

Convergence of Clinically Derived Diagnoses and Parent Checklists Among Inpatient Children¹

Alan E. Kazdin² and Irene E. Heidish

University of Pittsburgh School of Medicine

The extent to which parent rating scales differentiated children according to DSM III diagnoses was examined. A total of 113 psychiatric inpatient boys (ages 6-11) were rated by their mothers or maternal figures on the Child Behavior Checklist (CBCL) and the Behavior Problem Checklist (BPC). Children with DSM III diagnoses of conduct disorder or depression were compared to children without these diagnoses. Externalizing and internalizing scales of the parent checklists and additional measures of child aggression and depression differentiated children according to major diagnoses. The use of parent checklists to classify children indicated a high level of sensitivity for both CBCL and BPC scales for diagnosing conduct disorder and depression. However, specificity of the subscales, particularly for the CBCL, was relatively low, indicating a high rate of false positives. The need for further work that extends the range of diagnosis, that examines subtypes of disorders, and that increases the specificity of the measures for diagnostic purposes is discussed.

Classification of childhood (and adult) psychopathology has followed two major traditions. In the first, diagnosis is based on identifying discrete types of dysfunction. Disorders and the specific symptoms of which they are com-

¹Completion of this project was supported by a Research Scientist Development Award (MH00353) and a grant (MH35408) from the National Institute of Mental Health, and by a Clinical Research Center Grant for the Study of Affective Disorders (5 P50 MH30915) from the National Institute of Mental Health. The authors are grateful to Elaine Meyer, Karen Esveldt-Dawson, Antoinette Rodgers, and the clinical research team and staff of the Child Psychiatric Treatment Service.

²Address all correspondence to Alan E. Kazdin, Department of Psychiatry, University of Pittsburgh School of Medicine, Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15213.

posed are diagnosed as present or absent. Categorical diagnosis is reflected in the *Diagnostic and Statistical Manual of Mental Disorders* (DMS III; American Psychiatric Association, 1980), which focuses on different diagnostic entities. Categorical or qualitative diagnosis has emerged from the tradition of clinical observation where abstractions are culled from direct clinical experience. Diagnoses usually are made through clinical interviews. For children and adolescents, several structured and semistructured interviews have been developed for this purpose (e.g., Chambers, Puig-Antich, & Tabrizi, 1978; Herjanic, Herjanic, Brown, & Wheatt, 1975; Hodges, McKnew, Cytryn, Stern, & Kline, 1982).

In the second tradition, diagnosis is dimensional rather than categorical. Children are evaluated on several different dimensions (e.g., factors) usually based on standardized rating scales and checklists. Each child has a score for multiple characteristics, traits, or facets included in the measure. Dimensional assessment has relied heavily on checklists completed by parents and teachers (e.g., Achenbach, 1978; Conners, 1969; Quay, 1977). Standardized scales usually provide factor scores from empirically derived groups of items, normative data to permit comparisons with clinic populations, and quantitative information for a child on all dimensions in the scale (e.g., Achenbach & Edelbrock, 1981; Miller, 1972).

Qualitative and quantitative traditions, and clinical and multivariate methods with which they are associated, often are pitted against each other. However, the need to examine the interrelationship of these different approaches has become increasingly recognized if clinical diagnosis is to be placed on firm empirical ground. The emphasis on descriptive and objective diagnostic criteria within DSM III encourages the empirical evaluation of diagnostic entities using assessment and quantitative methods of the dimensional approaches. Multivariate techniques can be used to identify syndromes, and parallels between clinical and empirically derived systems can be examined (e.g., Achenbach, 1980; Quay, 1979). Similarly, comparisons can be made directly between clinical diagnoses and performance on standardized assessment techniques.

At present, it is unclear if quantitative assessment devices differentially reflect clinically derived diagnoses. Studies of different scales typically have evaluated differences between clinic and nonclinic populations, or delinquents and nondelinquents, or psychiatric and other (e.g., learning-disabled) populations (Quay & Peterson, 1979). Few studies have examined the relationship between clinical diagnosis (using DSM III criteria) and performance on standardized measures of psychopathology to assess if these latter measures can differentiate clinically identified groups.

The present investigation examined the relationship between two frequently used parent rating scales (the Child Behavior Checklist and the

Behavior Problem Checklist) and psychiatric diagnosis in a seriously disturbed clinical population. The purpose was to evaluate if clinically derived diagnostic groups could be differentiated on these measures. The diagnoses of conduct disorder and major depression were examined. Within categorical and dimensional assessment, these diagnoses have been considered to reflect different types of dysfunction (e.g., behavioral vs. emotional, externalizing vs. internalizing, under- vs. overcontrolled). The investigation examined the performance of conduct-disorder and depressed children on standardized measures and the sensitivity and specificity of subscales in identifying cases. To further examine the extent to which standardized measures differentiated among diagnostic groups, interview and paper-and-pencil measures that focused on specific features of conduct disorder and depression (i.e., aggression and depression) were also included.

METHOD

Subjects

Subjects were 113 boys and their mothers or maternal parent figures. The children were all inpatients of a psychiatric facility where children are hospitalized for 2 to 3 months. The facility houses 22 children (ages 5-13 years) at one time who are admitted for acute disorders including highly aggressive and destructive behavior, suicidal or homicidal ideation or behavior, psychotic episodes, or deteriorating family conditions. Inclusion criteria were a Verbal or Full Scale WISC-R IQ of 70 or above and no evidence of neurological impairment, acute confusional state, uncontrolled seizure disorder, or dementia. Children were selected from consecutive hospital admissions. The sample was restricted to boys because the factor structure of one of the primary measures, described below, has been shown to differ for girls and boys (Achenbach & Edelbrock, 1981) and because large numbers of girls with a diagnosis of conduct disorder were not available to permit evaluation of this diagnosis. The boys ranged in age from 6 through 11 yrs ($M = 9.4$) and in Full Scale IQ from 70 to 126 ($M = 92.3$). Eighty-two children were white; 31 were black.

Diagnoses of the children, based on DMS III criteria, were obtained from direct interviews with the children and their parent(s) immediately prior to admission and from psychiatric evaluation after the child had been admitted. Diagnoses were reached without reference to or use of the measures included in the present investigation. On the basis of the above

sources of information, two staff independently completed the diagnoses. Agreement on principal Axis I diagnosis was relatively high ($Kappa = .77$). In cases of disagreements, the child was discussed to reach consensus on the appropriate diagnosis. The principal Axis I diagnoses included major depression ($n = 11$), conduct disorder ($n = 50$), attention deficit disorder ($n = 21$), adjustment disorder ($n = 8$), anxiety disorder ($n = 4$), psychoses ($n = 6$), and other mental disorders ($n = 13$).

The children's mothers or maternal parent figures ranged in age from 23 to 59 years ($M = 33.7$ years). They included natural parents ($n = 93$), step-, foster, or adoptive mothers ($n = 16$), or other female relative or guardian ($n = 4$). Family social class, calculated by the Hollingshead two-factor index, yielded the following breakdown: Classes V (6%), IV (52%), III (30%), II (12%), and I (0%). Estimated monthly income for families ranged from 0 to \$500 to more than \$2,500 (Mdn. range = \$500 to \$1,000).

Assessment

Children and their parent(s), as available, were interviewed separately prior to admission, at which point initial diagnostic information was obtained. All standardized measures were obtained within the first 2 weeks of the child's admission to the hospital in meetings in which parents met with social workers for routine intake procedures. In addition to diagnostic interviews, the parent(s) completed rating scales to assess a broad range of child dysfunction. Because of interest in the present investigation in examining diagnoses of conduct disorder and depression, separate measures of aggression and depression were also completed by the parents, as described below.

Parent Checklists of Child Psychopathology. The children's mothers completed the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1981). The measure includes 118 items scored on a 0-to-2-point scale to cover multiple symptoms areas that have been derived through factor analyses separately by child age and gender. For boys, 6-11, the age group of the present patient sample, the first-order factors include schizoid, depressed, uncommunicative, obsessive-compulsive, somatic complaints, social withdrawal, hyperactive, aggressive, and delinquent. These first-order (narrow-band) factors constitute two broad-band factors (internalizing and externalizing), which are also examined. (The CBCL also includes three social competence scales that were not of direct interest and hence were not included in the present report.)

The parent also completed the Behavior Problem Checklist (BPC; Quay, 1977; Quay & Peterson, 1979). The BPC contains 55 items that describe diverse problem behaviors across a number of settings. For each

item, the parent scored the problem as present or absent. The items constitute five subscales, including conduct problem, personality problem, inadequacy-immaturity, socialized delinquency, and psychotic behavior or signs. (Six items are not encompassed by these scales and were not utilized for the present investigation.)

The CBCL and BPC yield raw scores for individual scales (factors). To facilitate comparison of data from these different measures, normalized *T* scores were used, with a mean of 50 and standard deviation of 10. The CBCL yields these *T* scores on the basis of data from normative samples for children of the same gender and within the same age range (6-11 years) of the patients in the present project (see Achenbach, 1978). For the BPC, normative data to derive *T* scores were obtained from Touliatos and Lindholm (1981), who assessed over 1,000 children from kindergarten through grade 8. For present purposes, means and standard deviations from mothers' ratings of boys ($n = 503$) within the age range of interest were utilized to provide the data to derive *T* scores for the present investigation.³ Thus, for both the CBCL and the BPC, a *T* score of 60 on a subscale reflects 1 standard deviation above the mean for nonreferred children.

Aggression and Depression Measures. In addition to checklists assessing multiple facets of dysfunction, paper-and-pencil and interview measures of aggression and depression were also administered to the parents. The measures, described below, have been examined in previous research evaluating their convergent, discriminant, and criterion validity for inpatient children (Kazdin, Esveldt-Dawson, Unis, & Rancurello, 1983; Kazdin, French, Unis, & Esveldt-Dawson, 1983).

The paper-and-pencil measure for aggression was the Hostility-Guilt Inventory (HGI), which consisted of 38 items in a true-false format. The parent identifies if the statement is characteristic of the child. The scale was derived from the Buss-Durkee Hostility-Guilt Inventory, which measures several areas of aggression, including assaultive behavior, negativism, irritability, resentment, verbal aggression, and guilt (Buss & Durkee, 1957). Items were selected from each facet to assess the full spectrum of aggressive behavior.

The *Interview for Aggression (IA)* is a semistructured interview (Kazdin, Esveldt-Dawson, Unis, & Rancurello, 1983). The measure includes 30 items that pertain to aggression, such as getting into fights, threatening others, starting arguments, teasing, and so on. Each item is rated on a 5-point scale for severity (1 = not at all, 5 = very much) and 3-point scale

³The authors are grateful to Dr. John Touliatos for making data from his research available to us to make these computations possible.

for duration (1 = recent or new problem, 3 = always). Total aggression was obtained by summing severity and duration scores.

The paper-and-pencil measure of depression was the Children's Depression Inventory (CDI, Kovacs, 1981). The measure is patterned after the Beck Depression Inventory and includes 27 items to assess affective, cognitive, and behavioral symptoms of depression. For each item, the parent selects one of three sentences (0-to-2-point scale) that best describes the child over the past 2 weeks.

The Bellevue Index of Depression (BID) is a semistructured interview for children or their parents (Petti, 1978). The modified version (Kazdin, French, Unis, & Esveldt-Dawson, 1983) includes 26 items that pertain to symptoms of depression such as looking sad, crying easily, thinking about death, losing interest in activities, and others. Symptoms were rated by the parent separately on a 5-point scale for severity and a 3-point scale for duration, as for the IA noted above. Total depression was obtained by summing severity and duration ratings.

RESULTS

Preliminary Analyses

To examine the relationship of patient and demographic variables and performance on the CBCL and BPC and other dependent measures, analyses of variance were completed for child age, gender, race, and Full Scale IQ and for parent age, employment status, income, and family Hollingshead classification. The only variable to emerge was that of race. White children were significantly higher than black children on the CBCL scales measuring depression ($F(1, 112) = 7.55, p < .01$) and somatic complaints ($F(1, 112) = 6.24, p < .05$).

A purpose of the investigation was to examine the extent to which the CBCL and BPC were related to childhood diagnosis and the extent to which subscales discriminated among independently derived diagnostic groups. The CBCL and BPC include factorally derived scales, some of which overlap in the sorts of problems that are assessed. Pearson correlations were computed between the subscales of the two measures to examine their interrelationships. High correlations between scales of the different measures could indicate that the scales provide redundant information.

The results (Table I) indicated several significant correlations in the low to moderate range. The highest correlations tend to be the expected relationships among scales reflecting similar dimensions. For example, correlations of the CBCL scales for aggression, delinquency, and externalizing behavior with the BPC conduct problem scale reflect the highest correla-

Table 1. Pearson Product-Moment Correlations Between Scales of the Child Behavior Checklist (CBCL) and the Behavior Problem Checklist (BPC) ($N = 113$)

	BPC				
	Conduct problem	Personality problem	Inadequacy immaturity	Socialized delinquency	Psychotic behavior
CBCL					
Schizoid	.06 ^c	.44 ^c	.28 ^b	-.04	-.03
Depressed	.28 ^c	.52 ^c	.30 ^c	-.03	.18 ^a
Uncommunicative	.26 ^b	.52 ^c	.38 ^c	.13	.13
Obsessive-compulsive	.53 ^c	.45 ^c	.39 ^c	.06	.33 ^a
Somatic complaints	.10	.17	.04	-.13	.07
Social withdrawal	.26 ^b	.42 ^c	.37 ^c	-.01	.13
Hyperactive	.32 ^c	.36 ^c	.37 ^c	-.02	.28 ^b
Aggressive	.68 ^c	.12	.19 ^a	.26 ^b	.23 ^a
Delinquent	.53 ^c	.02	.19 ^a	.36 ^c	.16
Internalizing	.40 ^c	.53 ^c	.36 ^c	.04	.22 ^a
Externalizing	.63 ^c	.15	.26 ^b	.26 ^b	.25 ^b

^a $p < .05$.^b $p < .01$.^c $p < .001$.

tions in the table. In general, the correlations indicate consistencies across the measures that would be expected given overlap in item content and the use of the same rater (parents). Interestingly, the correlations are not of such high magnitude to suggest heavily shared variance. In subsequent analyses, the subscales of each measure were analyzed separately.

Psychiatric Diagnosis and Parental Ratings

A major purpose of the study was to evaluate the extent to which parental ratings were consistent with and differentially reflected DSM III diagnoses. Subscales of the CBCL and BPC that reflected symptom areas consistent with the diagnoses of conduct disorder and major depression were evaluated. The expectation was that subscales reflecting acting-out behaviors ("externalizing" dimensions) would differ between conduct- and non-conduct-disordered children and that subscales reflecting inner-directed, emotional problems ("internalizing" dimensions) would differ between depressed and nondepressed children. Because several children received multiple Axis I diagnoses, comparisons were made separately for conduct disorder versus other diagnoses and depression versus other diagnoses. Preliminary analyses by these diagnostic groupings indicated no differences on subject and demographic variables.

The means and standard deviations for the entire sample and separate means by diagnostic groups are presented for each measure in Table II. Analyses of variance comparing children with ($n = 72$) and without ($n =$

Table II. *T* Score Means for the Total Sample as a Function of DSM III Diagnoses of Conduct Disorder and Depression

Measures	Total sample (<i>n</i> = 113)	Conduct disorder (<i>n</i> = 72)	No conduct disorder (<i>n</i> = 41)	<i>F</i> (1, 112)	Depression (<i>n</i> = 26)	No depression (<i>n</i> = 87)	<i>F</i> (1, 112)
CBCL							
Schizoid <i>SD</i>	66.7 9.8	65.5	68.9	2.91	68.3	66.2	< 1
Depressed <i>SD</i>	70.5 9.9	69.8	71.9	1.20	74.7	69.3	5.97 ^a
Uncommunicative <i>SD</i>	70.3 11.3	70.1	70.7	< 1	75.3	68.8	6.76 ^a
Obsessive-compulsive <i>SD</i>	71.0 10.0	69.4	72.0	< 1	74.2	70.0	3.45
Somatic complaints <i>SD</i>	64.8 9.9	64.4	65.6	< 1	70.0	63.3	9.53 ^b
Social withdrawal <i>SD</i>	70.1 9.3	70.1	68.6	1.67	73.1	69.2	3.39
Hyperactive <i>SD</i>	71.3 7.5	71.7	70.5	< 1	71.2	71.3	< 1
Aggressive <i>SD</i>	75.3 10.4	77.8	70.7	12.30 ^c	78.0	74.4	2.32
Delinquent <i>SD</i>	75.0 8.4	77.3	70.8	16.92 ^c	76.5	74.6	1.00
Internalizing <i>SD</i>	73.2 10.5	72.9	73.6	< 1	78.3	71.6	8.40 ^b
Externalizing <i>SD</i>	76.3 8.9	78.5	72.3	13.07 ^c	78.3	75.7	1.73

BPC									
Conduct problem	66.1	69.1	61.1	12.37 ^c	69.4	65.0	2.72		
SD	10.8								
Personality problem	59.2	57.8	61.5	2.53	61.3	58.5	1.08		
SD	10.6								
Inadequacy-immaturity	60.5	60.8	60.0	<1	63.7	59.5	1.31		
SD	14.9								
Socialized delinquency	67.0	73.7	56.0	9.82 ^b	74.6	64.5	2.34		
SD	26.5								
Psychotic behavior	66.4	64.7	69.2	<1	71.9	64.5	1.80		
SD	22.0								
Aggression									
HGI	25.1	26.6	22.5	10.90 ^b	26.8	24.5	2.36		
SD	6.0								
IA	113.4	118.9	104.4	4.05 ^c	123.6	110.2	2.61		
SD	33.5								
Depression									
CDI	19.9	19.2	20.6	<1	24.4	18.3	10.75 ^b		
SD	10.3								
BID	95.0	98.2	93.7	<1	109.8	92.5	8.27 ^b		
SD	27.9								

^a*p* < .05.

^b*p* < .01.

^c*p* < .001.

41) a primary or secondary Axis I diagnosis of conduct disorder yielded several differences. Conduct-disorder children, when compared to children without this diagnosis, were rated on the CBCL as significantly higher on the aggressiveness, delinquency, and broad-band externalizing scales, and on the BPC as significantly higher on the conduct problem and socialized delinquency scales. These differences are in the expected direction and convey that subscales of both the CBCL and the BPC reflect differences consistent with the diagnosis of conduct disorder. Moreover, subscales of the CBCL and BPC that reflect more internalizing dimensions (e.g., schizoid, uncommunicativeness, personality problem) and the internalizing scale do not differ between groups. Thus, parent ratings did not simply show greater severity of all symptom areas for conduct disorder children but only in those areas of acting-out behaviors. Parents also completed paper-and-pencil and interview measures of aggression and depression. Conduct-disordered children were rated as significantly more aggressive on the HGI and IA measures than non-conduct-disordered children but were not different on the measures of depression.

Analyses of variance also compared children with a primary or secondary Axis I diagnosis of major depression ($n = 26$) versus children with no depression ($n = 87$). As is evident in Table II, depressed children were significantly higher than nondepressed children on several internalizing scales of the CBCL, including depression, uncommunicativeness, somatic complaints, and the broad-band internalizing scale. Depressed and nondepressed children were not rated as different on the externalizing subscales. Depressed and nondepressed children also did not differ significantly on the BPC scales.

Parent ratings of paper-and-pencil and interview measures of depression also differentiated among diagnostic groups. Children with a diagnosis of major depression were significantly higher in CDI and BID depression than their peers (Table II). Yet they were not significantly different in parent ratings of aggression on the paper-and-pencil and interview measures.

Sensitivity and Specificity

The extent to which parent ratings distinguished diagnostic groups can be examined by computing sensitivity and specificity (Vecchio, 1966). Sensitivity is the percentage of persons who have the diagnosis who are positively identified by a measure (i.e., true positives); specificity is the percentage of persons who do not have the diagnosis and who are not identified by the measure (i.e., true negatives).

Sensitivity and specificity of identifying cases with diagnoses of conduct disorder or depression were evaluated by using subscales of the CBCL and BPC that reflected symptoms associated with these disorders. For the CLCL, three subscales were examined to evaluate conduct disorder children – namely, aggression, delinquency, and externalizing (broad-band). For the BPC, conduct problem and socialized delinquency scales were examined. For the diagnosis of depression, two CBCL scales (depression, internalizing) were examined. The BPC has no specific scale identified as a measure of depression. The scale measuring personality problems was selected on an a priori basis because the items encompassed multiple symptoms of depression (e.g., sadness, lack of self-confidence, anxiety, crying over minor annoyances, feelings of inferiority).

Contingency tables (2×2) were constructed separately by DSM III diagnosis (conduct disorder vs. no conduct disorder), and separate cutoff scores were explored in standard deviation units for classifying by diagnosis. The results indicated that above and below 1 standard deviation maximized sensitivity and specificity relative to other cutoff points.

Table III presents the sensitivity and specificity of the CBCL and BPC subscales by diagnosis. Sensitivity presents the proportion of children with the score equal to or greater than 1 standard deviation above the mean (T score = 60) who had the diagnosis of conduct disorder. Specificity presents the proportion of children with the score below this criterion who did not receive the diagnosis of conduct disorder. As evident in the table, both the CBCL and the BPC scales have a high degree of sensitivity in identifying

Table III. Sensitivity and Specificity of Selected CBCL and BPC Subscales for the Diagnosis of Conduct Disorder and Major Depression^a

Measure	Sensitivity	Specificity
Child Behavior Checklist		
Aggression	68/72 (94.4%)	5/41 (12.2%)
Delinquency	69/72 (95.8%)	3/41 (7.3%)
Externalizing	70/72 (97.2%)	3/41 (7.3%)
Behavior Problem Checklist		
Conduct Problem	64/72 (88.9%)	18/41 (43.9%)
Socialized Delinquency	54/72 (75.0%)	25/41 (60.1%)
		Depression
Child Behavior Checklist		
Depression	23/26 (88.5%)	17/87 (19.5%)
Internalizing	26/26 (100.0%)	13/87 (14.9%)
Behavior Problem Checklist		
Personality Problem	16/26 (61.5%)	39/87 (44.8%)

^aThe criterion score for each scale was 1 standard deviation above the mean (T score of 60 or greater).

conduct disorder children. (CBCL scales tend to be higher than BPC scales in sensitivity. On the other hand, the BPC scales tend to be higher in their specificity. This means that there were more false positives using the cutoff score on the scales for the CBCL than for the BPC.

A separate analysis was made that utilized information from both CBCL and BPC scales. If children received a score of at least 1 standard deviation above the mean on either the CBCL-aggression or BPC-conduct problem, sensitivity was 98.6%. But specificity was low (10%). Similarly, if children received a score at least 1 standard deviation above the mean on either CBCL externalizing or BPC conduct problem scales, sensitivity rose to 100%, but specificity was only 4.8%.

For the diagnosis of depression, CBCL depression and internalizing scales were evaluated in 2×2 contingency tables using a score of 1 standard deviation above the mean as the cutoff on each scale. Again, the CBCL scales showed high sensitivity but relatively low specificity (Table III). Evaluation of the BPC scale of personality problem that includes several symptoms of depression yielded a sensitivity of 61.5% and specificity of 45.3%. Although not predicted, the psychotic behavior scale of the BPC had relatively high sensitivity (80.8%) and specificity (36.0%).

DISCUSSION

The present study was designed to examine the relationship between DSM III and parent rating scales. The results indicated generally that parent ratings differentiated among diagnostic groups. As expected, scales of the CBCL (aggression, delinquency, externalizing) and BPC (conduct problem, socialized delinquency) that reflect symptoms of "acting-out" differentiated children with and without an Axis I diagnosis of conduct disorder. Similarly, scales of the CBCL (depression, uncommunicativeness, somatic complaints, internalizing) that reflected inner-directed or emotional problems distinguished children with and without a diagnosis of depression. Interview and paper-and-pencil measures of aggression and depression completed by the parents also were significantly different among diagnostic groups in the predicted direction.

Although parent rating scales encompassing many areas of child functioning and more narrowly focused measures of aggression and depression differentiated patients by diagnoses, the use of assessment devices for classification purposes is another matter. To examine the extent to which the parent scales could be used for classification, sensitivity and specificity of specific subscales were examined. Cutoff scores were identified to maximize the extent to which children with a diagnosis of conduct disorder or

depression could be selected. using a cutoff score of 1 standard deviation above the mean, sensitivity in identifying conduct disorder children was quite high (range = 75 to 97.2%) among the pertinent CBCL and BPC scales. Specificity of the measures was relatively high for the BPC scales (range = 43.9 to 60.1%) but quite low for the CBCL scales (range = 7.3 to 12.2%). Low specificity means that a high percentage of children without the diagnosis of conduct disorder also obtained scores above the criterion score (i.e., false positives). In general, the BPC scales were only slightly lower than the CBCL scales in sensitivity in identifying conduct disorder but much higher in specificity. The diagnosis of depression revealed that the CBCL subscales showed high sensitivity (using the criterion of 1 standard deviation). Indeed, 100% of depressed children exceeded the cutoff score on the broad-band internalizing scale. Specificity was relatively low, indicating a high rate of false positives on the depression and internalizing scales of the CBCL (range = 14.9 to 19.5%).

The present results illustrate the relationship between parent rating scales and clinically derived diagnoses. However, there are several important limitations and qualifications of the present findings. First, the relationships between clinically derived diagnoses and standard measures needs to be interpreted cautiously. The ability of parent ratings to distinguish groups and the general agreement among measures might be expected given the common source of information upon which the measures relied. In both diagnostic and rating measures, parents played a central role. Thus, convergence of different types of measures do not necessarily validate DSM III diagnoses or empirically derived factors. Without an independent criterion to evaluate the diagnoses, the results need to be interpreted in terms of agreement among alternative measurement strategies.

Second, the results were limited to two broad diagnostic categories, conduct disorder and depression, and to males. Subtypes of these diagnoses and other diagnoses were not evaluated. The extent to which rating scales can be used to make finer differentiations among given diagnostic groups needs to be explored.

Third, the research was conducted on an inpatient treatment service with patients who are likely to evince severe dysfunction. The children may represent more diffuse psychopathology (mixed syndromes) than seen in outpatient settings (see Lessing, Williams, & Gil, 1982). The diffuse psychopathology may have important implications for estimates of sensitivity and specificity. Children with more diffuse psychopathology may show a broad range of symptoms (and disorders) beyond a specific diagnostic entity. Specificity of the measures would be attenuated because children do not fall neatly into isolated diagnostic groups. Although the results may not be representative of outpatient clinic populations, the fact that parent rating

scales and symptom specific interviews and paper-and-pencil measures differentiated among diagnoses in the present population may be of special interest.

Finally, the study does not provide definitive information regarding the CBCL and BPC in relation to the DSM III diagnosis of psychiatric disorders. Both measures have undergone and probably will continue to undergo further development. For example, the BPC has been revised and expanded beyond the extremely well-researched 55-item version (Quay & Peterson, 1983). New scale items and derivative scales no doubt would generate different results. Similarly, the CBCL has been revised and recently expanded (Achenbach & Edelbrock, 1983) and includes versions beyond parent scales. The purpose of the present study was not to identify characteristics of the CBCL and BPC per se but rather to study the relationship of parent ratings and clinically derived diagnoses. The results suggest that ratings can readily differentiate among broad diagnostic groupings. Future research is needed to explore additional diagnostic entities, subtypes, and factors that enhance both sensitivity and specificity among the measures.

REFERENCES

- Achenbach, T. M. (1978). The Child Behavior Profile: I. Boys aged 6-11. *Journal of Consulting and Clinical Psychology, 46*, 478-488.
- Achenbach, T. M. (1980). DSM-III in light of empirical research on the classification of child psychopathology. *Journal of the American Academy of Child Psychiatry, 19*, 395-412.
- 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).
- Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington, Vermont: University Associates in Psychiatry.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, D.C.: Author.
- Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21*, 343-349.
- Chambers W., Puig-Antich, J., & Tabrizi, M. A. (1978). *The ongoing development of the Kiddie-SADS (Schedule for Affective Disorders for School-age Children)*. Paper presented at the American Academy of Child Psychiatry, San Diego.
- Connors, C. K. (1969). A teacher rating scale for use in drug studies with children. *American Journal of Psychiatry, 126*, 884-888.
- Herjanic, B., Herjanic, M., Brown, F., & Wheatt, T. (1975). Are children reliable reporters? *Journal of Abnormal Child Psychology, 3*, 41-48.
- Hodges, K., McKnew, D., Cytryn, L., Stern, L., & Kline, J. (1982). The Child Assessment Schedule (CAS) Diagnostic Interview: A report on reliability and validity. *Journal of the American Academy of Child Psychiatry, 21*, 468-473.
- Kazdin, A. E., Esveldt-Dawson, K., Unis, A. S., & Rancurello, M. D. (1983). Child and parent evaluations of depression and aggression in psychiatric inpatient children. *Journal of Abnormal Child Psychology, 11*, 401-413.

- Kazdin, A. E., French, N. H., Unis, A. S., & Esveldt-Dawson, K. (1983). Assessment of childhood depression: Correspondence of child and parent ratings. *Journal of the American Academy of Child Psychiatry, 22*, 157-164.
- Kovacs, M. (1981). Rating scales to assess depression in school aged children. *Acta Paedopsychiatrica, 46*, 305-315.
- Lessing, E. E., Williams, V., & Gil, E. (1982). A cluster-analytically derived typology: Feasible alternative to clinical diagnostic classification of children? *Journal of Abnormal Child Psychology, 10*, 451-482.
- Miller, L. C. (1972). School Behavior Checklist: An inventory of deviant behavior for elementary school children. *Journal of Consulting and Clinical Psychology, 38*, 138-144.
- 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.
- Quay, H. C. (1977). Measuring dimensions of deviant behavior: The Behavior Problem Checklist. *Journal of Abnormal Child Psychology, 5*, 277-287.
- Quay, H. C. (1979). Classification. In H. C. Quay & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (2nd ed., pp. 1-42). New York: Wiley.
- Quay, H. C., & Peterson, D. R. (1979). *Manual for the Behavior Problem Checklist*. Unpublished manuscript, University of Miami and Rutgers State University.
- Quay, H. D., & Peterson, D. R. (1983). *Interim manual for the Revised Behavior Problem Checklist*. Available from H. C. Quay, Box 248704, University of Miami, Coral Gables, Florida 33124.
- Touliatos, J., & Lindholm, B. W. (1981). Congruence of parents' and teachers' ratings of children's behavior problems. *Journal of Abnormal Child Psychology, 9*, 347-354.
- Vecchio, T. (1966). Predictive value of a single diagnostic test in unselected populations. *New England Journal of Medicine, 274*, 1171-1173.