic quality of their drawings.

(7) Abusing mothers tended to omit family members from their drawings more often than nonabusing mothers.

I am very grateful to you for the part you played in the completion of this research. Your time and interest were greatly appreciated. If you have any questions about the study, please feel free to contact me at 577-9300.

Sincerely yours,

Pamela Arnaut
APPENDIX O

SUMMARY LETTER SENT TO INTERESTED JUDGES
SUMMARY LETTER SENT TO INTERESTED JUDGES

Dear Research Participant,

As you may recall, you participated as a subject in a research project which was being conducted this past fall. The purpose of the study was to investigate (1) the validity of the Kinetic Family Drawing Technique as an assessment instrument in child-abuse evaluations; (2) some factors which might influence the ability to differentiate accurately the Kinetic Family Drawings of abusing and nonabusing mothers.

The results of the research were as follows:

(1) Judges were not able to differentiate between the drawings of abusing and nonabusing mothers.

(2) Judges were not able to differentiate between the drawings of Abusing versus Control versus Concerned mothers.

(3) Judges were able to make accurate judgments about the mothers' group membership on the basis of social history information.

(4) Using the drawings alone, the group of judges who were Speech Communication majors had the highest accuracy levels. The group of judges who were clinicians had the lowest accuracy levels. The P.A.C.T. counselors and clinical psychology students achieved accuracy levels which fell between those of the speech students and the clinicians.

(5) Training in the interpretation of drawings did not improve judges' accuracy.

(6) Judgments based on drawings plus social history were more accurate than judgments based on drawings alone.

(7) The mothers' groups did not differ in terms of the artistic quality of the drawings. Only one of the 12 judges appeared to be influenced by artistic quality in formulating judgments of mothers' group membership.

(8) Abusing mothers tended to omit family members from their drawings more often than nonabusing mothers. A higher percentage of mothers in the concerned group depicted barriers between themselves and another figure in their drawings.

I am very grateful to you for the part you played in the completion of this research. Your time and interest were greatly appreciated. If you have any questions about the study, please feel free to contact me at 577-9300.

Sincerely yours,

Pamela Arnaut
APPENDIX P

BASE-RATE SCORES ON THE 27 K-F-D
SCORING GUIDE VARIABLES
FOR 63 MOTHERS
BASE-RATE SCORES ON THE 27 K-F-D SCORING
GUIDE VARIABLES FOR 63 MOTHERS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance between mother and target child</td>
<td>$\bar{X} = 2.45$ inches</td>
</tr>
<tr>
<td>Mean distance between mother and all figures</td>
<td>$\bar{X} = 2.29$ inches</td>
</tr>
<tr>
<td>Height of mother figure</td>
<td>$\bar{X} = 2.75$ inches</td>
</tr>
<tr>
<td>Height of target child</td>
<td>$\bar{X} = 1.47$ inches</td>
</tr>
<tr>
<td>Ratio of mother to target child</td>
<td>$\bar{X} = 2.2:1$</td>
</tr>
<tr>
<td>Barriers present between figures</td>
<td>45/63 = 71.4%</td>
</tr>
<tr>
<td>Balls present</td>
<td>16/63 = 25.4%</td>
</tr>
<tr>
<td>Fire present</td>
<td>14/63 = 22.2%</td>
</tr>
<tr>
<td>Electrical appliances present</td>
<td>20/63 = 31.7%</td>
</tr>
<tr>
<td>X's present</td>
<td>3/63 = 4.8%</td>
</tr>
<tr>
<td>Pencil erasures present</td>
<td>40/63 = 63.5%</td>
</tr>
<tr>
<td>Arm extensions present</td>
<td>48/63 = 76.2%</td>
</tr>
<tr>
<td>Actions strange/unreal unexpected</td>
<td>5/63 = 7.9%</td>
</tr>
<tr>
<td>Figures distorted</td>
<td>6/63 = 9.5%</td>
</tr>
<tr>
<td>Figures in dangerous position</td>
<td>2/63 = 3.2%</td>
</tr>
<tr>
<td>Figures missing essential body parts</td>
<td>42/63 = 66.7%</td>
</tr>
<tr>
<td>Rotation</td>
<td>15/63 = 23.8%</td>
</tr>
<tr>
<td>Shading</td>
<td>33/63 = 52.4%</td>
</tr>
<tr>
<td>Compartmentalism</td>
<td>8/63 = 12.7%</td>
</tr>
<tr>
<td>Folding</td>
<td>0/63 = 0 %</td>
</tr>
<tr>
<td>Underlining of individual figures</td>
<td>0/63 = 0 %</td>
</tr>
<tr>
<td>Lining at bottom of page</td>
<td>4/63 = 6.3%</td>
</tr>
<tr>
<td>Lining at top of page</td>
<td>0/63 = 0 %</td>
</tr>
<tr>
<td>Variable</td>
<td>Score</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Encapsulation</td>
<td>13/63 = 20.6%</td>
</tr>
<tr>
<td>Edge placement</td>
<td>1/63 = 1.6%</td>
</tr>
<tr>
<td>Evasions</td>
<td>36/63 = 57.1%</td>
</tr>
<tr>
<td>One or more figures omitted</td>
<td>9/63 = 14.3%</td>
</tr>
<tr>
<td>One or more figures added</td>
<td>13/63 = 20.6%</td>
</tr>
<tr>
<td>Correct number depicted</td>
<td>41/63 = 65.1%</td>
</tr>
</tbody>
</table>
APPENDIX Q

PERCENTAGES OF DRAWINGS ELICITED BY FOUR EXAMINERS WHICH CONTAINED K-F-D FEATURES
Percentages of Drawings Elicited by Four Examiners Which Contained K-F-D Features

<table>
<thead>
<tr>
<th>Variable</th>
<th>E₁</th>
<th>E₂</th>
<th>E₃</th>
<th>E₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mom Present</td>
<td>100%</td>
<td>95.2%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>T/C Present</td>
<td>95.2</td>
<td>90.5</td>
<td>90.9</td>
<td>90</td>
</tr>
<tr>
<td>Barriers</td>
<td>85.7</td>
<td>81</td>
<td>45.5</td>
<td>50</td>
</tr>
<tr>
<td>Balls</td>
<td>33</td>
<td>19</td>
<td>27</td>
<td>20</td>
</tr>
<tr>
<td>Fire</td>
<td>33</td>
<td>14</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>Appliances</td>
<td>42.9</td>
<td>28.6</td>
<td>36.4</td>
<td>10</td>
</tr>
<tr>
<td>X's</td>
<td>0</td>
<td>9.5</td>
<td>9.1</td>
<td>0</td>
</tr>
<tr>
<td>Erasures</td>
<td>76.2</td>
<td>57.1</td>
<td>54.5</td>
<td>60</td>
</tr>
<tr>
<td>Arms Extended</td>
<td>81</td>
<td>76.2</td>
<td>81.8</td>
<td>60</td>
</tr>
<tr>
<td>Action Strange</td>
<td>14.3</td>
<td>9.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Distortion</td>
<td>14.3</td>
<td>4.8</td>
<td>9.1</td>
<td>10</td>
</tr>
<tr>
<td>Danger</td>
<td>9.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing Parts</td>
<td>76.2</td>
<td>57.1</td>
<td>63.6</td>
<td>70</td>
</tr>
<tr>
<td>Rotation</td>
<td>28.6</td>
<td>28.6</td>
<td>18.2</td>
<td>10</td>
</tr>
<tr>
<td>Shading</td>
<td>61.9</td>
<td>36.1</td>
<td>63.6</td>
<td>50</td>
</tr>
<tr>
<td>Compartmentalism</td>
<td>14.3</td>
<td>9.5</td>
<td>9.1</td>
<td>20</td>
</tr>
<tr>
<td>Folding</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lining Under</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lining Bottom</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Lining Top</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Encapsulation</td>
<td>14.3</td>
<td>14.3</td>
<td>27.3</td>
<td>40</td>
</tr>
<tr>
<td>Edging</td>
<td>4.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Evasion</td>
<td>57.1</td>
<td>38.1</td>
<td>81.8</td>
<td>70</td>
</tr>
<tr>
<td>Omitted Figures</td>
<td>9.5</td>
<td>4.8</td>
<td>36.4</td>
<td>20</td>
</tr>
<tr>
<td>Added Figures</td>
<td>23.8</td>
<td>9.5</td>
<td>36.4</td>
<td>20</td>
</tr>
<tr>
<td>Correct Figures</td>
<td>66.7</td>
<td>85.7</td>
<td>27.3</td>
<td>60</td>
</tr>
</tbody>
</table>

Note. Underlined figures indicate percentages which deviated from the mean by 16% or more.
APPENDIX R

MEANS FOR K-F-D FEATURES OF DRAWINGS ELICITED BY FOUR EXAMINERS
### MEANS FOR K-F-D FEATURES OF DRAWINGS ELICITED BY FOUR EXAMINERS

<table>
<thead>
<tr>
<th>Variable</th>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$E_3$</th>
<th>$E_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mom to T/C</td>
<td>2.98</td>
<td>2.09</td>
<td>2.31</td>
<td>2.15</td>
</tr>
<tr>
<td>Mom to Other</td>
<td>2.43</td>
<td>2.25</td>
<td>2.28</td>
<td>2.05</td>
</tr>
<tr>
<td>Mom Ht.</td>
<td>2.69</td>
<td>3.08</td>
<td>2.25</td>
<td>2.72</td>
</tr>
<tr>
<td>Child Ht.</td>
<td>1.5</td>
<td>1.38</td>
<td>1.76</td>
<td>1.26</td>
</tr>
<tr>
<td>Ratio</td>
<td>2.2:1</td>
<td>2.6:1</td>
<td>1.4:1</td>
<td>2.2:1</td>
</tr>
</tbody>
</table>

**Note.** Numbers indicate measurements in inches.
REFERENCES


VITA AUCTORIS

Pamela Susan Howitt, daughter of John Howitt and the late Vera Howitt, was born June 29, 1953, in Guelph, Ontario. Following graduation from Grand River Collegiate, Kitchener, Ontario, she attended Brock University in St. Catharines, Ontario, where she was the recipient of In-Course Scholarships and the Rose Horne Insurance Scholarship. In 1975, she graduated from Brock University with an Honours B.A. in Psychology, and was awarded the Governor General's Medal. She entered the graduate program in clinical psychology at the University of Windsor, and held University of Windsor Scholarships and Ontario Graduate Scholarships for the years 1975-1979. In 1977, she received a Master of Arts degree in Psychology. In 1979, she elected to continue her doctoral studies on a part-time basis, and accepted a staff position with the Wayne County Clinic for Child Study in Detroit, Michigan. Currently, she is employed at the Clinic as a staff psychologist.