



5-1-1995

Mothers' and Fathers' Responses on the Child Behavior Checklist: Are There Systematic Differences?

Ralph E. Casazza

Follow this and additional works at: <https://commons.und.edu/theses>

Recommended Citation

Casazza, Ralph E., "Mothers' and Fathers' Responses on the Child Behavior Checklist: Are There Systematic Differences?" (1995). *Theses and Dissertations*. 2727.
<https://commons.und.edu/theses/2727>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact und.common@library.und.edu.

MOTHERS' AND FATHERS' RESPONSES ON THE CHILD BEHAVIOR
CHECKLIST:
ARE THERE SYSTEMATIC DIFFERENCES?

by

Ralph E. Casazza
Bachelor of Arts, State University of New York, 1971
Master of Business Administration, Long Island University,
1978
Master of Science, California State University, 1992

A Dissertation
Submitted to the Graduate Faculty
of the
University of North Dakota
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Grand Forks, North Dakota
May
1995

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

UMI Number: 9709987

UMI Microform 9709987
Copyright 1997, by UMI Company. All rights reserved.

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Copyright by
Ralph E. Casazza
1995

ii

This dissertation, submitted by Ralph E. Casazza in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done, and is hereby approved.

George A. Jenby
(Chairperson)

Ray Kowbe PhD

Charles R. Baskin, PhD

Blank

Ronald J. Gyle

This dissertation meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

Harvey Kull
Dean of the Graduate School
May 15, 1995

PERMISSION

Title Mothers' and Fathers' Responses on the Child Behavior Checklist: Are there Systematic Differences?

Department Department of Counseling Psychology

Degree Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the Chairperson of the Department or the Dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Signature Roger Cagg
Date 5/10/95

TABLE OF CONTENTS

LIST OF FIGURES.....v

LIST OF TABLES.....ix

ACKNOWLEDGEMENTS.....x

ABSTRACT.....xi

CHAPTER

1. INTRODUCTION.....1

2. REVIEW OF THE LITERATURE.....4

 Factors Influencing Parental Ratings.....5

 Parent Characteristics Influencing Ratings.....9

 Child Characteristics Influencing Ratings.....15

 Interaction Effects Between Parent and
 Child Variables on Parent Ratings.....19

 Statistical Considerations Regarding
 Interrater Reliability.....28

 Summary.....38

3. METHOD.....38

 Statement of the Problem.....38

 Research Questions.....38

 Subjects.....39

 Design.....40

 Instruments.....41

Procedure.....	43
Data Analysis.....	43
4. RESULTS.....	48
Differences Between Mothers and Fathers.....	48
Reliability of Parent Ratings.....	52
Diagnostic Classification Agreement.....	64
Item Objectivity and Reliability of Ratings....	66
Interaction Between Parent and Child Gender....	67
5. DISCUSSION.....	70
Conclusions.....	73
Limitations.....	75
Directions for Future Research.....	77
APPENDICES.....	78
APPENDIX A: Child Behavior Checklist.....	79
APPENDIX B: Objectivity Rating Questionnaire.....	84
REFERENCES.....	88

LIST OF FIGURES

Figure	Page
1. Histogram of signed score differences; Scale 1 (Withdrawn).....	53
2. Scatterplot of mother vs. father scores; Scale 1 (Withdrawn).....	53
3. Histogram of signed score differences; Scale 2 (Somatic Complaints).....	54
4. Scatterplot of mother vs. father scores; Scale 2 (Somatic Complaints).....	54
5. Histogram of signed score differences; Scale 3 (Anxious/Depressed).....	55
6. Scatterplot of mother vs. father scores; Scale 3 (Anxious/Depressed).....	55
7. Histogram of signed score differences; Scale 4 (Social Problems).....	56
8. Scatterplot of mother vs. father scores; Scale 4 (Social Problems).....	56
9. Histogram of signed score differences; Scale 5 (Thought Problems).....	57
10. Scatterplot of mother vs. father scores; Scale 5 (Thought Problems).....	57
11. Histogram of signed score differences; Scale 6 (Attention Problems).....	58
12. Scatterplot of mother vs. father scores; Scale 6 (Attention Problems).....	58
13. Histogram of signed score differences; Scale 7 (Delinquent Behavior).....	59

14.	Scatterplot of mother vs. father scores; Scale 7 (Delinquent Behavior).....	59
15.	Histogram of signed score differences; Scale 8 (Aggressive Behavior).....	60
16.	Scatterplot of mother vs. father scores; Scale 8 (Aggressive Behavior).....	60
17.	Histogram of signed score differences; Internalizing broadband.....	61
18.	Scatterplot of mother vs. father scores; Internalizing broadband.....	61
19.	Histogram of signed score differences; Externalizing broadband.....	62
20.	Scatterplot of mother vs. father scores; Externalizing broadband.....	62

LIST OF TABLES

Table	Page
1. Age Distribution of Sample by Gender.....	40
2. Mother and Father CBCL Scores: All Children.....	49
3. Differences Between Ratings of Mothers and Fathers...	50
4. Intraclass Correlation Reliability Estimates for Parent CBCL Scale Scores.....	63
5. Diagnostic Classification of Mother (M) and Father (F) Ratings.....	65
6. CBCL Scale Item Objectivity Ratings, Kappa Coefficients, and Reliability Estimates.....	67
7. Mother and Father CBCL Scores by Child Gender.....	68
8. Differences Between Mother and Father Ratings by Child Gender.....	69

ACKNOWLEDGEMENTS

The author would like to thank George Henly, Ph.D., for his enthusiastic involvement in this study. With his guidance, this research incorporated new and innovative methods for the study of interrater agreement.

The author is also grateful to Eileen Searcy, Ph.D., UND Rehabilitation Hospital for allowing, and encouraging this research to proceed as smoothly as it did.

ABSTRACT

This study examined parental responses on the Child Behavior Checklist (CBCL), and considered variables that may affect rating reliability. Subjects for the study were 82 matched parent-pairs who completed CBCL protocols on their clinic-referred children. Research variables included parent gender, child gender, and level of item objectivity. The data was analyzed at the CBCL scale (narrowband) and broadband (Externalizing and Internalizing) levels, utilizing analysis of mean differences, repeated-measures analysis of variance, and clinical classification agreement.

Results showed little or no mean difference between mothers' and fathers' scores, but a large degree of discrepancy for parent-pairs was evident. None of the associated ANOVA *F*-tests were statistically significant. On analyses that investigated parent-gender by child-gender interaction effects, no significant gender differences were found between mother vs. father discrepancies.

Diagnostic classification analyses revealed a tendency for mothers to classify children in the clinical range more often than fathers, although differences were not statistically

significant. Kappas and Occurrence Agreement Indices were low to moderate (κ ranged from .30 to .55, and OAIs ranged from .28 to .53, respectively). Regarding item objectivity, Externalizing Scales (composed of items rated as more objective or observable), demonstrated greater parental agreement (higher kappas) and higher reliabilities than Internalizing Scales.

These results revealed no systematic differences between mother and father ratings, and therefore continued use of common norms is supported. However, the low to moderate reliabilities and indices of agreement suggest that the CBCL be used with caution when making diagnostic or treatment decisions.

CHAPTER 1

INTRODUCTION

Child behavior rating scales and checklists have seen increased use in both clinical and research settings. These instruments ask a knowledgeable person (e.g., parent, therapist, teacher) to rate the extent of a child's behavioral difficulties across a variety of domains (Achenbach & Edelbrock, 1983; Kazdin, French, & Unis, 1983; Quay, 1977, McDevitt & Carey, 1978).

Among the various rater groups, parental perceptions of child behavior are especially important. Parents typically spend more time with the child than anyone else, and thus become the primary data source in many referrals to mental health clinics and special education services (Leblanc & Reynolds, 1989). Consistent with this importance, a number of investigators have examined the variables that influence parental perceptions or ratings. Most of these studies have focused on characteristics of the parent (Griest, Wells & Forehand, 1979; Webster, Stratton & Hammond, 1988), or characteristics of the child (Brody & Forehand, 1986; Schaughency & Lahey, 1985), although some have chosen to look at other variables such as environmental or situational context (Alvarez, 1985; Duncan & Kilpatrick, 1990; Weinraub & Frankel, 1977).

Although the data gathered from these ratings has contributed significantly to both developmental and child clinical psychology, there are validity issues associated with the use of parents as raters, including the effects of parental values, biases, psychopathology, as well as situational and contextual factors (Marsh, Stoughton & Williams, 1985).

One concern has been the tendency to rely on mothers' ratings or evaluations of child behavior with a corresponding neglect of paternal perceptions (Conrad & Hammen, 1989; Phares, 1992; Richters & Pellegrini, 1989). This trend is reflected in the widespread use of maternal norms, or "parental" norms which combine mothers' and fathers' data in the standardization process. However, over the past three decades, rapid changes have occurred in family and societal structures, necessitating that fathers be considered as important informants in this research (Glick, 1988).

Another concern in the validity of behavioral checklist ratings is the nature and extent of differences between mothers' and fathers' reports of children's behavioral problems. This is basically an issue of parent reliability or agreement. If systematic and consistent parent differences exist, it may be appropriate to call for the development of separate mother/father norms in order to increase the accuracy and validity of behavioral checklists.

In this study, questions about differential parent perceptions/ratings are addressed. For example, do mothers or fathers tend to overreport child behavior problems in comparison to the other parent? Does parental agreement differ with the gender of the child? How does the statistical method used to measure interparent agreement affect the outcome of research studies? To date, a relatively small amount of research has addressed these questions, with mixed results and equivocal conclusions. Furthermore, much of the research is dated, and may have diminished relevance for contemporary societal applications. The present study attempts to extend and update our knowledge about these issues.

CHAPTER 2

LITERATURE REVIEW

There has been a proliferation of children's symptom checklists and behavior problem assessment instruments designed for completion by parents, teachers, therapists, or children (Jensen, Traylor, Xenakis, & Davis, 1988). Although a number of methodological and conceptual problems associated with their use are coming to light, these instruments have become an integral part of the diagnosis of emotional and behavioral disorders in children and adolescents.

Most notably, researchers have found disappointingly low correlations between parent and child self-report ratings (Kazdin & Esveldt-Dawson, 1983; Leon, Kendall & Garber, 1980). Similar difficulties have been found between different rater groups including teachers, parents, and clinicians, and across various settings (Christensen, Margolin, & Sullaway, 1992; Marsh, Stoughton, & Williams, 1985; Martin & Halverson, 1991).

For example, the interrater correlations for mothers and fathers reported in the Temperament Assessment Battery for Children (TABC) manual (Martin, 1988) generally ranged from .00 to .70, with a mean around .40. Jacobs, Grounds, and Haley (1982) reported correlations ranging from .18 to .73 across three scores on the Behavior Problem Checklist (Quay,

1977), and Lyon and Plomin (1981) found correlations between mothers and fathers on the EASI (Emotionality, Activity, Sociability, & Impulsivity Scale) ranging from .27 to .57. LeBlanc & Reynolds (1989) found parental correlations for the Child Behavior Checklist (Achenbach & Edelbrock, 1983) ranging from .15 to .80 with an average r of .35. In a brief review of parental agreement studies, Bates (1980) reported correlations between .06 and .69.

Even when stronger correlations between mothers' and fathers' ratings are found on certain scales or subscales, confidence in interparent reliability estimates may not be well founded because reliability estimates (usually based on the Pearson correlation coefficient) do not take into account the error variance between parents (Shrout & Fleiss, 1979). Thus, estimates may be spuriously high given the variability in factors that contribute to a parent's rating process such as gender, age, education, mental status, mood, etc. Of course, the instrument and its administration may also be subject to validity concerns, although these matters are not pursued here.

Factors Influencing Parental Ratings

When assessing the behavior of children, the use of behavioral symptom checklists completed by parent-raters in clinical or experimental conditions may introduce a degree of inaccuracy that doesn't exist with "observational" methods in

naturalistic settings. A number of potential variables might be expected to influence parental ratings such as age, sex, cognitive ability, socioeconomic status, characteristics of the behavior being rated, and rating "sets" that are unique to the parent.

Each of these influences is a potential source of error in the research effort. Some might be expected to contribute error in a more random fashion than others. For example, Burrows & Kelley (1983) found that parental ratings of children across a variety of different situations did not show a consistent pattern of differences in interparent agreement. On the other hand, variables such as parent psychopathology, and objectivity of items rated have revealed consistent differences in parent ratings of children's behavior in some studies (Christensen et al., 1992; Griest et al., 1979; Webster-Stratton & Hammond, 1988).

Demographic Factors

One problem that most clinicians and scientists who use behavioral checklists are aware of is that the rater and instrument can become commingled parts of the measurement process, and raters are known to be subject to a variety of influences including socioeconomic status, education, and mood. For example, maternal depression has been associated with lower interparent agreement on symptoms tapped by child behavior checklists (Johnston, 1991).

Another source of discrepancy between parents' ratings can be the characteristics of the child itself. For example, if parents value boys more than girls, as some research indicates (Peterson & Peterson, 1973), they might scrutinize boys more closely than girls, and discuss boys' behavior to a greater degree. Both the scrutiny and discussion could lead to greater parental agreement on sons' behavior versus daughters' behavior.

Societal Factors

Gender and parenting roles have been in a constant state of evolutionary flux over the last four decades, bringing a validity concern for the traditional reliance on maternal norms with child behavior ratings (Phares, 1992). Fathers have become increasingly active in child care and domestic responsibilities (Grief, 1985), yet unfortunately, previous studies often rely on maternal ratings exclusively (Brody & Forehand, 1986; Weissman, Orvaschel, & Padian (1980), or refer to "parental" ratings, when they are actually maternal ratings (Montemayor & Brownlee, 1987; Worobey, 1987). Nevertheless, enough research of fathers' influences on child and adolescent psychopathology has suggested that when fathers are included in studies, paternal effects are usually found (Earls, 1976; Lamb, 1975; Novick, Rosenfeld, Black & Dawson, 1966; Phares & Compas, 1992).

Concurrent with the evolutionary changes in parenting and sex roles, even the family structure itself has become quite variable. The traditional notion of a family composed of an employed father and a homebody mother who takes care of the children now constitutes less than 10% of American families (Richardson, 1988). More prevalent today is the dual-provider home, and various nontraditional household formations including single-parent and blended families.

In summary, there are numerous societal and demographic factors that can influence parental ratings, some of which may be consistent and systematic in nature. However, most child behavior scales depend on a restricted range of variables for consideration in the norming process. As an example, the Child Behavior Checklist (Achenbach & Edelbrock, 1983) consolidates mothers' and fathers' responses in the norm group, apparently assuming equivalence of mothers' and fathers' responses/perceptions.

As a partial response to these sociological changes, and the increasing importance of paternal involvement in domestic life, it is probably no longer appropriate to rely solely on mothers for information regarding family functioning and child psychopathology. Therefore, it may be prudent at this point in time to encourage separate norms for mothers and fathers if we are to interpret behavioral checklists accurately.

The balance of this review will examine research findings in four areas: (a) parent characteristics that influence

ratings of child behavior, (b) child characteristics that influence parental ratings, (c) interaction effects between parent and child variables on parental ratings, and (d) statistical considerations regarding interrater reliability, and item content objectivity.

Parent Characteristics Influencing Ratings

In looking at parental perceptions of child behavior, an important consideration is characteristics of the rater (Jacob, Grounds, & Haley, 1982). There are many relevant variables including parent gender, socioeconomic status, ethnic background, parent psychopathology, and the level of marital distress.

Parent Psychopathology

Several studies have demonstrated that, at least for mothers, depressive symptoms contribute unique variance, beyond that accounted for by observed child behavior, to perceptions of child problems (Griest et al., 1979; Johnston, 1991; Webster-Stratton & Hammond, 1988;). Christensen, Sullaway & King (1983) also found that happy couples had greater agreement than unhappy couples on behavioral checklists for dyadic interaction.

Webster-Stratton (1988) looked at the relationship of parental adjustment (using measures of depression, marital satisfaction, and parenting stress) to mothers' and fathers'

perceptions of their children's deviant behaviors. Two measures of child adjustment were used, the CBCL, and the Eyberg Child Behavior Inventory (ECBI; Robinson, Eyberg, & Ross, 1980). The ECBI yields a Total Problem score that permits the parent to indicate with a yes/no response whether the behavior is a problem for the parent, and an Intensity score that permits the parent to rate on a scale of 1-7 the frequency of the behavior problem.

On the CBCL, Webster-Stratton observed no significant differences between mothers' and fathers' reports; mothers' and fathers' perceptions of child deviance were significantly correlated (highest t value=1.67, $p>.001$ with experimentwise error controlled at .05 level) on the CBCL Internalizing, Externalizing, and Depression scales but not on the CBCL Hyperactive scale. On the ECBI, mothers reported significantly more behavior problems ($p<.001$) that they were concerned about, as well as more frequent occurrences of the problems, than did their husbands.

Webster-Stratton (1988) also found that father reports correlated well with teacher reports, but mother reports did not. Father reports were relatively less contaminated by personal adjustment measures. On the other hand, mother reports of child misbehavior were more highly correlated with maternal depression scores, and stress due to marital problems. The author concluded by calling for more research that includes father reports, and for increased understanding

of the paternal "perspective" in order to help serve families with problem children. Unfortunately, gender of child was not a variable in this study, so we do not know if there were any child-gender, or parent-gender by child-gender interaction effects.

Johnston (1991) examined the contributions of parental depression and child behavior to predicting mothers' and fathers' perceptions of clinic-referred children. All families in the sample were seeking assistance in reducing externalizing child problems such as noncompliance, hyperactivity and inattention. Although paired *t*-tests for CBCL Internalizing and Externalizing scores revealed no significant inter-parent differences, results supported the hypothesis that maternal perceptions of child problems were predicted both by the measure of child behavior (CBCL), and by mothers' report of their own depressive symptoms (BDI; Beck, 1967).

In Johnston's study, mothers with fewer depressive symptoms showed greater agreement with fathers ($r=.63$ and $.72$ for Externalizing and Internalizing scores respectively), than mothers with more depressive symptoms ($r=.28$ and $.32$). The depressed mothers also had more negative perceptions of their children's behavior. This was not the case for fathers' perceptions, where only child behavior accounted for significant, unique variance. Marital adjustment, as measured by the Short Marital Adjustment Test (SMAT; Kimmel &

Vanderveen, 1974) did not contribute to either mothers' or fathers' perceptions. Gender of child was not examined as a variable in this study.

By contrast, other studies (Conrad & Hammen, 1989; Richters & Pellegrini, 1989) have suggested that, although parental depression is associated with elevated reports of child problems, these reports are accurate, and not negatively influenced by the depression. In the Conrad & Hammen (1989) study, 64 children of unipolar or bipolar mothers, chronic medically ill, and normal mothers comprised the sample. Mothers completed the CBCL, and also participated in a 5 minute videotaped discussion with their child on a topic which usually involved disagreement and conflict in order to assess the number of negative utterances made by the mother.

Hierarchical multiple regression analyses were utilized by Conrad and Hammen (1989) to analyze experimental data, and results indicated that there was no overall tendency for depressed mothers to perceive their children as more maladaptive than other children. In fact, depressed mothers were found to be more accurate when children did present with symptoms. The authors concluded that a possible reason for dysphoric mothers to be more accurate in their perceptions was related to the construct of "depressive realism", in which nondepressed adults are postulated to hold rosier views than warranted by reality. It also may be that a parent's own

painful symptoms of dysphoria make them more sensitive to maladaptive behaviors in their children.

Parental/Gender Role

Lyon and Plomin (1981) looked at two issues that concern adequacy of parental ratings of young children's temperament; (a) parent gender, and (b) the extent to which parents project their own personality in ratings of children. Using a sample consisting of 137 families with twin children, the EASI Temperament Survey was employed to measure child traits. In addition to the parental rating of young children's temperament, the EASI includes a parallel self-report and rating form for adults.

Results of the Lyon and Plomin (1981) study revealed that parents do not substantially project their own personality in their ratings of children. The average parent-offspring correlation for the 11 EASI scales was .17. The largest discrepancy between mothers' and fathers' parent-offspring correlations was found for the quality or affectionateness component of sociability. This suggests that mothers are more prone to rate the affection of their children in terms of their own affectionateness. The extent of parent agreement was only moderate at best, with an average correlation of .51. Effects of child gender were not considered in this study.

Eisenstadt, McElreath, Eyberg, and McNeil (1994) looked at the effects of parent gender on interparent agreement using

the Eyberg Child Behavior Inventory (ECBI; Eyberg, 1974). Subjects were parents of 44 children referred to three university-based psychology clinics for treatment of disruptive behavior problems.

Although both Intensity and Problem Scales of the ECBI demonstrated adequate interparent reliability (Pearson r ranged from .50 to .87), significant mean differences between maternal and paternal reports were found. Mothers rated their children's disruptive behavior as more frequent and more problematic than did fathers. Eisentadt et al. concluded that discrepancies between parental reports may stem from mothers serving as primary caregivers, and therefore having more exposure to their children's problem behaviors. They also suggested that behavior problem children are more likely to obey fathers than mothers, resulting in fathers being less likely to see disruptive behavior. Effects of child gender were not considered in this study.

Burrows & Kelley (1983) looked at parental agreement in conjunction with two variables: situational specificity, and familiarity (biological offspring vs. unknown) with the child. Agreement was computed under several experimental conditions including (a) ratings of a parent's own child's behavior in general, using the Children's Behavior Checklist (CBC; Ferguson, Partyka, & Lester, 1974), (b) ratings of own child based on observations of him/her in a videotaped sample, and

(c) ratings of an unknown child based on observations in a videotaped sample.

Parents achieved fairly high rates of agreement (mean interrater reliability rating across experimental conditions=.81) in Burrows & Kelley's (1983) study, using an interrater index computed by dividing the number of agreements by agreements plus disagreements for each parent pair.

Increasing the specificity of the behaviors rated in Burrows and Kelley's study did not significantly affect agreement, and those parents who agreed the most did not necessarily spend a large amount of time in the same kind of situations with their child. Agreement was significantly greater ($F=9.48$, $p<.01$) when parents rated their own children's videotaped sample as opposed to that of an unknown child. A conclusion offered by the authors is that high parental agreement appears to be derived from a shared familiarity with their child's behavior across a variety of situations.

Child Characteristics Influencing Ratings

Characteristics of the child may be related to variation in parental agreement also. A variety of child attributes, including temperament, cognitive abilities, gender, age, etc. can impact upon parental ratings of the child's behavior, and family functioning in general. According to Patterson (1980), child behavior problems can undermine parental confidence and

effectiveness, color parental perceptions or behavior toward the child, and result in significant parental stress.

Child Psychopathology

One interesting finding was reported by Pfeffer and Martin (1983). Samples composed of "normal" preschool and elementary school children had higher levels of parental agreement (r ranged from .4 to .5) than samples referred for psychological evaluation due to parental concern (r generally not significantly different from 0.0).

Pfeffer and Martin used a modified version of the Parent Temperament Questionnaire, and found only one main effect for parental gender, on the "Threshold" scale which is designed to determine the level of stimulation required to produce a response from the child. Fathers rated threshold consistently higher (in the direction of less sensitivity) than mothers.

In contrast to Pfeffer's findings, Duncan, Gale and Kilpatrick (1990) found "good agreement" ($r=.65$) for mothers and fathers of emotionally disturbed children placed in a residential treatment center. The authors used the CBRS as a rating instrument, and parents tended to show slightly better agreement for boys (average $r=.69$) than for girls (average $r=.58$).

Konstantareas and Homatidis (1989) examined parental perception of learning-disabled (LD) children's adjustment problems using the CBCL. In this study, the authors

postulated that parental stress, locus of control, and self-concept were characteristics thought to correlate with the perception of a child's dysfunctional behavior. Findings indicated that although mothers did not differ from fathers in behavior ratings, there was a child gender effect, suggesting that female LD children are seen as displaying fewer behavior problems than males.

An interesting outcome of Konstantareas and Homatidis' study involved mothers reporting greater stress in response to their children's behavior. This was particularly true of mothers of middle or upper socioeconomic status with an external locus of control. Konstantareas and Homatidis state that this finding is corroborated by research (Bristol, Gallagher, & Schopler, 1988) reporting that fathers tend to be increasingly less involved with the disabled child as the severity of the handicap increases. Considering the likelihood of being entrusted with more child-rearing responsibilities vis-à-vis LD children, it may not be surprising that mothers, particularly those with an external locus of control, feel more distressed.

Kazdin, French, and Unis (1983) looked at the extent to which children, mothers and fathers agree on the children's psychological symptoms. In this study, 48 children who were in-patients of a psychiatric intensive care facility comprised the sample. The Children's Depression Inventory (CDI; Kovacs, 1981), Bellevue Index of Depression (BID; Petti, 1978), and

the Depression Symptom Checklist (DS-CL; Weinberg, Rutman, Sullivan, Penick, & Dietz, 1973) were used to examine child and parent demographic variables on depression scores including child age, race, gender, IQ; mother and father age, race, family Hollingshead (1975) class, income level, welfare status, marital status (married or not), relationship of rater to the child (natural parent vs. other), and living arrangement (living together or not). Pearson product-moment correlations, and means and standard deviations were computed to measure correspondence of parental ratings.

Kazdin et al. found few significant effects, probably due to the small sample and large number of variables considered. Nevertheless, the authors reported the following trends: (a) measures completed by the children tended not to correlate with the same or related measures completed by their parents, (b) different measures of depression completed by the same rater (child, mother, or father) were moderately to highly correlated with each other; (c) mothers' and fathers' ratings of children's depression tended to correlate moderately to highly with each other; and (d) no significant effects for child gender emerged on measures completed by mothers or fathers.

Martin and Halverson (1991) studied mothers' and fathers' temperament ratings of their children using the Temperament Assessment Battery for Children-Parent Form (TABCP; Martin, 1988). They found that in normal samples, mothers tend to

rate positive temperamental characteristics more positively than did fathers, and in samples of more "troubled" children, mothers tend to rate negative characteristics more negatively than did fathers. These results led them to argue that mothers are more discriminating raters than fathers. Martin and Halverson further speculated that mothers may have more experience with children, especially in the early years (infancy through grade school), and have a better knowledge of the typical behavior of other children so that comparative judgements are made more accurately.

Interaction Effects Between Parent and Child

Variables on Parent Ratings

The reasons underlying low to moderate agreement between parents' ratings of their own children are probably complex and involve many variables. One possibility may be that mothers and fathers differ in the way they act toward their children, depending upon the sex of the child. In other words, mothers and fathers may hold different expectations of their sons and daughters based upon societal sex-typing of boys and girls, among other possibilities.

In reviewing the literature, it has been difficult to assess such cross-sex effects because of myriad experimental approaches to the problem, the lack of father data in many studies, and the lumping of offspring into a "child" classification rather than accounting for boys and girls

separately. Nevertheless, some research has included parent and child gender as variables, and can be sorted into two groups: (a) severity of parental ratings or (b) nature of sex-typing observed.

Severity of Parental Ratings

An interesting phenomenon in research that concentrates on behavioral differences between boys and girls is that boys usually display more problems than girls (Duncan & Kilpatrick, 1991; Lindholm & Touliatos, 1981; Speer, 1971). In a study examining the effect of sex of rater and sex of child upon interrater agreement, Kilpatrick and Duncan (1985) found that girls were seen as being better adjusted than boys by both male and female child care workers when differences did occur.

The authors used the CBRS as a rating instrument, with a sample of children in a residential treatment center. Results also showed that female child care workers were more consistent with each other than male-female rater pairs, especially when rating girls.

Jensen et al. (1988) examined the effects of parent and child gender in conjunction with parental psychiatric symptoms on the reliability and agreement of children's and parents' reports of children's behavior problems, using the CBCL. The authors used a military sample, consisting of 90 elementary-school children from intact families, all of whom were headed

by officers or senior enlisted personnel on active duty with the U.S. Army.

Parents' mean ratings differed significantly in this study (Jensen et al., 1988), with mothers reporting more problem behavior in both sons and daughters. The relationships between parental symptoms and their CBCL responses were less clear; Pearson r s were moderate and ranged between .31 to .59 (mean r = .46). The authors suggested that mothers' reports of daughters are less affected by mothers' symptoms than are reports on sons. Fathers' symptoms explained a larger portion of the variance between parental reports.

Lindholm & Touliatos (1981) also found differences when comparing mothers' and fathers' perceptions of their children's behavior problems. The Behavior Problem Checklist was utilized on a sample that consisted of 1008 white children enrolled in kindergarten through eighth grade in a suburban school district. The data analysis looked at differences between the means of mothers' and fathers' ratings. Two basic conclusions were drawn from the study: (a) mothers perceived more behavior problems in their children than fathers did, and (b) relationships between the parents' ratings were only low to moderate.

Moving to the subscale, or dimension level, as it is called in the BPC, the Lindholm and Touliatos study found that both parents reported more problems for boys than girls on the

Psychotic Signs dimension. On the Conduct Problems dimension, mothers noted more difficulties in boys than fathers did, and on the Personality Problems dimension, mothers perceived more difficulties with girls than fathers did.

Thompson, Curry, and Yancy (1979) looked at the utility of the Missouri Children's Behavior Checklist (Sines, Pauker, Sines, & Owen, 1969) in describing clinically relevant dimensions of behavior of children with developmental disabilities, and discriminating these children from normal controls. They also examined interparent reliability in this project, and found significant Pearson correlations on all seven MCBC scales for ratings of boys, and on five of seven for girls (Inhibition and Sex scales had $r=.31$ and $.23$ respectively). However, analysis of means and standard deviations resulted in the observation of significantly higher aggression ratings of boys by mothers than fathers.

On the other hand, some researchers have found better agreement between parents when rating children's behavior. For example, Thompson and McAdoo (1973) found consensus on the part of parents rating emotionally disturbed children using a revised version of the Missouri Children's Behavior Checklist. The *t*-test for correlated means was used to compare mothers' and fathers' ratings of their clinic-referred boys and girls. There were no significant differences between mothers' and fathers' ratings of boys on any of the scales, and only one significant difference occurred for ratings of girls

(Sociability), with fathers reporting more sociability than mothers. The authors also found that average ratings given boys were not significantly different than those given girls, except on the Sex scale where both mothers and fathers reported significantly more items for girls than for boys.

Finken and Amato (1993) also found no appreciable gender effects in a large ($n=1,624$), nationally representative study on the relationship between parental self-esteem and parents' perception of child behavior problems. The authors developed, and utilized a seven item questionnaire that surveyed various types of child behavior problems. MANCOVAs were performed with parent gender and child gender included among the control variables. Results revealed that behavior problems among children were associated with low self-esteem among parents, although none of the moderating variables had a significant impact. No appreciable impact was observed on the interaction of child problems and child gender, and there was no significant main effect for parental gender in any analysis.

Some interrater agreement research has yielded mixed findings for interaction effects between parent and child gender. For example, Wierson, Armistead, Forehand, Thomas and Fauber (1990) examined parent-adolescent conflict using adolescent age and gender, and parent gender as research variables. Mothers and fathers completed the Conflict Behavior Questionnaire (CBQ; Prinz, Foster, Kent, & O'Leary, 1979) which is a 75 dichotomous (yes/no) item instrument that

assesses conflict behavior between adolescents and their parents. The authors did not find significant effects for gender of adolescent or gender of adolescent by gender of parent in their study, supporting Steinberg's (1987) view that the case for sex of adolescent main effects has been overstated in recent gender research.

Wierson et al. did however, find a significant main effect for mothers scoring higher than fathers ($F=5.14$, $p<.05$), or reporting more conflict with their adolescent than fathers report. This result is similar to the findings of Montemayor (1982) that mothers argue more with their adolescents than do fathers. An explanation for such results is offered by the authors concluding that mothers may simply be more involved in parenting an adolescent than fathers. That is, it is not that the mother-adolescent relationship is more proportionately stressful or heated, but that mothers interact more frequently with their adolescents and, thus, are more likely to discuss topics which lead to conflict.

Duncan and Kilpatrick (1991) explored the similarities and differences in mother-father ratings of normal boys and girls, ages 6-12 years, on two different behavioral rating scales; the Behavior Problem Checklist (BPC; Quay, 1977) and the Child Behavior Rating Scales (CBRS; Duncan & Kilpatrick, 1975). Their results indicated that on the CBRS, mothers tend to rate their daughters more favorably than did fathers, but there was no difference in parental ratings of boys (hence, a

parent by child gender interaction took place). On the BPC there were no significant differences between parents' ratings of either boys or girls. These findings were expected by Duncan and Kilpatrick, as the 55 items on the BPC are more descriptive of problem behaviors that are characteristic of maladjusted children, whereas the CBRS contains items descriptive of both prosocial and problem behaviors.

Sex-Typing in Parental Ratings

Two studies (Meyer & Sobieszek, 1972; Rothbart & Maccoby, 1966) found that parents are better able to define the behavior of same-sex children, but are also inclined to be less tolerant of same-sex child behavior, especially in the areas of aggression and dependency. In the Rothbart and Maccoby (1966) study, parents' reactions to specific child behaviors were examined. Their method included a questionnaire adapted from Sheriffs and Jarrett (1953), and audiotaped samples of a child's voice that was ambiguous with respect to sex.

Parents were put in a hypothetical situation with a child, and asked to record their immediate reactions to what the child did and said. The questionnaire had two parts, one measuring parents' opinions about differences they felt actually existed between boys and girls, and one measuring absolute differences between ratings of an item's importance for girls and importance for boys (the higher the differences

score, the higher the sex-role differentiation that was indicated).

A pattern of results emerged in Rothbart and Maccoby's (1966) study with fathers showing generally more permissiveness toward girls than boys for both dependency and aggression (however, the only significant main effect was for "permissiveness for autonomy"). Mothers showed greater permissiveness toward boys than girls, although no discrepancies were significant. Parents' sex-role differentiation scores (as measured by the questionnaire) revealed that high-differentiation parents did not show a stronger tendency to promote stereotyped sex-role behavior in their responses to the child's voice.

Cohen, Dibble, and Grawe (1977) studied fathers' and mothers' perceptions of children's personality using a sample of families with twin children (monozygotic and dizygotic pairs were included). The authors used the Childhood Personality Scale (CPS; Dibble & Cohen, 1974). On most dimensions tapped by the CPS, results were consistent with predictions about sex-related or stereotypic perceptions of boys and girls. Girls were seen as far more talkative, placid, and socially introverted, while boys were perceived as more active, outgoing, and ebullient. On one dimension, attentiveness, parents' ratings went against this stereotypic trend, finding boys more attentive than girls. Differences between fathers and mothers were also found. In their

fathers' eyes, children were less interesting, less talkative, and less ebullient.

Leblanc and Reynolds (1989) studied parental responses on the CBCL for 200 adolescents attending a summer institute for gifted students. They found modest to low interparent correlations on most scales (r ranged from .15 to .24) with Delinquency (Scale VII) being the only major exception ($r=.80$). In addition, mothers' ratings of problems for sons were higher than the fathers' on every scale but Hostile Withdrawal (Scale VI). Mothers' ratings of problems for daughters were higher than the fathers' on every scale but Immature Hyperactive (Scale V).

The highest correlations in Leblanc and Reynolds' study were obtained for ratings of "acting out" or "externalizing" behavior of boys (i.e., Delinquent, Aggressive, Hyperactive). The authors concluded that there is a clear interaction between parent and child gender on parental ratings, especially given the diminished sensitivity of the CBCL with normal populations (Achenbach & Edelbrock, 1983).

In a study examining parents' perceptions of the ideal child, Paguio (1983) used the Ideal Child Checklist (ICC; Torrance, 1975), an instrument that contains 66 items representing personality characteristics that have been found in empirical studies to differentiate productive, creative persons from persons who are less productive and creative (Torrance, 1962; 1975). Paguio's sample consisted of 522

parents (295 mothers and 227 fathers) with one or more children, and all were of Caucasian descent and middle-class.

Results of Paguio's study failed to reveal differences in perceptions of the ideal child. Boys and girls were equally expected to be confident, aggressive, well-adjusted (Factor 1); socially virtuous (Factor 2); and creative, intuitive (Factor 4). Boys and girls were also equally discouraged to be negativistic and critical (Factor 3). The author did, however, find small but significant gender-of-parent effects on responses to the checklist. Mothers most strongly encouraged sociability and creativeness/intuitiveness. Also, mothers discouraged negativistic, critical behaviors more than fathers. Paguio offers an interpretation that mothers may have perceived the importance of these traits in interpersonal relationships due to their own developmental process and skillfulness in these areas (Kagan, 1984).

Statistical Considerations Regarding Interrater Reliability

Cronbach (1970) categorized statistical concerns that are inherent in rating instruments into three areas: (a) ambiguity in response alternatives; (b) ambiguity in traits or behaviors to be rated; and (c) judgmental errors or individual peculiarities of response (e.g., rater response sets, "halo" effects).

Choice of Interrater Index

A significant variable in measuring response sets and ambiguity in rated behaviors may be the choice of interrater index. Differences in measurement procedure can have an important effect upon the degree, nature, and meaning of parent agreement reported. Indeed, any index one adopts is likely to carry distinct validity threats and perhaps yield different configurations of results, as a function of the index's particular statistical properties.

Jacob et al. (1982) studied interparent agreement on the Behavior Problem Checklist (BPC), using a sample consisting of 96 families with male children between the age of 10-18. This sample was composed of "normal" (control) subjects, and "disturbed" subjects, the great majority of whom were recruited from community-based probation offices. The authors explored a variety of factors that may influence interparent agreement including item content, reliability index, diagnostic status of the sample, and selected demographic variables. Five indices of interrater agreement were used; correlation coefficients, *t*-tests, percentage agreements, effective percentage agreements, and absolute differences.

In this study, Jacob et al. found parent agreement to be (a) at moderate levels, (b) higher in nondistressed than in distressed samples, (c) lower for Personality Problem (PP) items than Conduct Problem (CP) items, (d) very low in both disturbed and control samples regarding presence of specific

behaviors, (e) weakly related to demographic variables after severity of disturbance was statistically controlled, and (f) strongly related to severity of disturbance with some reliability indexes but not others.

Jacob et al. concluded that alternative reliability indices produce different patterns of results, and propose that one index, absolute difference (AD), appears to be less affected by frequency of endorsement (level of chance agreement) than other indices. This characteristic of the AD index is proposed to be a more appropriate and sensitive measure of parental agreement with a factor-analytically derived rating instrument like the BPC. Child gender was not considered as a research variable.

Martin & Halvorsen (1991) examined the issue of parental agreement in their examination of four samples of children that varied by severity of clinical pathology, socioeconomic status, and referral status (referred or not referred). Child temperament was assessed using the TABC, and four indices were investigated in the study: (a) comparisons of differences in means and standard deviations of ratings provided by mothers and fathers for each sample; (b) correlation between ratings of mothers' and fathers' of the same characteristic (e.g., activity level) for each sample; (c) calculation of absolute differences between mothers' and fathers' ratings across all characteristics rated for each child; and (d) intraclass correlations for mothers' and fathers' ratings for each child.

Although Martin & Halvorsen found clear variation across samples and rating pairs in levels of agreement (no matter what index was used), the observed variation was determined by unique aspects related to each of the four rating situations or samples. This interpretation was supported by the lack of strong and consistent findings for any index across samples.

Hulbert, Gdowski, & Lachar (1986) looked at interparent agreement on the Personality Inventory for Children (PIC; Wirt et al., 1984) for clinically referred children and adolescents. They used several indices of agreement including Pearson correlations, matched *t*-tests, classifications by level of scale scores as defined by the PIC actuarial guide (Lachar & Gdowski, 1979), and calculations of the number of disagreements within each profile pair.

Hulbert et al. found that correlation coefficients were relatively high (average $r=.66$), compared with those for other similar instruments (in a 1987 meta-analysis, Achenbach, McConaughy, & Howell found an average inter-rater correlation of .59). Using clinically relevant indices determined by classification of the level of scale scores, and the number of disagreements within each parent-pair suggested that fathers did not observe the same, or as many negative behaviors as mothers. Results also revealed that tests of mean differences were of even less utility than correlation indices, since mean values obscured individual profile-pair differences that were of primary clinical interest. The authors concluded that

fathers did not observe the same, or as many behaviors as mothers, and that low or high correlations do not translate into poor or good correspondence between parental scale scores, or interpretation of test results based on these scores. Child gender was not considered as a research variable in this study.

Duncan & Kilpatrick (1989) examined the use of extreme responses on the Child Behavior Rating Scale (CBRS; Duncan & Kilpatrick, 1975) and the Behavior Problem Checklist by parents and child care workers. The samples differed in this study, depending on the instrument used, and whether the children being rated were normal or maladjusted. Two methods of analysis were applied to the data: (a) a frequency count was made of all items checked in the two extreme categories, and the proportion of male extreme positive or negative responses was compared with those of female raters, and (b) the proportion of female raters who met the criterion of "extreme rater" was tested against the proportion of male extreme raters, for each of the 16 scales of the CBRS, with the corrected chi square test for independent samples (Siegel, 1956).

Duncan & Kilpatrick found only partial support for their hypothesis that female raters would tend to be more extreme in their ratings of both prosocial and negative child behaviors. Mothers and fathers did not differ in the frequency with which they checked extreme negative responses on both the CBRS and

the BPC ratings. The authors concluded that attempts to attribute response bias in the use of extreme categories to a single variable such as sex of rater are too simplified. Other variables, such as intensity and type of relationship to the individual being rated, or the degree of maladjustment on the part of the ratee interact with the sex of rater variable. Child gender was not considered for interaction effects with sex of rater.

Objectivity of Item Content

As Cronbach (1970) noted, another statistical source of error is ambiguity in response alternatives and/or ambiguity in behaviors being rated. In parent agreement research, ambiguity relating to the descriptions of symptoms for many childhood disorders can lead to problems in ratings or statements about the child.

Kazdin (1988) suggested that parents achieve greater agreement on observable behaviors because, presumably, less speculation is necessary, and the definitions of overt behaviors are clearer and allow for less interpretation than non-overt behaviors. On the other hand, Achenbach et al. (1987), reported slightly greater agreement between raters for internalizing (overcontrolled), rather than externalizing (undercontrolled) behaviors.

Consider the "clinical" items of the CBCL. Many contain ambiguous referents such as item # 5, "Behaves like opposite

sex". A parent's response to this item is likely to be affected by a variety of individual-specific factors including cultural background, gender-role identification, age, etc. In effect, a parent is forced to conjecture about how the opposite sex behaves, since no guidelines are given. However, for an item like # 72, "sets fires", the behavioral description contains more definable and observable terms that may foster agreement between raters.

Even the CBCL response alternatives (0=not true, as far as you know, 1=somewhat or sometimes true, and 2=very true or often true) seem to invite uncertainty. How frequent is sometimes or often? Perhaps specific frequencies of symptom behavior might be incorporated into test items so that a consistency between related diagnostic tools such as DSM-IV (American Psychiatric Association, 1994) is developed.

Christensen et al. (1992) looked at interparental agreement on childhood behavior problems (using the CBCL) utilizing five variables; sex of parent, sex of child, distress level of the marriage, distress level of the child, and type of behavioral problem (according to it's overtness). To measure parental agreement, Christensen et al. calculated Cohen's (1960) Kappa, which indicates the percentage of agreement beyond chance between two informants.

"Discrepancy Indexes" were also computed by the authors, consisting of (a) the number of disagreements in which one parent endorsed a problem behavior and the other parent did

not, and (b) the number of disagreements in which one parent endorsed a problem behavior as occurring more frequently than the other parent. Christensen et al. found that mothers consistently reported more problem behavior than fathers, but these differences were not affected by gender of the child.

Results also indicated that CBCL items on the externalizing scale of the checklist achieved higher interparental agreement. Kappas (Cohen, 1960) for Externalizing items were higher than kappas for Internalizing items ($M=.32$, and $.24$ respectively).

Marsh et al. (1985) looked at the effects of a range of rater characteristics (role, gender, age, and parental status—either a parent or not a parent) on ratings of childhood behavior problems. However, instead of rating children, raters were instructed to rate CBCL items on a scale of one to five in terms of psychological significance (a rating of one was used for normal behaviors that usually don't require mental health services, and a rating of five reflected more serious psychological problems).

Marsh et al. examined four groups of raters: teachers, parents, school psychologists, and clinical psychologists. A multiple discriminant function analysis was used to determine whether ratings could be used to differentiate between the four rater groups. Unfortunately, sex of rater was collapsed across all four groups, and no meaningful inferences could be obtained for sex of the parent group.

Results indicated that items of an internalizing nature such as confusion and fearfulness, were significantly affected by various rater characteristics, particularly role and age. Only fifteen items were unaffected by any rater characteristic (pointing to substantial consistency of perception for those items), which tended to focus on externalizing problems such as disobedience and stealing. There was also a trend for clinical and school psychologists to imbue many childhood behaviors with greater psychological significance than parents and teachers.

Summary

The way parents perceive their childrens' behavior appears to be a function of several key factors including parent gender, interaction of parent and child gender, parent and child mental status, and a variety of other demographic variables. Although a number of studies have reported interparent agreement information in analysis sections, the data is meager, and there is a need for more information about degree of parental agreement.

For the first two categories of research findings in this literature review (parental and child characteristics that influence ratings of child behavior), results are highly variable, and often show only low to modest correlations between mother and father ratings of child behavior problems. In the third category, parent gender by child gender

interaction effects, results were again highly variable, with no clear inferences emerging. The amount of research conducted in this area has been minimal (especially regarding child gender as a variable), and the issue of interparental agreement was often neglected in the research designs.

The last category of research findings addressed statistical considerations surrounding interrater agreement analysis. Most studies utilized Pearson correlations to demonstrate degree of parental agreement, although a few studies compared multiple indices of interrater agreement, and their relative merits or shortcomings for investigations of this nature.

One trend did materialize, however, being a tendency for fathers to underreport behavior problems (Christensen et al., 1992; Eisenstadt et al., 1994; Hulbert et al., 1986; Jensen et al., 1988; LeBlanc & Reynolds, 1989; Lindholm & Touliatos, 1981; Webster-Stratton, 1988); Wierson et al., 1990; Wirt et al., 1984). If future research confirms this trend, then using one combined parental norm to measure the significance of a rating may mask important information in the assessment of a child.

CHAPTER 3

METHOD

Statement of the Problem

Parental ratings are a primary source of information in many mental health and special education referrals for children. A parent's rating is considered important because he/she has the primary role in caretaking, and spends more time with the child than anyone else.

These ratings help clinicians to identify relevant dimensions of children's behavior, to discriminate between different illnesses (and their etiologies), and to choose appropriate interventions. In view of the importance accorded parent ratings of child behavior, the reliability of those ratings is an important concern for clinicians who use behavioral checklists to aid in diagnosis or case management.

Research Questions

This study addresses several issues regarding the reliability of parent Child Behavior Checklist ratings by pursuing the following questions:

1. Are there significant systematic differences in mothers' and fathers' responses to CBCL scales?

2. What is the inter-parent reliability of CBCL scale scores?

3. To what extent do parents' scale scores translate into equivalent clinical classification levels?

3. Does objectivity of item content influence the reliability of parental responses?

4. Does child gender influence parental rating discrepancies on CBCL scales?

Subjects

This research involved the analysis of Child Behavior Checklist protocols collected at the Child Evaluation and Treatment Program (CETP) at the University of North Dakota Rehabilitation Hospital. Data on 82 clinic-referred, 4- to 17-year-old children, for whom both parents or primary caregivers had completed the CBCL, were used in the study.

Of the children, 54 (66%) were boys and 28 (34%) were girls. The children ranged in age from 4 to 17 years ($M=8.71$, $SD=3.26$). Table 1 presents age distributions of children in the sample by gender. In the sample, 71 children (87%) were of Caucasian descent; 4 children (5%) were Native American; 6 children (7%) were Hispanic; and one child was Asian-American (1%). The families had come to CETP seeking assistance in reducing child behavior problems such as noncompliance, hyperactivity, inattention, depression, etc.

Table 1

Age Distribution of Sample by Gender

Age	Boys n	Girls n	Total n	(%)
4	6	1	7	(8.5)
5	4	5	9	(11.0)
6	4	3	7	(8.5)
7	8	3	11	(13.4)
8	7	2	9	(11.0)
9	6	2	8	(9.8)
10	3	3	6	(7.3)
11	5	3	8	(9.8)
12	3	4	7	(8.5)
13	2	-	2	(2.4)
14	3	1	4	(4.9)
15	1	-	1	(1.2)
16	1	1	2	(2.4)
17	1	-	1	(1.2)
Total	54 (65.9%)	28 (34.1%)	82	

Design

This is an *ex post facto* correlational study utilizing a one-between subjects factor, one-within-subjects factor, repeated-measures design. The within-subjects factor, rating source consisted of two levels (mothers and fathers). The

between-subjects factor, child gender, had two levels (male and female). Repeated measures (or within-subjects) designs typically have greater sensitivity due to the lack of error associated with subject variance that is found in randomized or between-subjects designs (Keppel, 1991).

Instruments

The Parent Form of the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1991) was used as a measure of child behaviors. This checklist (Appendix A), completed independently by mothers and fathers, is a commonly used clinical measure of child adjustment. The CBCL consists of two major sections: a social competence section, and a behavior problems section. The focus of this study was on the latter.

The 113 behavioral problem, or "clinical" items utilize a three-category response format: "Not true"=0; "Somewhat or sometimes true"=1; and "Very true or often true"=2. These items are aggregated into eight "narrowband" scales yielding scores on various dimensions of child behavior including attention problems, thought problems, delinquent behavior, and somatic complaints. Achenbach & Edelbrock (1991) indicated that these scales have demonstrated satisfactory test-retest reliability (mean of .89 for 1 week) and parent agreement (mean Pearson r s for four age by sex groups ranged from .65

for girls 4-11 to .75 for boys 4-11), and effectively discriminate clinic-referred children from normal children.

Two "broadband" CBCL scales, Internalizing and Externalizing, are computed by summing scales 1-3 and 7-8, respectively. These groupings were derived from principal factor analyses of the correlations among scale scores separately for each sex by age group of children. The CBCL also provides cutting scores for each of the narrowband scales that demarcate "clinical" and normal range problem levels.

Item objectivity.

In order to assess the objectivity of CBCL clinical items, a questionnaire was devised by the author (Appendix B). Objectivity refers to the overtness or observability of a behavior. Ratings of each item, based upon a three-point scale ("Low objectivity"=1; "Moderately objective"=2; and "Highly objective"=3), were provided by four doctoral students (two male, two female) in counseling psychology at the University of North Dakota. All four judges were familiar with the particular constructs under study. For each scale, a mean item objectivity score was obtained by averaging the objectivity ratings for items included in the scale. The reliability of the scale mean item objectivity ratings (for the composite of four raters) was estimated by intraclass correlation methods (Ebel, 1951) to be .91.

Procedure

Secondary use was made of CBCL protocols from CETP client folders. As part of standard clinic procedure, parents (i.e., biological or adoptive parents, step-parents, or foster-parents) each had independently completed a CBCL on the day they arrived at the clinic, or at home after the clinical interviews and testing, but prior to receiving evaluation results. The following demographic variables were collected for each child: age, race, and gender.

Data Analysis

CBCL item responses for mothers and fathers were computer-entered, and scores on each CBCL scale were computed for each parent. Based upon the CBCL manual instructions, missing item responses on parent forms were scored as 0. The obtained scale scores were used in conjunction with CBCL clinical range cutoff scores to determine whether each parent's rating of the child was of clinical significance.

Signed difference scores were computed for each child on each scale by subtracting the father's scale score from the mother's score. Absolute difference scores were computed by taking the absolute value of the signed difference scores.

Means and standard deviations were obtained for all scale scores. Histograms of signed difference scores and scatterplots of mother vs. father scores were obtained for each scale.

Level differences.

In order to test for mean differences between mothers and fathers in the severity of their ratings, a one-way repeated analysis of variance (ANOVA) was run for each of the CBCL narrowband and broadband scales, with mother's and father's ratings being the repeated measure for each child. In order to maintain a family-wise Type I error rate below .10, each ANOVA *F*-test was run using an alpha-level criterion of .01.

Interrater reliability.

Most studies comparing mothers and fathers as informants have relied on the use of Pearson correlations as the index of reliability (Hulbert et al., 1986). Calculation of a Pearson *r* from the scores of two parents provides an index of variation in one set of scores that can be accounted for/predicted from variations in the second set of scores. However, product-moment correlations may be misleading for evaluating clinical decision-making or for investigations of relationships within parental dyads because they ignore consistent or systematic mean or variance differences between raters. (Jacob et al., 1982).

An alternative approach is the use of intraclass correlation estimates of the reliability of ratings (Shrout and Fleiss, 1979). Intraclass correlation estimates are superior to product-moment estimates because intraclass correlations include rater level differences in the estimation

of error variance. Using the results of the repeated measures ANOVA, the intraclass correlation estimate of reliability, r_{ic} , was computed to be $(BMS-WMS)/(BMS+WMS)$, where BMS is between-subjects mean square, and WMS is within-subjects mean square. The coefficient r_{ic} estimates the reliability of a single rater's score.

Diagnostic classification agreement.

In addition to assessing the interrater reliability of scale scores, it may also be useful to examine the consistency of results when parents' scores are used to determine the child's diagnostic classification. To ascertain this, CBCL scale scores for each parent were examined in relation to the cutoff points which separate normal-range scores from clinically significant elevations. When parents' scores both fall above, or both fall below cutoff points, they may be considered to be in diagnostic agreement; otherwise, they are in diagnostic disagreement.

Crosstabulations of the clinical significance (clinical vs. normal range) of mother and father scores were obtained for each scale. In order to test for parent differences in the proportions of children identified as in the clinical range, McNemar's test (Fleiss, 1973, p. 74) was used. This test is suitable when analyzing data from matched pairs on a dichotomous (normal vs. clinical) outcome.

Two measures of interrater agreement were also calculated, Cohen's (1960) Kappa, κ , and the Occurrence Agreement Index (OAI; Suen & Ary, 1989, pp. 110-111). Kappa measures the proportional improvement beyond chance in the percentage of agreement between two informants. It has a possible range of -1.0 to 1.0, with $\kappa=0$ when observed agreement equals that expected by chance. Greater than chance agreement leads to positive values of κ , while less than chance agreement leads to negative values. A κ of 1.0 is generated when perfect agreement is observed.

Parent diagnostic classification consistency may also be evaluated using an occurrence agreement index, which takes the following form:

$$\frac{\text{number of occurrence agreements}}{\text{number of occurrence agreements} + \text{number of disagreements}}$$

According to Suen & Ary (1989), this index is especially appropriate when the base rate for a clinical diagnosis is low. In the present study, the children referred for psychological evaluation bring a diverse set of problems, and therefore, the proportion of clinical-range cases for a given CBCL scale may be small.

Child gender and parent discrepancies.

In order to determine if child gender influences mother-father rating discrepancies, a two-groups (boys vs. girls) *t*-test was conducted for each scale, with the signed difference

score as the dependent variable. This procedure is statistically equivalent to testing for a significant, child gender by parent interaction through analysis of variance. As was done in the earlier ANOVA's, overall Type I error rate was controlled to be less than .10 by running each significance test at a nominal .01 level.

CHAPTER 4

RESULTS

Differences Between Mothers and Fathers

Means and standard deviations of parents' CBCL ratings are presented in Table 2. Although the mothers' mean ratings for children's behavior problems were slightly higher than the fathers' on every scale except Scale 2 (Somatic Complaints), these differences were small in comparison to the variation observed across children. In terms of effect size, the most substantial observed difference (Scale 8-Aggressive Behavior) represented less than .15 standard deviation units.

Means and standard deviations for absolute and signed difference scores on the ten scales, and results of the repeated-measures ANOVAs are presented in Table 3. The mean signed differences were generally small (the largest was 1.45), and none of the associated ANOVA *F*-tests were statistically significant. In other words, no systematic tendency for mothers or fathers to give more severe ratings was found.

At the same time, however, there was substantial evidence of rating discrepancies for individual mother-father dyads. The average absolute parent difference scores on the scales ranged from .91 (Somatic Complaints) to 7.11 (Externalizing

Table 2

Mother and Father CBCL Scores: All Children.

CBCL Scale	Raters			
	Mothers		Fathers	
	M	SD	M	SD
1 Withdrawn	3.65	2.66	3.50	2.78
2 Somatic Complaints	1.17	.98	1.23	1.27
3 Anxious/Depressed	6.62	5.16	6.07	4.76
4 Social Problems	4.58	2.88	4.16	3.01
5 Thought Problems	2.22	2.17	2.06	1.94
6 Attention Problems	8.38	4.23	8.07	4.83
7 Delinquent Behavior	4.29	3.81	4.01	3.76
8 Aggressive behavior	16.68	9.42	15.51	9.00
Internalizing Broadband	10.95	6.80	10.34	6.99
Externalizing Broadband	20.98	11.97	19.52	11.56

broadband), with mean absolute differences of greater than 3 points occurring on four of the scales.

In Figures 1 through 20, histograms of mother vs. father signed score differences and scatterplots of mother vs. father scores for all scales and broadbands are presented. Points on the scatterplots are depicted using a "sunflower" schema, where each petal represents one case, allowing for precise indication of multiple occurrences. Deviations from the diagonal line on the scatterplot reflect the extent to which parent-pairs diverged from consensus.

Table 3

Differences Between Ratings of Mothers and Fathers.

CBCL Scale	Absolute Difference		Signed Difference		F	p
	Mean	S.D.	Mean	S.D.		
1-Withdrawn	2.05	1.51	.15	2.55	.27	>.60
2-Somatic Comp	.91	.83	-.06	1.24	.20	>.60
3-Anxious/Depr	2.55	2.57	.55	3.59	1.92	.17
4-Social Probl	2.18	1.70	.43	2.74	1.98	.16
5-Thought Prob	1.52	1.43	.16	2.09	.47	.49
6-Attention Pr	3.01	2.66	.31	4.02	.47	.49
7-Del Behavior	1.82	1.89	.28	2.61	.94	.33
8-Agg Behavior	5.83	5.28	1.17	7.80	1.85	.18
Internalizing	4.46	3.53	.61	5.68	.94	.33
Externalizing	7.11	6.53	1.45	9.58	1.88	.17

* Difference was calculated by subtracting father's scale score from mother's scale score.

When looking at these figures, the extent of parent disagreement is markedly evident. For example, on Figure 1 (histogram of Scale 1-Withdrawn), it is apparent that individual parent-pairs are dispersed throughout a wide range

of score differences (-7.5 to +5.5), yet the average difference score is less than one point! Analogously, the corresponding scatterplot (Figure 2) reveals a paucity of cases falling on the perfect agreement diagonal line, further documenting the trend for parental discrepancy on Scale 1 ratings.

Histograms and scatterplots also help clarify the direction of mother vs. father differences. With histograms, the distribution above zero difference indicates the number of mothers reporting higher psychopathology ratings for children. With scatterplots, a predominance of points above or below the perfect agreement line indicate which parent-gender has higher ratings. This is reflected in Figure 3 (histogram for Scale 2- Somatic Complaints), where difference scores approximate a symmetrical distribution around zero. This indicates that no trend emerged for mothers or fathers to be more severe in their ratings on this scale. The scatterplot (Figure 4) corroborates these results with a relative balance in occurrences above and below the perfect agreement line.

Graphs for Scale 3-Anxious/Depressed (Figures 5 and 6), Scale 4- Social Problems (Figures 7 and 8), Scale 5-Thought Problems (Figures 9 and 10), Scale 6-Attention Problems (Figures 11 and 12), Scale 7-Delinquent Behavior, and the Internalizing broadband (Figures 17 and 18) exhibited qualities similar to those of Scale 1. The histograms are characterized by a large range of score differences that are

evenly distributed around zero mean difference. Scatterplots reflect the large standard deviations in observed scores, with relatively few cases falling on, or close to the perfect agreement line. Thus, on these CBCL scales, parents are generally portrayed in discrepancy rather than agreement.

Graphs for Scale 8-Aggressive Behavior (Figures 15 and 16) and the Externalizing broadband (Figures 19 and 20) reveal a slight tendency (not statistically significant) for mothers to rate their children more severely, and this difference is apparent in the scatterplots.

Reliability of Parent Ratings

Intraclass correlation estimates of the reliability of parent ratings on the CBCL scales are reported in Table 4. The estimates ranged from .42 (Somatic Complaints) to .77 (Delinquent Behavior). These coefficients represent the reliability of the ratings from a single parent. (The corresponding reliabilities of averaged ratings from both parents would be .59 and .87.) In view of the weight given to CBCL profiles in clinical decision-making, the size of the observed parent CBCL rating discrepancies and the estimated reliabilities of the rating scales may be of concern.

It should be noted further that reliabilities will generally be even less when the range of scores is restricted. That is, when decisions are being made with respect to a subset of individuals who have elevated scores on a scale, the

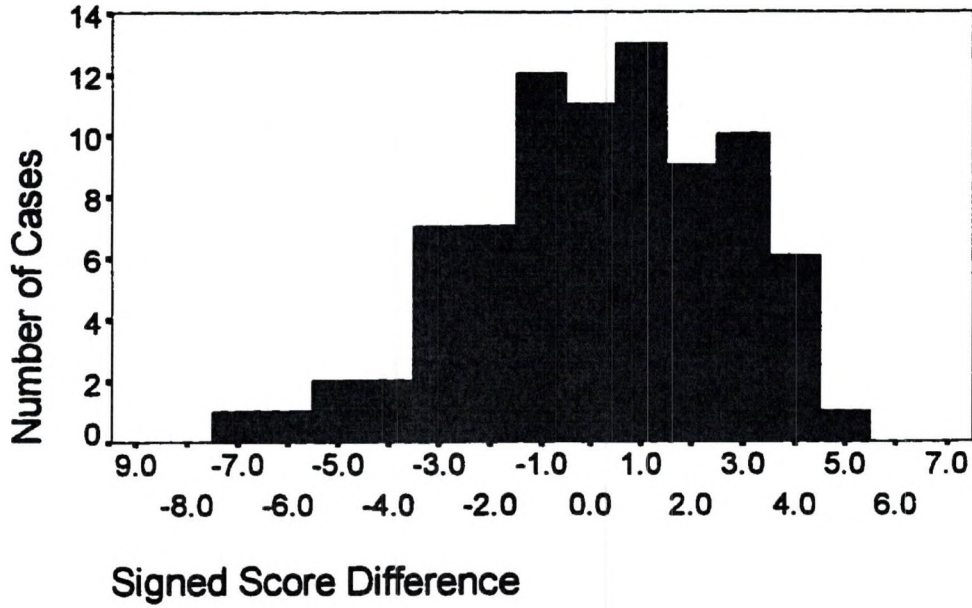


Figure 1. Histogram of signed score differences (mother score minus father score) for Scale 1-Withdrawn.

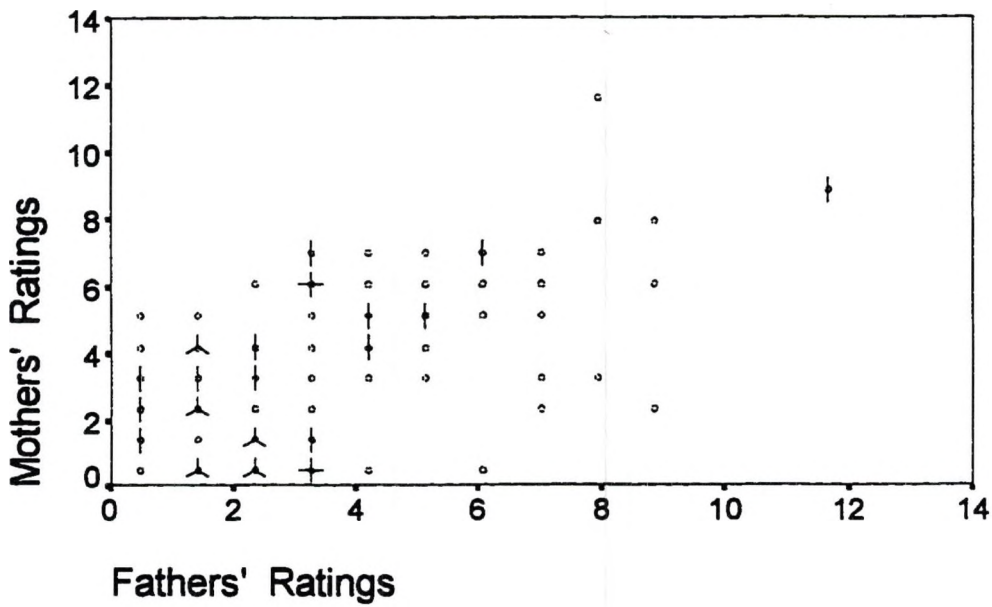


Figure 2. Scatterplot of mother vs. father scores for Scale 1-Withdrawn.

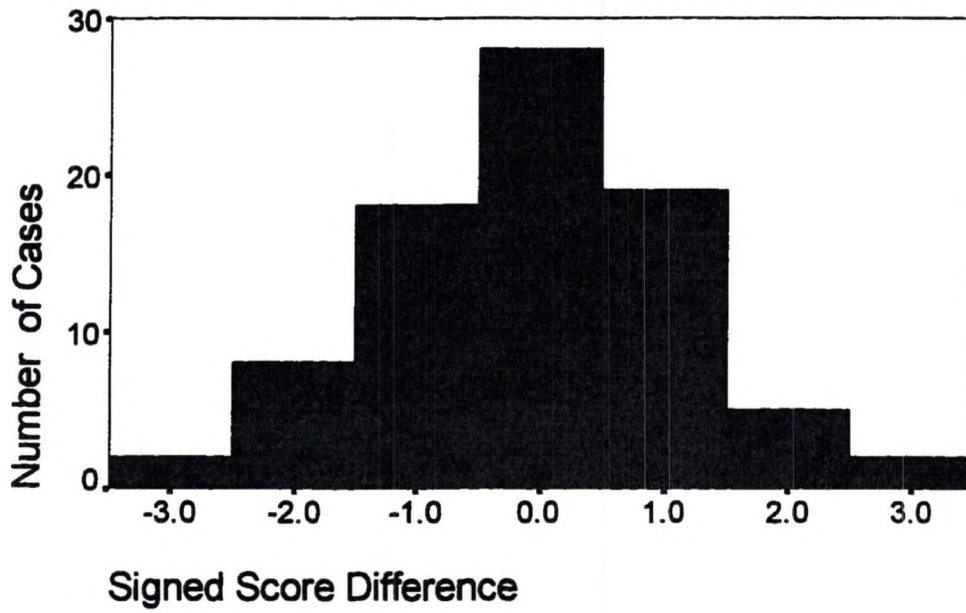


Figure 3. Histogram of signed score differences (mother score minus father score) for Scale 2-Somatic Complaints.

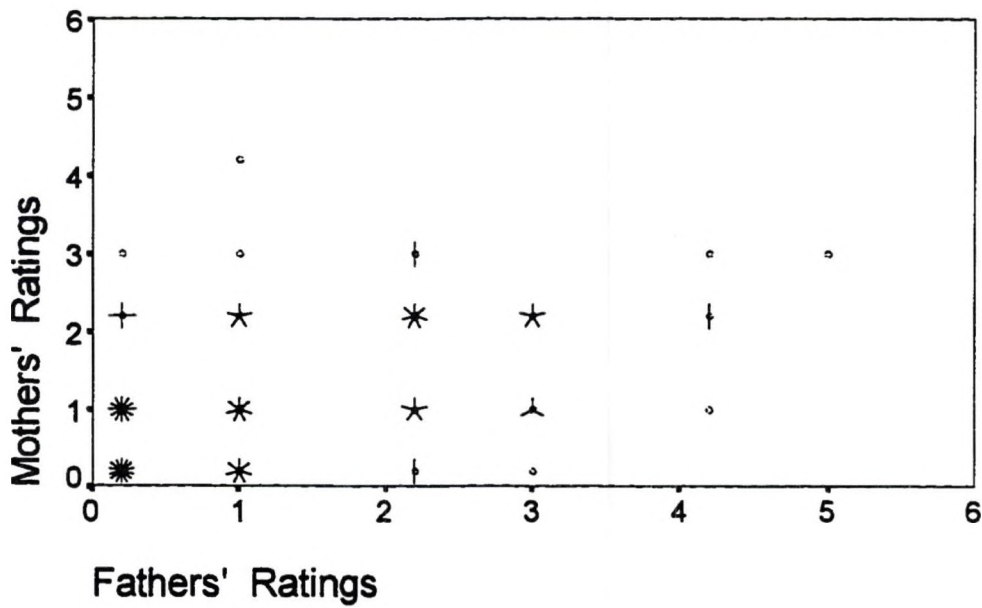


Figure 4. Scatterplot of mother vs. father scores for Scale 2-Somatic Complaints.

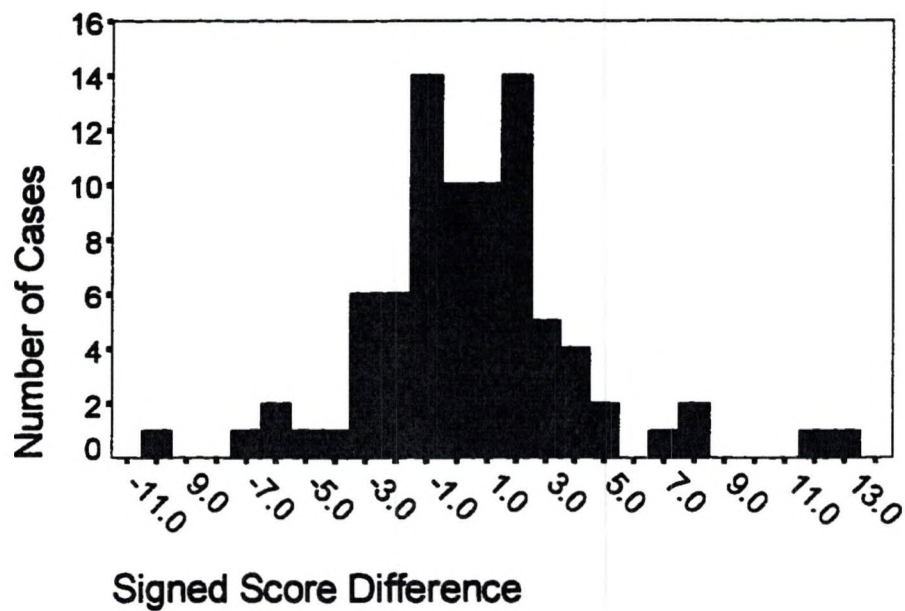


Figure 5. Histogram of signed score differences (mother score minus father score) for Scale 3-Anxious/Depressed.

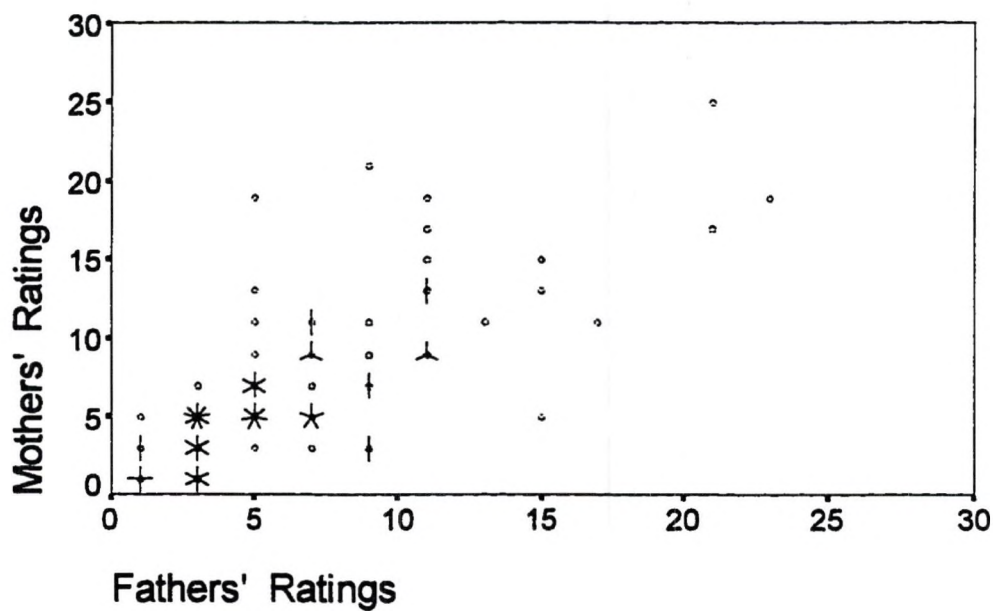


Figure 6. Scatterplot of mother vs. father scores for Scale 3-Anxious/Depressed.

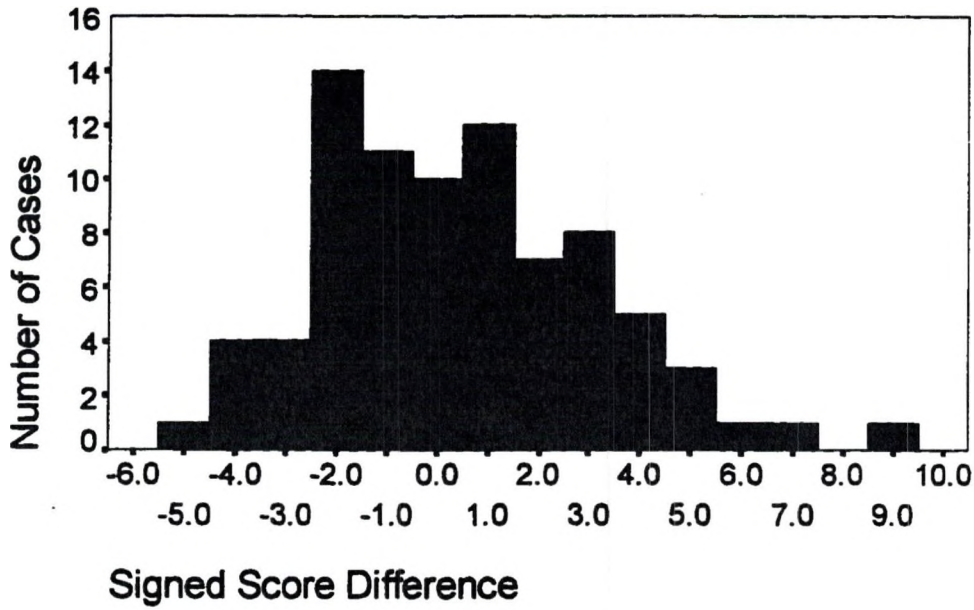


Figure 7. Histogram of signed score differences (mother score minus father score) for Scale 4-Social Problems.

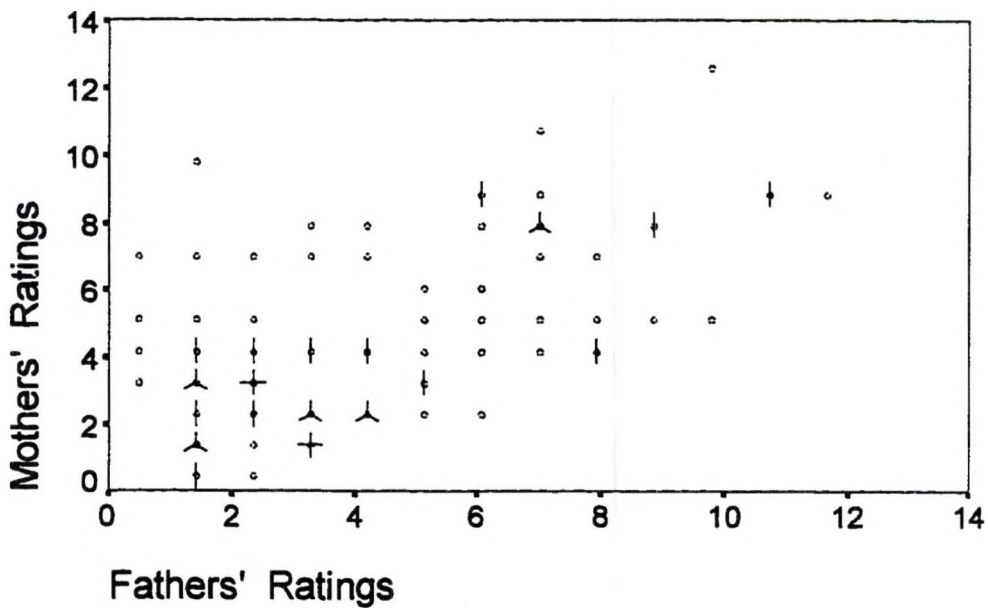


Figure 8. Scatterplot of mother vs. father scores for Scale 4-Social Problems.

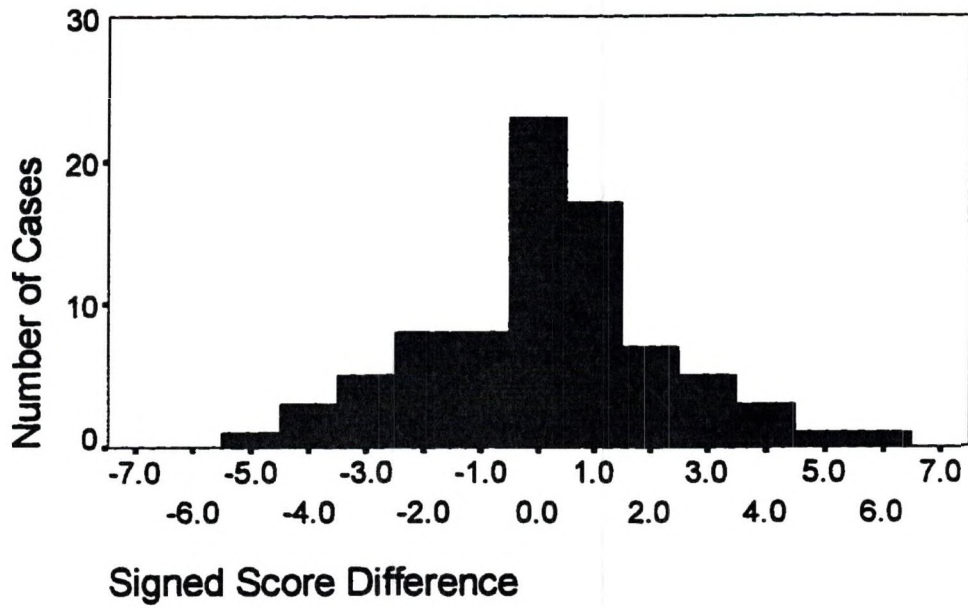


Figure 9. Histogram of signed score differences (mother score minus father score) for Scale 5-Thought Problems.

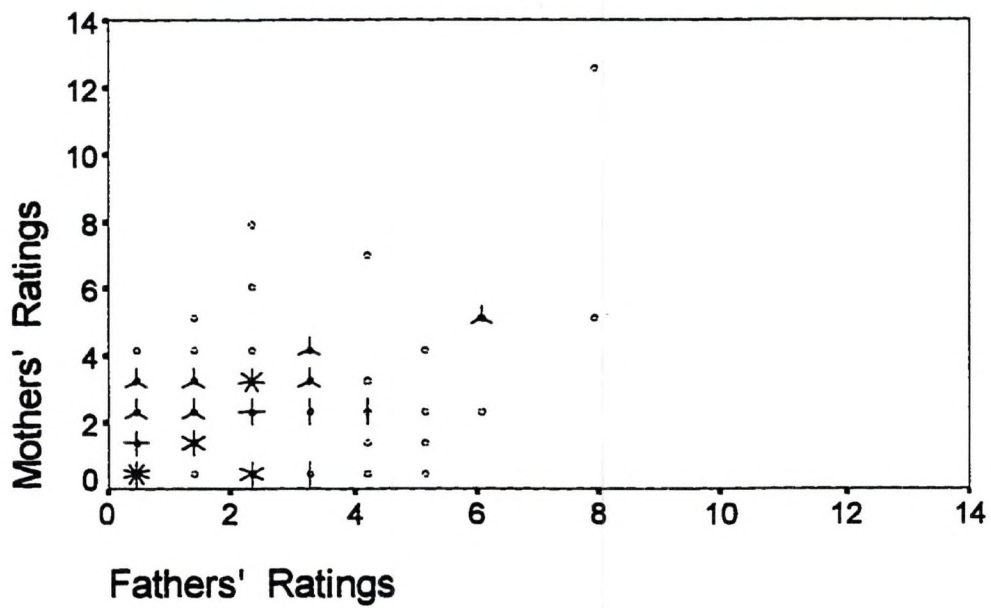


Figure 10. Scatterplot of mother vs. father scores for Scale 5-Thought Problems.

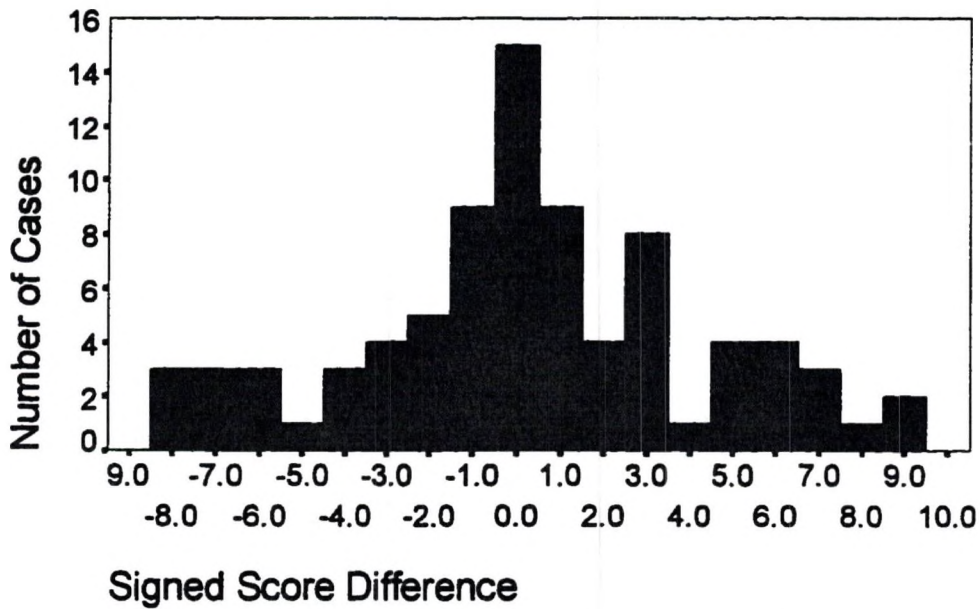


Figure 11. Histogram of signed score differences (mother score minus father score) for Scale 6-Attention Problems.

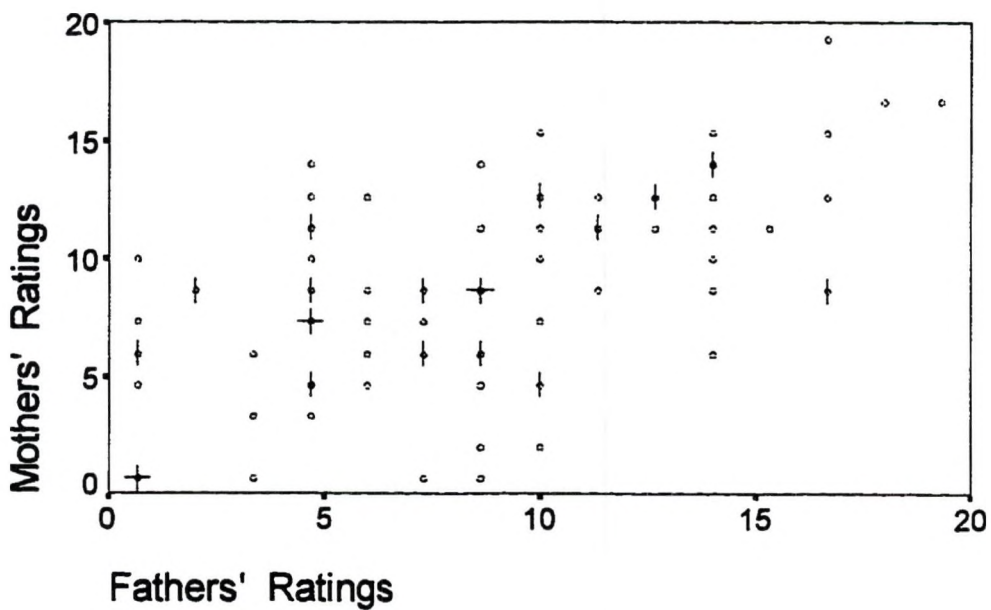


Figure 12. Scatterplot of mother vs. father scores for Scale 6-Attention Problems.

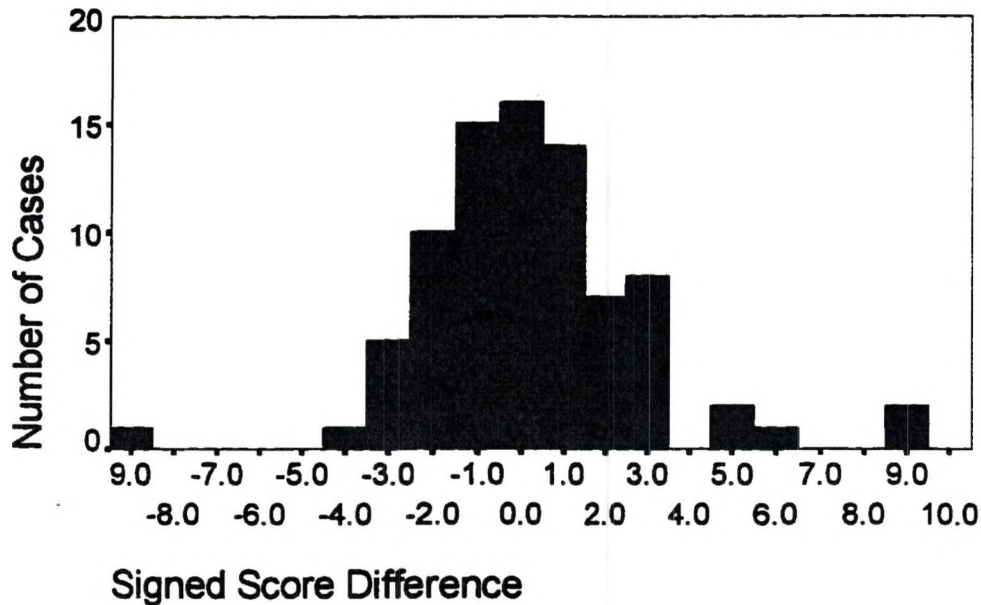


Figure 13. Histogram of signed score differences (mother score minus father score) for Scale 7-Delinquent Behavior.

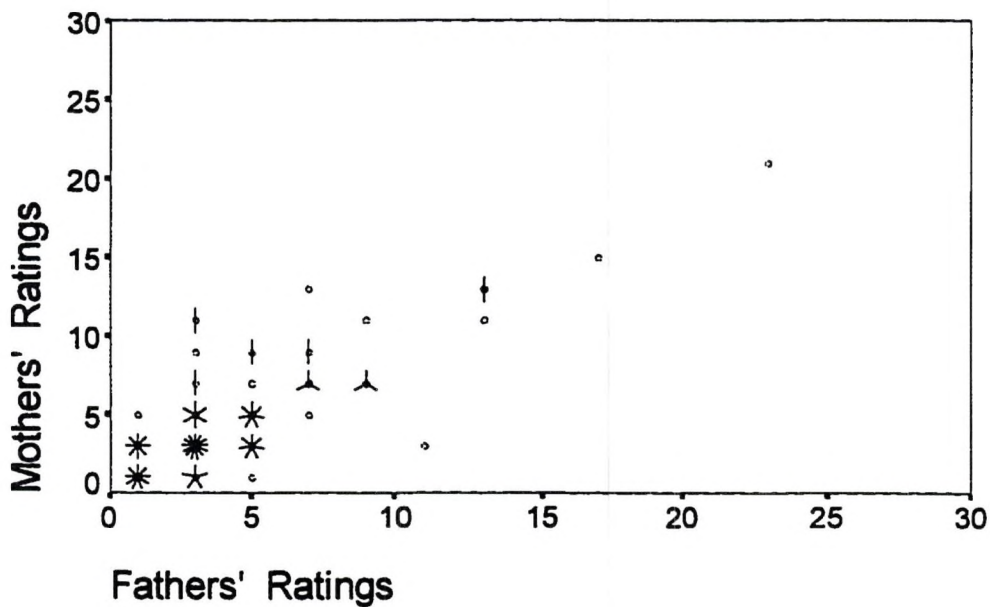


Figure 14. Scatterplot of mother vs. father scores for Scale 7-Delinquent Behavior.

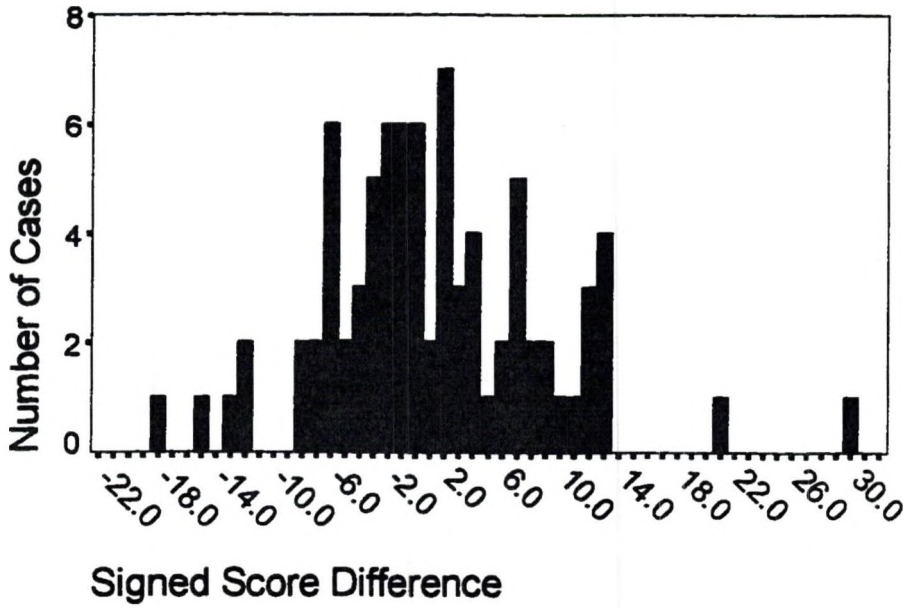


Figure 15. Histogram of signed score differences (mother score minus father score) for Scale 8-Aggressive Behavior.

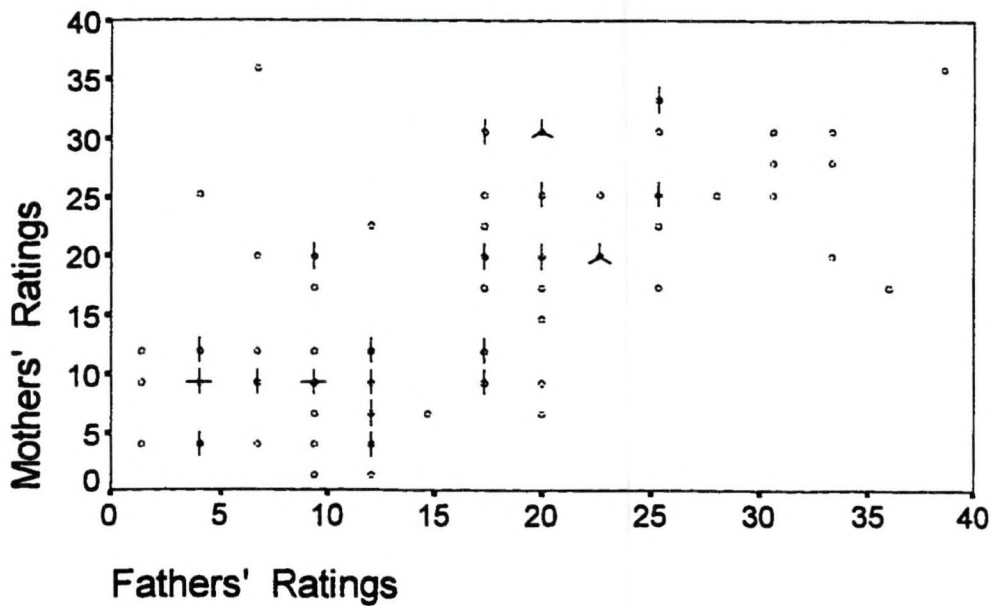


Figure 16. Scatterplot of mother vs. father scores for Scale 8-Aggressive Behavior.

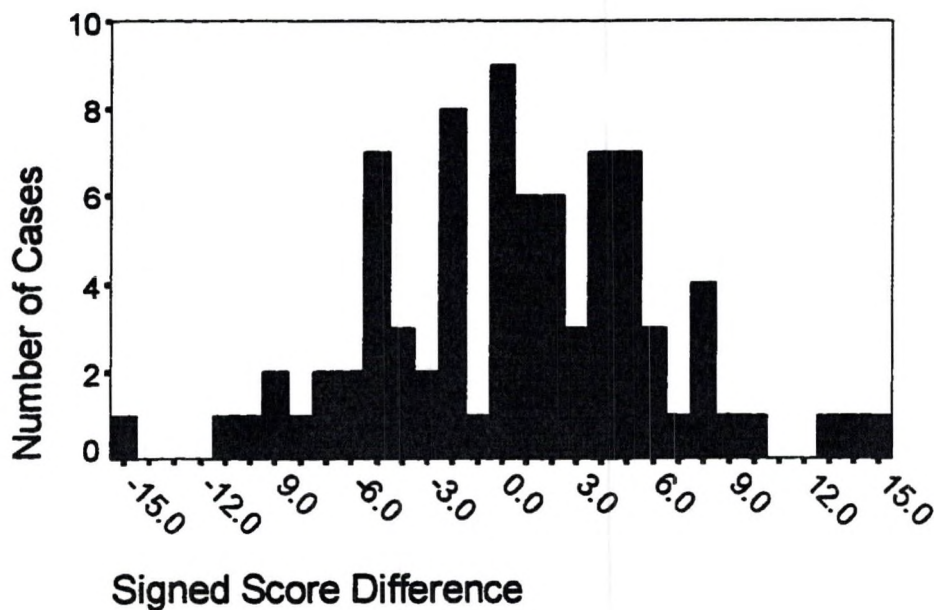


Figure 17. Histogram of signed score differences (mother score minus father score) for the Internalizing broadband.

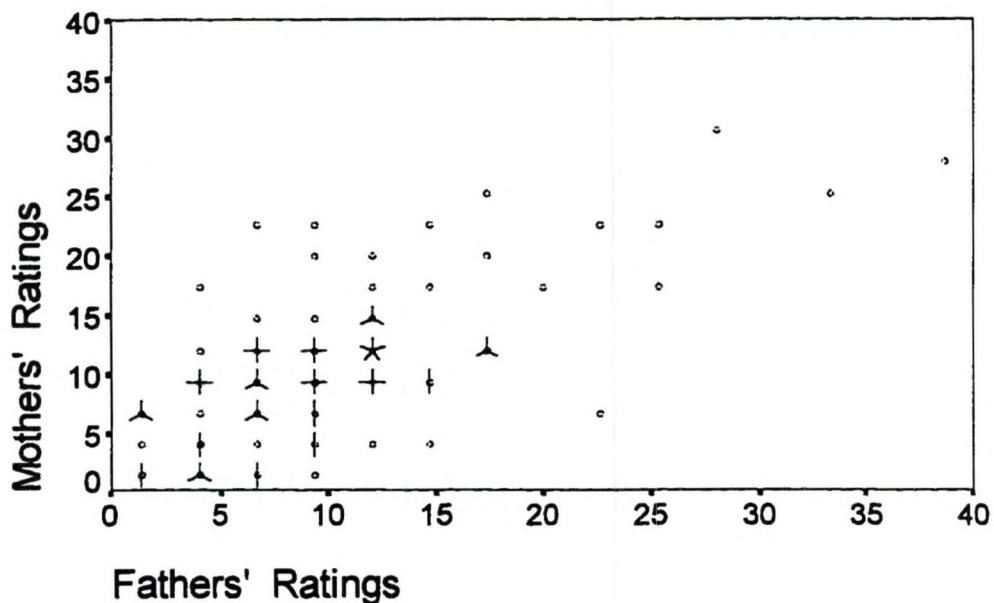


Figure 18. Scatterplot of mother vs. father scores for the Internalizing broadband.

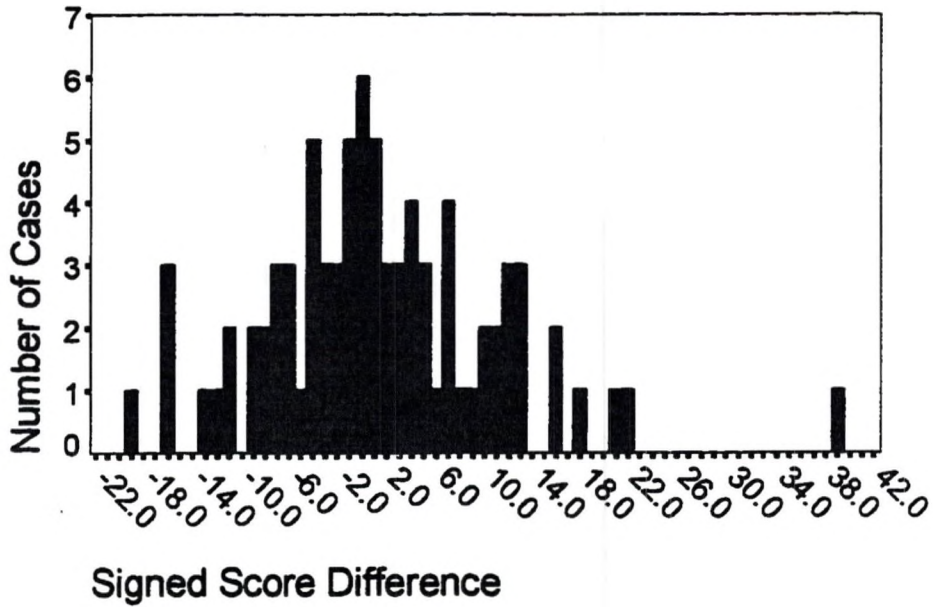


Figure 19. Histogram of signed score differences (mother score minus father score) for the Externalizing broadband.

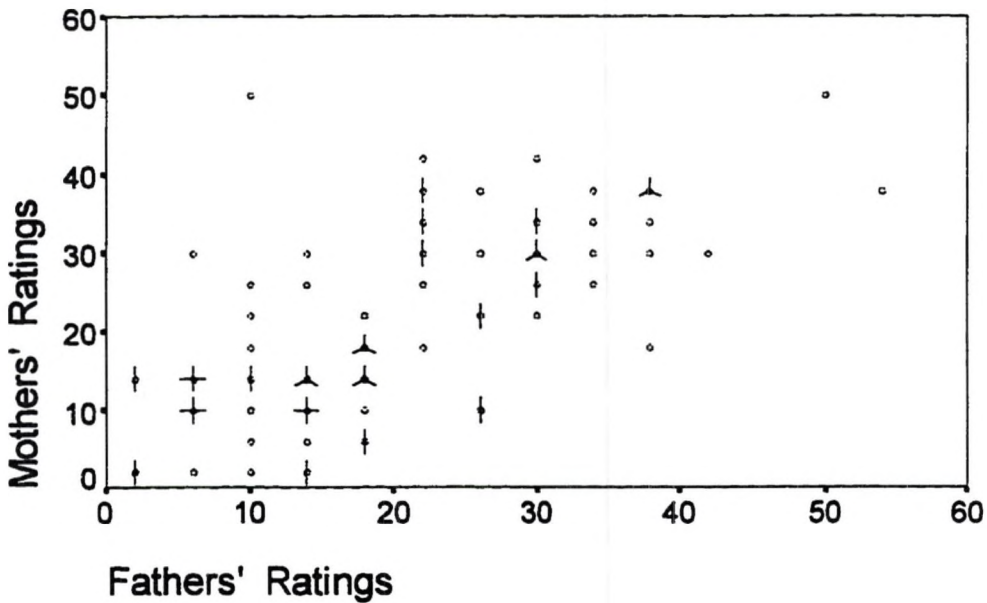


Figure 20. Scatterplot of mother vs. father scores for the Externalizing broadband.

Table 4

Intraclass Correlation Reliability Estimates for Parent CBCL Scale Scores.

CBCL Scale	Reliability ^a
1--Withdrawn	.56
2--Somatic Complaints	.42
3--Anxious/Depressed	.52
4--Social Problems	.57
5--Thought Problems	.48
6--Attention Problems	.62
7--Delinquent Behavior	.77
8--Aggressive Behavior	.68
Internalizing Band	.56
Externalizing Band	.69

^aReliability is estimated reliability of rating by one parent.

reliability of those scores will be less than that for the total group. Elevated score groups were defined for each CBCL scale by selecting subjects whose summed parent scores exceeded the sum of parent means on the scale. Intraclass correlation estimates of single rater reliabilities for elevated score groups on the ten CBCL scales and broadbands were computed, and ranged from -.04 (Scale 2) to .63 (Scale 7), with a median of .30.

Diagnostic Classification Agreement

Using cutoff points defined in the CBCL manual (Achenbach, 1991), mother and father scores on each narrowband scale were classified as either "normal range" or "clinical." Table 5 presents results on the agreement of mother and father score classifications for each of the scales.

For 7 of 8 scales, mothers' ratings fell in the clinical range more often than those of fathers, but none of these differences were detected as statistically significant by McNemar's test. On Scale 2 (Somatic Complaints) no children were rated in the clinical range by either parent.

Cohen's (1960) Kappa, κ , was calculated for all CBCL scales to augment our understanding of Table 5. Kappa reflects the proportional level of agreement beyond that expected by chance. Across the eight scales, κ was low to moderate in value, ranging from .30 on Scale 5 (Thought Problems) to .55 on Scale 3 (Anxious/Depressed). All obtained kappa values were significantly different from 0.0 ($p < .05$). The median κ value was .46. Thus, the typical level of agreement observed reflected less than 50% of the improvement possible beyond chance agreement. Gelfand and Hartmann (1975) recommended a kappa of .60 as the minimally acceptable level of interobserver agreement. Landis and Koch (1977) suggested that a kappa of .80 is an indication of "good" reliability.

The Occurrence Agreement Index (OAI) provides additional information on classification agreement in instances where at

Table 5

Diagnostic^a Classification of Mother (M) and Father (F) Ratings.

CBCL Scale	M Normal		M Clinical		M Clinical		M Normal		κ	(ASE) ^b	OAI ^c
	F Normal n	F Normal (%)	F Clinical n	F Clinical (%)	F Normal n	F Normal (%)	F Clinical n	F Clinical (%)			
1	68	(83)	4	(5)	6	(7)	4	(5)	.38	(.16)	.28
2	82	(100)	0	-	0	-	0	-	-	-	-
3	64	(78)	8	(10)	8	(10)	2	(2)	.55	(.13)	.44
4	50	(61)	14	(17)	11	(13)	7	(9)	.46	(.11)	.44
5	57	(70)	7	(8)	9	(11)	9	(11)	.30	(.13)	.28
6	44	(54)	20	(24)	11	(13)	7	(9)	.52	(.10)	.53
7	59	(72)	10	(12)	9	(11)	4	(5)	.51	(.12)	.43
8	46	(56)	16	(19)	13	(16)	7	(9)	.44	(.11)	.44

^a Classification levels (clinical versus normal) based upon cutoff points as defined in the CBCL Manual (Achenbach, 1991).

^b Asymptotic Standard Error of Kappa.

^c Occurrence Agreement Index (Suen & Ary, 1987).

least one parent's rating is in the clinical range. The lowest OAI values were observed on Scale 1 (Withdrawn) and Scale 5 (Thought Problems), where only 28% of all judgements in the clinically significant category involved agreement between parents. Mothers and fathers exhibited greater consensus (OAI=.53) on Scale 6 (Attention Problems). The median OAI value for the eight scales was .44. For seven of the eight scales, among children having at least one parent rating in the clinical range, less than half had both parents in agreement.

Item Objectivity and Reliability of Ratings

The means and standard deviations of mean item objectivity ratings were calculated for all scales, and appear in Table 6. Kappas and reliabilities for the scales are presented again in this table to assist in the examination of the relationship between item objectivity and parental reliability or parental agreement across the CBCL scales.

There is some evidence for objectivity of item content increasing parental agreement. As expected, each of the Internalizing scales (Withdrawn, Somatic Complaints, and Anxious/Depressed) were rated as less objective than Externalizing scales (Delinquent Behavior and Aggressive Behavior). Externalizing Scales, composed of items rated as more objective or observable, demonstrated greater parental agreement (higher kappas) and higher reliabilities than Internalizing Scales.

Table 6

CBCL Scale Item Objectivity Ratings, Kappa Coefficients, and Reliability Estimates.

CBCL Scale	Item Objectivity		Kappa	Reliability
	Mean	S.D.		
1-Withdrawn	2.08	.21	.41	.56
2-Somatic Complaints	1.67	.47	--	.42
3-Anxious/Depressed	1.73	.30	.42	.52
4-Social Problems	2.19	.30	.47	.57
5-Thought problems	1.61	.27	.27	.48
6-Attention Problems	1.96	.22	.54	.62
7-Delinquent Behavior	2.54	.11	.52	.77
8-Aggressive Behavior	2.38	.38	.47	.68
Internalizing Band	1.83	.30		.56
Externalizing Band	2.46	.23		.69

Interaction Between Parent and Child Gender

In Table 7, the means and standard deviations for mothers' and fathers' ratings on all CBCL scales are presented by gender of the child. Signed differences (mother vs. father scores) were computed separately for boys and girls on all scales, and the results appear in Table 8. Inspection of these means suggests a slight tendency for mother vs. father rating discrepancies to be greater for boys than for girls, but none of these were detected as statistically significant. Thus, there is not support for the existence of an interaction

between parent and child gender in the level of CBCL ratings given.

Table 7

Mother and Father CBCL Scores by Child Gender.

CBCL Scale	Daughters				Sons			
	Mother		Father		Mother		Father	
	M	SD	M	SD	M	SD	M	SD
1	4.21	3.01	4.32	3.23	3.35	2.44	3.07	2.44
2	1.43	1.10	1.54	1.43	1.04	0.89	1.07	1.16
3	7.07	5.17	6.54	5.52	6.39	5.20	5.83	4.35
4	4.57	3.35	4.43	3.05	4.59	2.65	4.02	3.01
5	2.46	2.59	2.39	2.44	2.09	1.93	1.89	1.62
6	7.54	5.26	7.93	5.81	8.82	3.57	8.15	4.28
7	3.89	3.98	3.75	3.44	4.50	3.74	4.15	3.94
8	13.50	8.73	14.39	8.20	18.33	9.42	16.09	9.40
Int	12.18	7.30	11.86	8.42	10.32	6.51	9.56	6.05
Ext	17.39	11.76	18.14	10.18	22.83	11.76	20.24	12.25

Table 8

Differences Between Mother and Father Ratings by Child Gender.

CBCL Scale	Signed Difference [*]		t-Value	p
	Mean	S.D.		
1-Withdrawn				
Boys	.278	2.587	.64	.52
Girls	-.107	2.514		
2-Somatic Complaints				
Boys	-.037	1.149	.24	.81
Girls	-.107	1.423		
3-Anxious/Depressed				
Boys	.556	3.601	.02	.98
Girls	.536	3.636		
4-Social Problems				
Boys	.574	2.639	.67	.50
Girls	.143	2.965		
5-Thought Problems				
Boys	.204	2.158	.27	.79
girls	.071	1.999		
6-Attention Problems				
Boys	.667	3.475	1.13	.26
Girls	-.393	4.894		
7-Delinquent Behavior				
Boys	.352	2.593	.34	.73
Girls	.143	2.690		
8-Aggressive Behavior				
Boys	2.241	7.219	1.75	.09
Girls	-.893	8.582		
Internalizing Band				
Boys	.756	5.600	.33	.74
Girls	.321	5.926		
Externalizing Band				
Boys	2.593	8.699	1.51	.14
Girls	-.750	10.906		

*Signed difference was calculated by subtracting father's scale score from mother's scale score.

CHAPTER 5

DISCUSSION

The purpose of this study was to examine parent agreement on the CBCL from several perspectives including interrater reliability and diagnostic agreement, and to examine the impact of item objectivity and child gender on rater agreement.

Parent Level Differences

The mean levels of ratings given by mothers and fathers were generally quite similar, and were not found to differ significantly on any of the CBCL scales. All of the observed mean differences were small compared to the variation in scores among the children rated. Thus, no evidence was found for either mothers or fathers being more severe raters.

Since much of the research in this area (Christensen et al., 1992; Eisenstadt et al., 1994; Hulbert et al., 1986; Jensen et al., 1988; LeBlanc & Reynolds, 1989; Lindholm & Touliatos, 1981; Webster-Stratton, 1988; Wiersen et al., 1990; Wirt et al., 1984) has found that mothers report more problematic child behavior problems, the present results were somewhat unexpected.

Interrater Reliability

Even though no mean level differences were observed between the two parent groups, substantial differences often were found in the ratings of parent pairs. These discrepancies are reflected in the moderate levels of interrater reliability (.42 to .77, mean=.59) obtained for the 10 scales. Although these reliabilities might be adequate for use in research settings, Nunnally (1967) indicates that when important clinical decisions are being made about individuals, scale reliabilities in the range observed in this study are not adequate.

Most studies comparing mothers and fathers as informants have reported correlations as the index of agreement. In the review of literature, a broad range of interparent correlations across a variety of instruments was observed. High correlations were often presented as suggesting that either parent can be an adequate informant. Achenbach (1991) reported CBCL scale mean r s ranging from .52 (Somatic Complaints) to .80 (Externalizing) and a mean r =.67, with the comment that "good" interparent agreement was in evidence (those results exceeded the mean r of .59 found in his meta-analysis of parental agreement studies; Achenbach et al., 1987).

On the other hand, low correlations were typically presented as supporting evidence for further inquiry into the identification and nature of factors influencing interparent

agreement, or the revision of norms to include separate mother and father rating standards. The results obtained in the present study would seem to fall in the latter category.

Diagnostic Classification Agreement

The extent of parental discrepancy is also reflected in the diagnostic classifications made based on mothers' and father's scores. The low to moderate kappa coefficients (.27 to .54) obtained indicated that the proportion of agreement between parents significantly exceeded chance expectation, but fell well below the levels recommended by Cohen (1960), Gelfand and Hartmann (1975) and Landis and Koch (1977).

Low to moderate Occurrence Agreement Indices (ranging from .28 to .53, with a median value of .44) indicated that when one parent's rating placed the child in the clinical range, that classification was corroborated by the second parent's rating in less than half of the cases.

The present results are similar to those reported by Hulbert et al. (1986), who employed "interpretive" diagnostic classification as an index for parental agreement using the PIC instead of the CBCL. Mothers rated children in the clinical range more often than fathers (11 of 16 scales on the PIC versus 7 of 8 CBCL scales in the present study).

These figures suggest a lower-than-expected level of interpretive agreement, given the finding of no mean-level differences between parents. In other words, diagnostic

agreement can have a poor correspondence with correlational analyses.

Item Objectivity

In general, the externalizing scales (Aggressive Behavior and Delinquent Behavior) manifest higher rated item objectivity, and higher levels of reliability and agreement. This corroborates the findings of Christensen et al. (1992), and is consistent with the general consensus that agreement between informants is higher for objective child behaviors than for subjective states (Kazdin, 1988).

Interaction Between Parent and Child Gender

Results of this study failed to reveal a relationship between the level of mother-father rating discrepancies and the sex of the child. This suggests that parents of different gender do not adopt different sex-specific norms in evaluating their children's behavior.

Conclusions

The purpose of this study was to explore the issue of parent agreement on the CBCL. Two basic sorts of questions were addressed. The first one concerned the presence of systematic differences between mothers' and fathers' ratings of children. The second concerned the extent of interrater reliability or agreement of two parents rating the same child.

From the perspective of CBCL users, the results of the study provide both reassuring and troubling news. No evidence for systematic differences in the ratings of mothers and fathers was found. On average, mothers and fathers were equally severe in their ratings. This suggests that there is little benefit to be had by constructing separate norms for mothers and fathers; therefore the current practice of using a common set of norms appears appropriate.

Unfortunately, in many instances, there are very large discrepancies in the ratings provided by the two parents of children referred for psychological evaluations. The extent of these disagreements is reflected in the modest interrater reliability coefficients and agreement indices (Kappas and OAI) that were observed.

Given the importance of the clinical assessment process, and the weight given to CBCL ratings in that process, these results are somewhat troubling. For many of the children in the study, the independent reports of each parent, taken by themselves, would lead to very different clinical and diagnostic decisions.

To some extent, this concern could be overcome by obtaining ratings from both parents. The average of two parent's ratings would be expected to be more reliable than either parent's rating alone, and when substantial discrepancies were found they could provide a basis for further exploration. In clinical practice, however, the great

majority of cases involve children for whom only one parent's report is available at the time of assessment. Depending upon which parent accompanies the child, the clinical outcome can be greatly affected.

Limitations

A major limitation of this research study lies in its use of a convenience sample which may limit the extent to which these results generalize to other settings and populations. Estimates of reliability always depend upon the heterogeneity of the sample involved. In the present case, we are advised that inferences from research findings may only be relevant to clinic-referred, behavior problem children in a north-midwestern, predominantly caucasian locale.

Another limitation of reliability analyses is that they generally require the assumption that raters are equally reliable. In this study, mothers and fathers were presumed to make the same magnitude of error in their judgements of offspring. It may be the case that mothers (or fathers) are much more accurate informants, but that cannot be determined from this study.

Although parents were asked to independently complete the questionnaires, no means were available to assess potential interparent collaboration in the home or in the clinic waiting room. In other words, results obtained may even overestimate parents' actual level of independent agreement.

Similarly, it is possible that contextual influences surrounding the child evaluation process also serve to encourage higher parental agreement. For example, before a child is admitted to a treatment center, numerous conferences and interviews with the child's parents often take place. Consequently, aspects of the child's behavior that are of high concern may be brought to the attention of both parents, and a relative consensus could develop regarding the child's behavior. Furthermore, parents presenting their children for psychological evaluation might have a vested interest in demonstrating a "united front" in their depictions of a child's behavior in order to obtain social or mental health services.

An additional limitation is the study's focus upon a single measure of child behavior, the CBCL. Therefore, the results obtained may be unique to this measure. Some of the agreement observed between parents may reflect CBCL method variance rather than actual child behavior.

Another limitation is that girls were underrepresented in the sample. Discrepancies between maternal and paternal perceptions of child misbehavior may differ as a function of child gender. The smaller number of female cases in the study may have reduced the statistical power of tests for those differences. Achenbach (1991) reported greater interparent correspondence on the CBCL for boys than for girls.

Directions for Future Research

The absence of systematic differences between mothers and fathers indicates that parent role/gender is not the primary source of rating error or rating unreliability. Consequently, other factors that contribute to mother vs. father rating discrepancies need to be identified or examined further. Some of these factors include parent psychopathology and parent involvement; other pertinent variables remain to be identified.

It would be desirable for future research to include multiple measures in the experimental design, and it would certainly seem prudent for researchers to obtain data from both parents and other informants in order to increase understanding of child behavior problems. Therefore in future studies it would be appropriate to use additional measures such as the Conners' Parent Rating Scale (CPRS) and/or the Eyberg Child Behavior Inventory (ECBI) to show if there are any across-measure and between-measure effects on interrater agreement.

Research studies of the sort proposed can do more than document the methodological weaknesses of self-report; they can provide substantial information on parenting, human cognition, gender-related effects, and family dynamics.

APPENDICES

Appendix A

CBCL

CHILD BEHAVIOR CHECKLIST FOR AGES 4-18

For office use only
ID #

CHILD'S NAME			PARENTS' USUAL TYPE OF WORK, even if not working now. (Please be specific—for example, auto mechanic, high school teacher, homemaker, laborer, is the operator, shoe salesman, army sergeant.)		
SEX <input type="checkbox"/> Boy <input type="checkbox"/> Girl	AGE	ETHNIC GROUP OR RACE	FATHER'S TYPE OF WORK: _____		
TODAY'S DATE Mo _____ Date _____ Yr _____		CHILD'S BIRTHDATE Mo _____ Date _____ Yr _____	MOTHER'S TYPE OF WORK: _____		
GRADE IN SCHOOL _____	Please fill out this form to reflect your view of the child's behavior even if other people might not agree. Feel free to write additional comments beside each item and in the spaces provided on page 2.		THIS FORM FILLED OUT BY:		
NOT ATTENDING SCHOOL <input type="checkbox"/>			<input type="checkbox"/> Mother (name: _____) <input type="checkbox"/> Father (name: _____) <input type="checkbox"/> Other—name & relationship to child: _____		

I. Please list the sports your child most likes to take part in. For example: swimming, baseball, skating, skate boarding, bike riding, fishing, etc. <input type="checkbox"/> None	Compared to others of the same age, about how much time does he/she spend in each?				Compared to others of the same age, how well does he/she do each one?			
	Don't Know	Less Than Average	Average	More Than Average	Don't Know	Below Average	Average	Above Average
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	a. _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b. _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

II. Please list your child's favorite hobbies, activities, and games, other than sports. For example: stamps, dolls, books, piano, crafts, cars, singing, etc. (Do not include listening to radio or TV.) <input type="checkbox"/> None	Compared to others of the same age, about how much time does he/she spend in each?				Compared to others of the same age, how well does he/she do each one?			
	Don't Know	Less Than Average	Average	More Than Average	Don't Know	Below Average	Average	Above Average
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	a. _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b. _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

III. Please list any organizations, clubs, teams, or groups your child belongs to. <input type="checkbox"/> None	Compared to others of the same age, how active is he/she in each?			
	Don't Know	Less Active	Average	More Active
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	a. _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b. _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

IV. Please list any jobs or chores your child has. For example: paper route, babysitting, making bed, working in store, etc. (Include both paid and unpaid jobs and chores.) <input type="checkbox"/> None	Compared to others of the same age, how well does he/she carry them out?			
	Don't Know	Below Average	Average	Above Average
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	a. _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b. _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Copyright 1991 T.M. Achenbach, U. of Vermont.

1 S. Prospect St., Burlington, VT 05401 UNAUTHORIZED REPRODUCTION FORBIDDEN BY LAW

1-91 Edition

PAGE 1

- V. 1. About how many close friends does your child have? None 1 2 or 3 4 or more
(Do not include brothers & sisters)
2. About how many times a week does your child do things with any friends outside of regular school hours?
(Do not include brothers & sisters) Less than 1 1 or 2 3 or more

VI. Compared to others of his/her age, how well does your child:

- | | Worse | About Average | Better | |
|---|--------------------------|--------------------------|--------------------------|---|
| a. Get along with his/her brothers & sisters? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> Has no brothers or sisters |
| b. Get along with other kids? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| c. Behave with his/her parents? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| d. Play and work by himself/herself? | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |

VII. 1. For ages 6 and older—performance in academic subjects. If child is not being taught, please give reason _____

- | | Failing | Below average | Average | Above average |
|---|--------------------------|--------------------------|--------------------------|--------------------------|
| a. Reading, English, or Language Arts | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| b. History or Social Studies | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| c. Arithmetic or Math | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| d. Science | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Other academic subjects—for example: computer courses, foreign language, business. Do not include gym, shop, driver's ed., etc. | | | | |
| e. _____ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| f. _____ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| g. _____ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

2. Is your child in a special class or special school? No Yes—what kind of class or school?

3. Has your child repeated a grade? No Yes—grade and reason

4. Has your child had any academic or other problems in school? No Yes—please describe

When did these problems start?

Have these problems ended? No Yes—when?

Does your child have any illness, physical disability, or mental handicap? No Yes—please describe

What concerns you most about your child?

Please describe the best things about your child:

Below is a list of items that describe children and youth. For each item that describes your child now or within the past 6 months, please circle the 2 if the item is very true or often true of your child. Circle the 1 if the item is somewhat or sometimes true of your child. If the item is not true of your child, circle the 0. Please answer all items as well as you can, even if some do not seem to apply to your child.

0 = Not True (as far as you know)			1 = Somewhat or Sometimes True	2 = Very True or Often True			
0	1	2	1. Acts too young for his/her age	0	1	2	31. Fears he/she might think or do something bad
0	1	2	2. Allergy (describe): _____ _____	0	1	2	32. Feels he/she has to be perfect
0	1	2	3. Argues a lot	0	1	2	33. Feels or complains that no one loves him/her
0	1	2	4. Asthma	0	1	2	34. Feels others are out to get him/her
0	1	2	5. Behaves like opposite sex	0	1	2	35. Feels worthless or inferior
0	1	2	6. Bowel movements outside toilet	0	1	2	36. Gets hurt a lot, accident-prone
0	1	2	7. Bragging, boasting	0	1	2	37. Gets in many fights
0	1	2	8. Can't concentrate, can't pay attention for long	0	1	2	38. Gets teased a lot
0	1	2	9. Can't get his/her mind off certain thoughts; obsessions (describe): _____ _____	0	1	2	39. Hangs around with others who get in trouble
0	1	2	10. Can't sit still, restless, or hyperactive	0	1	2	40. Hears sounds or voices that aren't there (describe): _____ _____
0	1	2	11. Clings to adults or too dependent	0	1	2	41. Impulsive or acts without thinking
0	1	2	12. Complains of loneliness	0	1	2	42. Would rather be alone than with others
0	1	2	13. Confused or seems to be in a fog	0	1	2	43. Lying or cheating
0	1	2	14. Cries a lot	0	1	2	44. Bites fingernails
0	1	2	15. Cruel to animals	0	1	2	45. Nervous, highstrung, or tense
0	1	2	16. Cruelty, bullying, or meanness to others	0	1	2	46. Nervous movements or twitching (describe): _____ _____
0	1	2	17. Day-dreams or gets lost in his/her thoughts	0	1	2	47. Nightmares
0	1	2	18. Deliberately harms self or attempts suicide	0	1	2	48. Not liked by other kids
0	1	2	19. Demands a lot of attention	0	1	2	49. Constipated, doesn't move bowels
0	1	2	20. Destroys his/her own things	0	1	2	50. Too fearful or anxious
0	1	2	21. Destroys things belonging to his/her family or others	0	1	2	51. Feels dizzy
0	1	2	22. Disobedient at home	0	1	2	52. Feels too guilty
0	1	2	23. Disobedient at school	0	1	2	53. Overeating
0	1	2	24. Doesn't eat well	0	1	2	54. Overtired
0	1	2	25. Doesn't get along with other kids	0	1	2	55. Overweight
0	1	2	26. Doesn't seem to feel guilty after misbehaving	0	1	2	56. Physical problems without known medical cause:
0	1	2	27. Easily jealous	0	1	2	a. Aches or pains (not headaches)
0	1	2	28. Eats or drinks things that are not food — don't include sweets (describe): _____ _____	0	1	2	b. Headaches
0	1	2	29. Fears certain animals, situations, or places, other than school (describe): _____ _____	0	1	2	c. Nausea, feels sick
0	1	2	30. Fears going to school	0	1	2	d. Problems with eyes (describe): _____ _____
				0	1	2	e. Rashes or other skin problems
				0	1	2	f. Stomachaches or cramps
				0	1	2	g. Vomiting, throwing up
				0	1	2	h. Other (describe): _____ _____

0 = Not True (as far as you know)			1 = Somewhat or Sometimes True			2 = Very True or Often True			
0	1	2	57.	Physically attacks people	0	1	2	84.	Strange behavior (describe): _____
0	1	2	58.	Picks nose, skin, or other parts of body (describe): _____					_____
					0	1	2	85.	Strange ideas (describe): _____

0	1	2	59.	Plays with own sex parts in public	0	1	2	86.	Stubborn, sullen, or irritable
0	1	2	60.	Plays with own sex parts too much	0	1	2	87.	Sudden changes in mood or feelings
0	1	2	61.	Poor school work	0	1	2	88.	Sulks a lot
0	1	2	62.	Poorly coordinated or clumsy	0	1	2	89.	Suspicious
0	1	2	63.	Prefers being with older kids	0	1	2	90.	Swearing or obscene language
0	1	2	64.	Prefers being with younger kids	0	1	2	91.	Talks about killing self
0	1	2	65.	Refuses to talk	0	1	2	92.	Talks or walks in sleep (describe): _____
0	1	2	66.	Repeats certain acts over and over; compulsions (describe): _____					_____
					0	1	2	93.	Talks too much
0	1	2	67.	Runs away from home	0	1	2	94.	Teases a lot
0	1	2	68.	Screams a lot	0	1	2	95.	Temper tantrums or hot temper
0	1	2	69.	Secretive, keeps things to self	0	1	2	96.	Thinks about sex too much
0	1	2	70.	Sees things that aren't there (describe): _____					
					0	1	2	97.	Threatens people
					0	1	2	98.	Thumb-sucking
					0	1	2	99.	Too concerned with neatness or cleanliness
0	1	2	71.	Self-conscious or easily embarrassed	0	1	2	100.	Trouble sleeping (describe): _____
0	1	2	72.	Sets fires					_____
0	1	2	73.	Sexual problems (describe): _____	0	1	2	101.	Truancy, skips school
					0	1	2	102.	Underactive, slow moving, or lacks energy
					0	1	2	103.	Unhappy, sad, or depressed
0	1	2	74.	Showing off or clowning	0	1	2	104.	Unusually loud
0	1	2	75.	Shy or timid	0	1	2	105.	Uses alcohol or drugs for nonmedical purposes (describe): _____
0	1	2	76.	Sleeps less than most kids					_____
0	1	2	77.	Sleeps more than most kids during day and/or night (describe): _____	0	1	2	106.	Vandalism
					0	1	2	107.	Wets self during the day
0	1	2	78.	Smears or plays with bowel movements	0	1	2	108.	Wets the bed
0	1	2	79.	Speech problem (describe): _____					
					0	1	2	109.	Whining
0	1	2	80.	Stares blankly	0	1	2	110.	Wishes to be of opposite sex
0	1	2	81.	Steals at home	0	1	2	111.	Withdrawn, doesn't get involved with others
0	1	2	82.	Steals outside the home	0	1	2	112.	Worries
0	1	2	83.	Stores up things he/she doesn't need (describe): _____					
								113.	Please write in any problems your child has that were not listed above:
					0	1	2		_____
					0	1	2		_____
					0	1	2		_____

PLEASE BE SURE YOU HAVE ANSWERED ALL ITEMS.

PAGE 4

UNDERLINE ANY YOU ARE CONCERNED ABOUT.

APPENDIX B
RATING QUESTIONNAIRE

Name of Judge/Rater:

Date:

Rating Questionnaire

Please rate the following items for their "objectivity", by writing the appropriate number next to the item using the following scale:

1-low objectivity

2-moderately objective

3-highly objective

For this task, "objective/objectivity" is operationally defined as the overtness and/or observability of the behavior. For example, a highly objective behavior is likely to be of a more concrete, quantifiable, or observable nature such as "runs around the block every day", and might not be as influenced by personal feelings of the respondent. On the other hand, an item that may have "low objectivity" (i.e. subjective) is not as easily quantified, observable, or concrete, such as "concerned about low self-esteem", and may be more susceptible to the influence of a respondent's personal feelings.

ITEMS

1. Acts too young for his/her age
2. Allergy (item includes space for description)
3. Argues a lot
4. Asthma
5. Behaves like opposite sex
6. Bowel movements outside toilet
7. Bragging, boasting
8. Can't concentrate, can't pay attention for long
9. Can't get his/her mind off certain thoughts; obsessions (item includes space for description)
10. Can't sit still, restless, or hyperactive
11. Clings to adults or too dependent
12. Complains of loneliness
13. Confused or seems to be in a fog
14. Cries a lot
15. Cruel to animals
16. Cruelty, bullying, or meanness to others
17. Day-dreams or gets lost in his/her thoughts
18. Deliberately harms self or attempts suicide
19. Demands a lot of attention
20. Destroys his/her own things

21. Destroys things belonging to his/her family or others
22. Disobedient at home
23. Disobedient at school
24. Doesn't eat well
25. Doesn't get along with other kids
26. Doesn't seem to feel guilty after misbehaving
27. Easily jealous
28. Eats or drinks things that are not food--don't include sweets (item includes space for description)
29. Fears certain animals, situations, or places, other than school (item includes space for description)
30. Fears going to school
31. Fears he/she might think or do something bad
32. Feels he/she has to be perfect
33. Feels or complains that no one loves him/her
34. Feels others are out to get him/her
35. Feels worthless or inferior
36. Gets hurt a lot, accident prone
37. Gets in many fights
38. Gets teased a lot
39. Hangs around with others who get in trouble
40. Hears sounds or voices that aren't there (includes space for description)
41. Impulsive or acts without thinking
42. Would rather be alone than with others
43. Lying or cheating
44. Bites fingernails
45. Nervous, highstrung, or tense
46. Nervous movements or twitching (includes space for description)
47. Nightmares
48. Not liked by other kids
49. Constipated, doesn't move bowels
50. Too fearful or anxious
51. Feels dizzy
52. Feels too guilty
53. Overeating
54. Overtired
55. Overweight
56. Physical problems without known medical cause (item includes several options such as headaches, rashes cramps, and other)
57. Physically attacks people
58. Picks nose, skin, or other parts of body (includes space for description)
59. Plays with own sex parts in public
60. Plays with own sex parts too much
61. Poor school work
62. Poorly coordinated or clumsy
63. Prefers being with older kids
64. Prefers being with younger kids
65. Refuses to talk

66. Repeats certain acts over and over; compulsions (includes space for description)
67. Runs away from home
68. Screams a lot
69. Secretive, keeps things to self
70. Sees things that aren't there (includes space for description)
71. Self-conscious or easily embarrassed
72. Sets fires
73. Sexual problems (includes space for description)
74. Showing off or clowning
75. Shy or timid
76. Sleeps less than most kids
77. Sleeps more than most kids during day and/or night (includes space for description)
78. Smears or plays with bowel movements
79. Speech problems (includes space for description)
80. Stares blankly
81. Steals at home
82. Steals outside the home
83. Stores up things he/she doesn't need (includes space for description)
84. Strange behavior (includes space for description)
85. Strange ideas (includes space for description)
86. Stubborn, sullen, or irritable
87. Sudden changes in mood or feelings
88. Sulks a lot
89. Suspicious
90. Swearing or obscene language
91. Talks about killing self
92. Talks or walks in sleep (includes space for description)
93. Talks too much
94. Teases a lot
95. Temper tantrums or hot temper
96. Thinks about sex too much
97. Threatens people
98. Thumb sucking
99. Too concerned with neatness or cleanliness
100. Trouble sleeping (includes space for description)
101. Truancy, skips school
102. Underactive, slow moving, or lacks energy
103. Unhappy, sad, or depressed
104. Unusually loud
105. Uses alcohol or drugs for nonmedical purposes (includes space for description)
106. Vandalism
107. Wets self during the day
108. Wets the bed
109. Whining
110. Wishes to be of opposite sex
111. Withdrawn, doesn't get involved with others
112. Worries

References

- Achenbach, T.M. (1991). Manual for the Child Behavior Checklist/4-18 and 1991 Profile. Burlington, VT.: University Associates in Psychiatry.
- Achenbach, T.M., & Edelbrock, C.S. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. Monographs of the Society for Research in Child Development, 46, (Serial No.188), 1-82.
- Achenbach, T.M., & Edelbrock, C. (1983). Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, Vt: Queen City Publishers.
- Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychological Bulletin, 101, 213-232.
- Alvarez, W.F. (1985). The meaning of maternal employment for mothers and their perceptions of their three-year-old children. Child Development, 56, 350-360.
- American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, D.C.: Author.
- Barglow, P., Vaughn, B.E., & Molitor, N. (1987). Effects of maternal absence due to employment on the quality of infant-mother attachment in a low-risk sample. Child Development, 58, 945-954.
- Bates, J.E. (1980). The concept of difficult temperament. Merrill-Palmer Quarterly, 26, 299-319.
- Beck, A.T. (1967). Depression: Causes and Treatment. Philadelphia: University of Pennsylvania Press.
- Berman, P.W. (1980). Are women more responsive than men to the young? A review of developmental and situational variables. Psychological Bulletin, 88, 668-695.

- Bristol, M.M., Gallagher, J.J., & Schopler, E. (1988). Mothers and fathers of young developmentally disabled and nondisabled boys: Adaptation and spousal support. Developmental Psychology, 24, 441-451.
- Brody, G. & Forehand, R. (1986). Maternal perceptions of child maladjustment as a function of the combined influence of child behavior and maternal depression. Journal of Consulting and Clinical Psychology, 54, 237-240.
- Bronstein, P. (1984). Differences in mothers' and fathers' behaviors toward children: A cross-cultural comparison. Developmental Psychology, 20, 995-1003.
- Burrows, K.R., & Kelley, C.K. (1983). Parental interrater reliability as a function of situational specificity and familiarity of target child. Journal of Abnormal Child Psychology, 11, 41-48.
- Christensen, A., Margolin, G. & Sullaway, M. (1992). Interparental agreement on child behavior problems. Psychological Assessment, 4, 419-425.
- Christensen, A., Sullaway, M. & King, C. E. (1983). Systematic error in behavioral reports of dyadic interaction: Egocentric bias and content effects. Behavioral Assessment, 5, 129-140.
- Cohen, D.J., Dibble, E., & Grawe, J.M. (1977). Fathers' and mothers' perceptions of children's personality. Archives of General Psychiatry, 34, 480-487.
- Cohen, F.A. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.
- Conrad, M., & Hammen, C. (1989). Role of maternal depression in perceptions of child maladjustment. Journal of Consulting and Clinical Psychology, 57, 663-667.
- Cowan, C.P. & Cowan, P.A. (1988). Who does what when partners become parents: Implications for men, women, and marriage. Marriage and Family Review, 12, 105-131.

- Crocker, L., & Algina, J. (1986). Introduction to Classical & Modern Test Theory. Orlando, Fl.: Holt, Rinehart and Winston, Inc.
- Cronbach, L.J. (1970). Essentials of Psychological Testing. (3rd ed.). New York: Harper Row.
- Dibble, E., & Cohen, D.J. (1974). Companion instruments for measuring children's competence and arental style. Archives of General Psychiatry, 30, 805-815.
- Duncan, P., Gale, C., & Kilpatrick, D. (1990). Ratings of children in a residential treatment center by child care workers and parents. Journal of Child and Youth Care, 4, 59-71.
- Duncan, P. & Kilpatrick, D.L. (1989). Use of extreme rating categories in ratings of child behavior. Child Study Journal, 19, 51-64.
- Duncan, P., & Kilpatrick, D.L. (1975). The Child Behavior Rating Scale. Unpublished. (Available from P. Duncan, Department of Psychology, University of Victoria, Victoria, B.C., Canada.
- Earls, F. (1976). The fathers (not the mothers): Their importance and influence with infants and young children. Psychiatry, 39, 209-226.
- Ebel, R.L. (1951). Estimation of the reliability of ratings. Psychometrika, 16, 407-424.
- Eisenstadt, T.H., McElreath, L.H., Eyberg, S.M., & McNeil, C.B. (1994). Interparent agreement on the Eyberg Child Behavior Inventory. Child & Family Behavior Therapy, 16, 21-27.
- Eyberg, S. (1974). Eyberg Child Behavior Inventory. (Available from S. Eyberg, Dept. of Clinical and Health Psychology, Box 100165 HSC, University of Florida, Gainesville, FL 32610).
- Fagot, B.I. (1978). The influence of sex of child on parental reactions to toddler children. Child Development, 49, 459-465.
- Ferguson, L.R., Partyka, L.B., & Lester, B.M. (1974). Patterns of parent perception differentiating clinic from nonclinic children. Journal of Abnormal Child Psychology, 2, 169-181.

- Finken, L.L., & Amato, P.R. (1993). Parental self-esteem and behavior problems in children: Similarities between mothers and fathers. Sex Roles, 28, 569-582.
- Fleiss, J.L. (1973). Statistical methods for rates and proportions. New York: John Wiley & Sons.
- Gelfand, D.M., & Hartmann, D.P. (1975). Child behavior analysis and therapy. New York: Pergamon.
- Glick, P.C. (1988). Fifty years of family demography: A record of social change. Journal of Marriage and the Family, 50, 861-873.
- Grieff, G.L. (1985). Single fathers. Lexington, MA: D.C. Heath.
- Griest, D.L., Wells, K.C., & Forehand, R. (1979). An examination of predictors of maternal perceptions of maladjustment in clinic-referred children. Journal of Abnormal Psychology, 88, 277-281.
- Hoffman, L.W. (1977). Changes in family roles, socialization, and sex differences. American Psychologist, 32, 644-657.
- Hollingshead, A.B. (1975). Four factor index of social status. New Haven, CN: Yale University Press, Dept. of Sociology.
- Hulbert, T.A., Gdowski, C.L., & Lachar, D. (1986). Interparent agreement on the Personality Inventory for Children: Are substantial correlations sufficient? Journal of Abnormal Child Psychology, 14, 115-122.
- Jacob, T., Grounds, L., & Haley, R. (1982). Correspondence between parents' reports on the Behavior Problem Checklist. Journal of Abnormal Child Psychology, 10, 593-608.
- Jensen, P.S.; Traylor, J.; Xenakis, S.N. & Davis, H. (1988). Child psychopathology rating scales and interrater agreement: Parents' gender and psychiatric symptoms. Journal of the American Academy of Child and Adolescent Psychiatry, 27, 442-450.
- Johnston, C. (1991). Predicting mothers' and fathers' perceptions of child behavior problems. Canadian Journal of Behavioral Science, 23, 349-357.

- Kagan, J. (1964). Acquisition and significance of sex typing and sex role identity. In M.L. Hoffman & L.W. Hoffman (Eds.), Review of child development research (Vol. 1). New York: Russell Sage Foundation.
- Kagan, J. & Lemkin, I. (1960). The child's differential perception of parental attributes. Journal of Abnormal and Social Psychology, 61, 440-447.
- Kazdin, A.E. (1988). The diagnosis of childhood disorders: Assessment issues and strategies. Behavioral Assessment, 10, 67-94.
- Kazdin, A.E. & Esveldt-Dawson, R. (1983). Assessment of childhood depression: correspondence of child and parent ratings. Journal of the American Academy of Child Psychiatry, 22, 157-164.
- Kazdin, A.E., French, N.H., & Unis, A.S. (1982). Child, mother, and father evaluations of depression in psychiatric inpatient children. Journal of Abnormal Child Psychology, 11, 167-179.
- Keppel, G. (1991). Design and Analysis: A Researcher's Handbook (3rd ed.). New Jersey: Prentice Hall.
- Kilpatrick, D., & Duncan, P. (1985). The effect of sex of rater and sex of child on adult ratings of child behavior. Child Study Journal, 15, 237-249.
- Kimmel, D.C., & Vanderveen, F. (1974). Factors of marital adjustment in Locke's Marital Adjustment Test. Journal of Marriage and the Family, 36, 57-63.
- Konstantareas, M.M. & Homatidis, S. (1989). Parental perception of learning-disabled children's adjustment problems and related stress. Journal of Abnormal Child Psychology, 17, 177-186.
- Kovacs, M. (1981). Rating scales to assess depression in school aged children. Acta Paedopsychiatrica, 46, 305-315.
- Lachar, D., & Gdowski, C.L. (1979). Actuarial assessment of child and adolescent personality: An interpretive guide for the Personality Inventory for Children profile. Los Angeles: Western Psychological Services.

- Lamb, M.E. (1975). Fathers: Forgotten contributors to child development. Human Development, 18, 245-266.
- Lamb, M.E. (1977). Father-infant and mother-infant interaction in the first year of life. Child Development, 48, 167-181.
- Lamb, M.E. (1986). The changing roles of fathers. In M.E. Lamb (Ed.), The father's role: Applied perspectives (pp.3-27). New York: Wiley.
- Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.
- LeBlanc, R. & Reynolds, C.R. (1989). Concordance of mothers' and fathers' ratings of children's behavior. Psychology in the Schools, 26, 225-229.
- Leon, G.R., Kendall, P.C. & Garber, J. (1980). Depression in children: parent, teacher, and child perspectives. Journal of Abnormal Child Psychology, 8, 221-235.
- Lindholm, B.W., & Touliatos, J. (1981). Mothers' and fathers' perceptions of their children's psychological adjustment. The Journal of Genetic Psychology, 139, 245-255.
- Lyon, M.E., & Plomin, R. (1981). The measurement of temperament using parental ratings. Journal of Child Psychology and Psychiatry, 22, 47-53.
- Maccoby, E.E. & Jacklin, C.N. (1974). The psychology of sex differences. Stanford: Stanford University Press.
- Marsh, D.T., Stoughton, N.L., & Williams, T.A. (1985). Effects of role, gender, age, and parental status on perception of childhood problems. Exceptional Children, 52, 170-177.
- Martin, R.P. (1988). The Temperament Assessment Battery for Children: Manual. Brandon, VT: Clinical Psychology Press.
- Martin, R. & Halvorsen, C. (1991). Mother-father agreement in temperament ratings. In J. Strelau & A. Angleitner (Eds.), Explorations in temperament (pp.235-248). New York: Plenum Press.
- McDevitt, S.C., & Carey, W.B. (1978). The measurement of

- temperament in 3- to 7-year old children. Journal of Child Psychology and Psychiatry, 19, 245-253.
- McMahon, R.J. (1984). Behavior checklists and rating scales. In T.H. Ollendick & M. Hersen (Eds.), Child behavioral assessment: Principles and procedures (pp. 80-105). New York: Pergamon Press.
- Meyer, J.W., & Sobieszek, B.I. (1972). Effect of a child's sex on adult interpretation of its behavior. Developmental Psychology, 6, 42-48.
- Mischel, W. (1966). A social learning view of sex differences in behavior. In E.E. Maccoby (Ed.), The development of sex differences. Stanford: Stanford University Press.
- Montemayor, R. (1982). The relationship between parent-adolescent conflict and the amount of time adolescents spend alone and with parents and peers. Child Development, 53, 1512-1519.
- Montemayor, R. & Brownlee, J. (1987). Fathers, mothers and adolescents: Gender based differences in parental roles during adolescence. Journal of Youth and Adolescence, 16, 281-291.
- Novick, J., Rosenfeld, E., Black, D.A., & Dawson, D. (1966). Ascertaining deviant behavior in children. Journal of Consulting Psychology, 30, 230-238.
- Nunnally, J.C. (1967). Psychometric Theory. New York: McGraw-Hill.
- Paguio, L.P. (1983). The influence of sex of child and parent on perceptions of the ideal child. Child Study Journal, 13, 187-194.
- Patterson, G.R. (1980). Mothers: The unacknowledged victims. Monographs of the Society for Research in Child Development, 45 (5, Serial No. 186).
- Peplau, L.A. (1983). Roles and gender. In H.H. Kelley, E. Berscheid, A. Christensen, J.H. Harvey, T.L. Huston, G. Levinger, E. McClintock, L.A. Peplau, & D.R. Peterson (Eds.), Close relationships (pp. 220-264). New York: Freeman.
- Peterson, C.C., & Peterson, J.L. (1973). Preference for sex of offspring as a measure of change in sex attitudes. Psychology, 10, 3-5.

- Petti, T.A. (1978). Depression in hospitalized child psychiatry patients: Approaches to measuring depression. Journal of the American Academy of Child Psychiatry, 17, 49-59.
- Pfeffer, J., & Martin, R.P. (1983). Comparison of mothers' and fathers' temperament ratings of referred and nonreferred preschool children. Journal of Clinical Psychology, 39, 1013-1020.
- Phares, V. (1992). Where's poppa? American Psychologist, 47, 656-664.
- Phares, V. & Compas, B.E. (1992). The role of fathers in child and adolescent psychopathology: Make room for daddy. Psychological Bulletin, 111, 387-412.
- Prinz, R.J., Foster, S., Kent, R.N., & O'Leary, K.D. (1979). Multivariate assessment of conflict in distressed and nondistressed mother-adolescent dyads. Journal of Applied Behavioral Analysis, 12, 691-700.
- Quay, H.C. (1977). Measuring dimensions of deviant behavior: The Behavior problem Checklist. Journal of Abnormal Child Psychology, 5, 277-287.
- Richardson, L. (1988). The dynamics of sex and gender: A sociological perspective (3rd ed.). New York: Harper & Row Publishers, Inc.
- Richters, J., & Pellegrini, D. (1989). Depressed mothers' judgements about their children: An examination of the depression-distortion hypothesis. Child Development, 60, 1068-1075.
- Robinson, E.A., Eyberg, S.M., & Ross, A.W. (1980). The standardization of an inventory of child conduct problem behaviors. Journal of Clinical Child Psychology, 9, 22-28.
- Rothbart, M.K. & Maccoby, E.E. (1966). Parents' differential reactions to sons and daughters. Journal of Personality and Social Psychology, 4, 237-243.
- Shaughency, E.A., & Lahey, B.B. (1985). Mothers' and fathers' perceptions of child deviance: Roles of child behavior, parental depression, and marital satisfaction. Journal of Consulting and Clinical Psychology, 53, 718-723.

- Sheriffs, A.C., & Jarrett, R.F. (1953). Sex differences in attitudes about sex differences. Journal of Psychology, 35, 161-168.
- Siegal, M. (1987). Are sons and daughters treated more differently by fathers than by mothers? Developmental Review, 7, 183-209.
- Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.
- Simpson, M. (1935). Parent Preferences of young children. Teachers College of Columbia University Contributions to Education, No. 652.
- Sines, J.O., Pauker, J.D., Sines, L.K., & Owen, D.R. (1969). Identification of clinically relevant dimensions of children's behavior. Journal of Consulting and Clinical Psychology, 33, 728-734.
- Speer, D.C. (1971). Behavior Problem Checklist (Peterson-Quay): Base-line data from parents of child guidance and nonclinic children. Journal of Consulting and Clinical Psychology, 36, 221-228.
- Steinberg, L. (1987). Recent research on the family at adolescence: The extent and nature of sex differences. Journal of Youth and Adolescence, 16, 191-197.
- Suen, H.K. & Ary, D. (1989). Analyzing quantitative behavioral observation data. Hillside, N.J.: Lawrence Erlbaum Associates.
- Tasch, R.G. (1952). The role of the father in the family. Journal of Experimental Education, 20, 319-361.
- Thompson, L., & Walker, A.J. (1991). Gender in families: Women, and men in marriage, work and parenthood. In A. Booth (Ed.), Contemporary Families: Looking Forward, looking back. Minneapolis, MN: National Council on Family Relations.
- Thompson, R.J., Curry, J.F., & Yancy, W.S. (1979). The utility of parents' behavior checklist ratings with developmentally disabled children. Journal of Pediatric Psychology, 4, 19-28.

- Thompson, R.J., & McAdoo, W.G. (1973). A comparison of mothers' and fathers' behavior checklist ratings of outpatient boys and girls. Journal of Community Psychology, 1, 387-389.
- Torrance, E.P. (1962). Guiding creative talent. Englewood Cliffs, NJ: Prentice-Hall.
- Torrance, E.P. (1975). Preliminary manual: Ideal child checklist. Athens, GA: Georgia Studies of Creative Behavior.
- Webster-Stratton, C. (1988). Mothers' and fathers' perceptions of child deviance: Roles of parent and child behaviors and parent adjustment. Journal of Consulting and Clinical Psychology, 56, 909-915.
- Webster-Stratton, C., & Hammond, M. (1988). Maternal depression and its relationship to life stress, perceptions of child behavior problems, parenting behaviors, and child conduct problems. Journal of Abnormal Child Psychology, 16, 299-315.
- Weinberg, W.A., Rutman, J., Sullivan, L., Penick, E.C., & Dietz, S.G. (1973). Depression in children referred to an educational diagnostic center: Diagnosis and treatment. Journal of Pediatrics, 83, 1065-1072.
- Weinraub, M., & Frankel, J. (1977). Sex differences in parent-infant interaction during free play, departure, and separation. Child Development, 48, 1240-1249.
- Weissman, M.M., Orvaschel, H., & Padian, N. (1980). Children's symptom and social functioning self-report scales: Comparison of mothers' and children's reports. The Journal of Nervous and Mental Disease, 168, 736-740.
- Wirt, R.D., Lachar, D., Klinedinst, J.K. & Seat, P.D. (1984). Multidimensional description of child personality: A manual for the Personality Inventory for Children. Los Angeles, CA: Western Psychological Services.
- Worobey, J. (1987). Temperament ratings in early childhood by parents, teachers, and students. Early Childhood Research Quarterly, 2, 169-179.