

Development of a Structured Psychiatric Interview for Children: Agreement on Diagnosis Comparing Child and Parent Interviews

Wendy Reich

*University of Missouri in St. Louis and
Washington University School of Medicine*

Barbara Herjanic¹

*Washington University School of Medicine and
St. Louis Children's Hospital*

Zila Welner and P. R. Gandhi

Washington University School of Medicine

Similar structured diagnostic interviews about the child were given by different interviewers to a cohort of 307 mother-child pairs. A diagnosis was made by computer on each interview, using specified criteria. Diagnoses on mother-child interviews were compared using the kappa statistic. Kappas of .30 or higher were found for the diagnosis of antisocial personality, conduct disorder, enuresis, mixed behavior-neurotic disorder, and possible depression. Comparisons were made for sex and age. Possible depression and enuresis were diagnosed reliably at all age levels and for both sexes. The limitations of the interview and diagnostic system used are discussed. The findings support the need for further efforts to develop diagnostic research interviews for use with children and adolescents.

The purpose of this study was to compare diagnoses made independently from the responses of children and their mothers to a structured interview schedule. The specific aim was to test the usefulness of direct systematic interviewing of children as a means of collecting reliable information

Manuscript received in final form September 24, 1981.

¹Address all correspondence to Barbara Herjanic, Department of Psychiatry, 4940 Audubon, Avenue, St. Louis, Missouri 63110.

from which a diagnosis could be made. For such an interview to be useful in an epidemiologic study of a general population, for instance, it would have to distinguish disturbed children from normal, and designate the type of disturbance as accurately as possible. It would also be useful to know if there are some diagnoses that could be reliably made by interviewing the child or adolescent alone. If a style of interviewing could be shown to be reliable using one diagnostic system, there is no reason to assume that the same style could not be adapted to classification systems different from the one utilized in this study.

The system of diagnosis chosen was a modified form of the International Classification, described by Rutter, Shaffer, and Shepherd (1973), and the diagnostic system used by Rutter, Tizard, and Whitmore (1970) in the Isle of Wight studies.

Because the interview schedule asked a series of questions about functioning in various areas, such as relationships at home, at school, and with peers, it was necessary to determine how many positive symptoms would distinguish a disturbed child from a well child. This was done by interviewing 50 well-matched pairs of pediatric and psychiatric clinic children and their mothers, and determining at what level of positive symptoms in each area there was a significant difference between the two groups. The cutoff points so established were incorporated into the diagnostic scheme (Herjanic & Campbell, 1977).

The number of diagnoses that could be made was limited by the simplicity and length of the interview. The study covered the age range from 6 through 16 and aimed to keep the interview time between 1 and 1½ hours. The following diagnoses were made: normal variation, adaptation reaction, hyperkinetic syndrome, specific learning disorder, conduct disorder, neurotic disorder, behavior disorder, mixed behavior-neurotic disorder, enuresis, encopresis, and psychosis (unspecified). Some diagnoses, such as specific learning disorder, were recognized as "suspected" and not final, since educational and psychological tests would be required. In addition, diagnoses of antisocial personality, possible depression, and possible hysteria were made according to the Washington University criteria (Feighner, Robins, Guze, Woodruff, Winokur, & Munoz, 1972). These were "possible" diagnoses because the criteria were based on experience with adults. The important feature is that the diagnostic system used was applied consistently to all 307 cases.

METHOD

The interview procedure, subjects, the interview itself, and the use of the kappa statistic have been described in a previous paper by Herjanic

and Reich (1982). Of the 307 subjects, ranging in age from 6 through 16, 50 were children who attended a pediatric clinic. The remaining 257 were children who were referred for psychiatric evaluation or consultation. Diagnoses were made by computer using the systems of diagnosis described above. Adaptation reaction, hyperkinetic syndrome, specific learning disorder, and psychosis were omitted from this analysis because of the impossibility of defining the criteria sufficiently using the information obtained from either the child or mother interview alone.

This study was not concerned with determining the primary diagnosis of a particular child, but in comparing agreement between mother and child on the presence or absence of symptoms that might lead to one or more diagnoses for the child. There is considerable overlap, therefore, in that one child may be included under the diagnosis of both conduct and behavior disorder and neurotic disorder, thereby making the child eligible for the mixed behavior-neurotic disorder category. The child could also have enuresis or encopresis or sufficient symptoms to be diagnosed as possible depression.

Some diagnoses could be made on the basis of either the mother or child interview alone, but not both, since some symptoms, as described in Herjanic and Reich (1982), were reported significantly more often by either mother or child. This asymmetrical reporting of symptoms produced the "diagnoses with asymmetrical responses," shown in Tables I, II, and III.

The kappa statistic was used to compare diagnoses made from pairs of interviews. An important problem was to assess the significance of the kappas. This was especially crucial because, with a few exceptions, they were not high. Helzer, Clayton, Pambakian, Reich, Woodruff, and Reveley (1977) tested interrater reliability on adult psychiatric interviews similar to ours and found diagnostic agreement with kappas of .55 and higher. That study, however, compared diagnoses made from interviews of the same person obtained by different interviewers. The present study sought to measure agreement on diagnoses for *one* child using *two* sources of interview material, one being the child and one the mother, each seen by two separate interviewers. The rate of agreement might be expected to drop when both interviewer and respondent were different.

Kappas obtained in this study were compared to those from studies that measured the agreement for the diagnosis of affective disorder made from interviews with adults, and diagnoses on those same adults made from separate, matching interviews of first-degree relatives conducted by different interviewers (Rimmer & Chambers, 1969; Winokur, Clayton, & Reich, 1969; Andreasen, Endicott, Spitzer, & Winokur, 1977). The kappas in these studies ranged from .30 to .51. Since the kappas in this study (see below) fell within the same range, we concluded that our child

and adolescent subjects performed at least as well as adults under similar circumstances. Therefore, the kappa of .30 was taken as a baseline. Diagnoses with kappas at that level and above were assumed to represent the reliability obtained to date in this type of interview situation.

RESULTS

The results are presented in Tables I to III. In Table I, diagnoses made on the basis of the child's interview, but not confirmed by the mother's interview, are under the heading Child's Interview Only, and diagnoses made on the basis of mother's interview, but not confirmed by the child's interview, are under the heading, Mother's Interview Only. Diagnoses, such as possible hysteria, with less than 10 subjects were omitted, since the kappa statistic is not useful with such small frequencies.

Kappas of .30 or higher were obtained between mothers and children in five diagnostic categories: antisocial personality, conduct disorder, enuresis, mixed behavior-neurotic disorder, and possible depression (Table I). Of these, antisocial personality and enuresis had the highest kappas (.58 and .54, respectively), while conduct disorder, mixed behavior-

Table I. Diagnoses Based on 307 Child-Mother Pairs

	Diagnosis positive				Kappa
	Both interviews	Mother's interview only	Child's interview only	Either or both	
Diagnoses with kappa of .3 or higher					
Antisocial personality	5	5	2	12	.58
Conduct disorder	11	14	16	41	.37
Enuresis	40	30	15	85	.54
Mixed behavior-neurotic disorder	11	14	16	41	.37
Possible depression	70	45	47	162	.36
Diagnoses with asymmetrical responses					
Behavior disorder	34	67 ^b	32	133	.19
Neurotic disorder	25	34	54 ^a	113	.18
Diagnoses with kappas below .3 and no asymmetry					
Encopresis	7	28	18	53	.15
Normal variation ^c	8	12	12	32	.00

^a $p = .05$.

^b $p = .01$.

^cPediatric group only ($N = 50$).

neurotic disorder, and possible depression had kappas of .37, .37, and .36, respectively. In none of the diagnoses did the number of pairs with full agreement exceed the total number of pairs in which there was disagreement. In only two categories, enuresis and possible depression, did the number of pairs that agreed on diagnosis exceed the number of Mother Only and Child Only diagnoses.

The diagnosis of possible depression was very frequent (162), involving 53% of the children (52% of the boys and 55% of the girls). As mentioned previously, the Feighner criteria were used to make the diagnosis of depression. However, the interview was not designed precisely to fulfill all of the adult criteria for depression as described by Feighner et al. (1972). To be diagnosed as depressed, a child had to have a "yes" response to one of two questions concerning dysphoric mood, corresponding to Feighner A criteria. In addition, the child had to have five "yes" responses corresponding to the areas included in Feighner B criteria. These areas were appetite change, sleep disturbance, tiredness (fewer than 10 responses), loss of interest, difficulty concentrating, and thoughts of death or suicide. Two Feighner criteria, agitation or retardation, and feelings of self-reproach, were omitted from the interview.

Many of the children and mothers interviewed reported many of these symptoms. What makes a definite diagnosis impossible is the lack of time specificity. The adult criteria require dysphoric mood, plus at least five symptoms of a month's duration or longer. Our interviews contained no question describing duration or severity of symptoms, or possible relationship to life events. These are deficits that are being remedied in new interview formats now being developed. Since the present study clearly suggests that children report the usual symptoms of depressive illness when directly asked, it is hoped that more precise handling of questions relating to duration and severity will enable researchers to separate the children who have transient symptoms from those with a definite diagnosis.

Four diagnoses had kappas of less than .30. Two of these, however, had asymmetrical responses. In reporting the symptoms of behavior disorder, mothers responded positively significantly more often than children, while in symptoms of neurotic disorder, the children's positive responses were significantly more common (see Herjanic & Reich, 1982). That children asymmetrically report neurotic symptoms is important, as it suggests that there is some information relevant to diagnosis that can be obtained more effectively from children themselves.

Two remaining diagnoses, encopresis and normal variation, had kappas below .3 for the group as a whole. The category normal variation was used for children who did not fall into any of the other diagnostic groups. This means that these children could be completely

healthy with only a few mild symptoms, or they could be considered "undiagnosed," in that they showed too many symptoms not to be considered disturbed in some way, but insufficient to fulfill the diagnostic criteria of any specific disorder. In order to sort out the true normals from the undiagnosed, we assumed that the majority of children with only a few mild symptoms would be found in the pediatric group, and there would be few, if any, undiagnosed children in this group. We therefore compared the mother-child responses in the pediatric group only (Table I, Normal variation) and found a kappa .00, i.e., the positive agreement was no better than chance. We also found the same number of diagnoses of normal variation (12) for mothers only and for children only. On the interviews of only eight mother-child pairs was there agreement that the child was not suffering from a psychiatric disorder as defined here. Since few children would be likely to exhibit no symptoms at all, the disagreement must lie in the number of symptoms reported by each pair. As mothers and children showed reasonable agreement in five of the categories shown in Table I, and since this is a pediatric group that should exhibit fewer severe psychiatric problems than children seen by a psychiatrist, it seems likely that mothers and children were disagreeing about the severity of a problem that was in itself mild.

There is some evidence to substantiate this line of reasoning. As each of the 50 pediatric children and their mothers were interviewed, a diagnostic assessment was done independently on each interview by one of two child psychiatrists. Then a consensus diagnosis between the two psychiatrists was made on each child. When there were differences between the parent and the child interviews, more weight was given to the parent responses in the area of behavior, and more weight was given to the child response in the area of affect. By this clinical method 30 of the 50 pediatric children were diagnosed as having "no disorder," 8 had a definite or possible behavior disorder, and 6 had a definite or possible neurotic disorder. Of the 6 remaining, 3 had a learning disorder and 3 mental retardation. In this clinical evaluation, only the primary diagnosis was listed. It is interesting to note that out of the same 50 children, the computer diagnosed a total of 32 as having "normal variation" by either parent or child interview or both. With the method used here there was no way to weigh the relative importance of the differences in reported symptoms. Children with few mild symptoms falling within the normal range have to be separated from those with more serious, long-lasting, and more numerous symptoms, but insufficient to reach designated diagnostic criteria. In future studies allowance has to be made for an "undiagnosed" group, as distinguished from a "no-disorder" group.

The lack of agreement on encopresis occurred even though a concrete question was asked of the child, such as "Do you remember ever having a bowel movement (or "having pooped" or "having messed") in your pants or bed?" which is the diagnostic symptom for encopresis in the absence of a medical illness. Mothers reported the symptom of soiling more than children, but the difference was statistically significant only in the 6-to-9-year age range.

The lack of agreement on the diagnosis of encopresis illustrates a major problem area in establishing the reliability of a diagnosis that depends basically upon the answer to a single question, especially a question that is associated with feelings of shame and can easily be denied. The kappas on encopresis were very low for all age groups. However, the data showed that mothers had almost as great a tendency to deny or to forget the symptom as their children, especially when the child was 10 to 16 years of age.

Differences by Sex

Boys in the study outnumbered girls by almost two to one. Interviews of both boys and girls showed sufficiently high rates of agreement with their mothers to result in kappas of .30 or higher in enuresis and possible depression. Thirty percent of the boys and 24% of the girls had the symptom of enuresis at some time since the age of 5. Only girls reported significantly more symptoms leading to the diagnosis of possible depression than their mothers (see Table II). This tendency of girls to report many symptoms in this area at a very early age is interesting, in view of the well-known preponderance of adult females with affective disorder.

Girls are primarily responsible for the asymmetrical reporting of neurotic disorder. However, the fact that boys do report neurotic symptoms, even if they do not agree with their mothers, plus the fact that possible depression, the diagnosis of which depends on the reporting of many subjective feelings, has essentially the same kappa (.36 and .37) for girls and boys indicates that boys are reporting subjective feelings, even if they do not do so as often as the girls.

Girls' agreement with their mothers on the symptom of encopresis reached a kappa of .33, and involved 12% of the girls. Twenty percent of the boys, on the other hand, were reported by one or both interviews to have the symptom of soiling, but the agreement was very poor, the kappa being less than .01. Boys, therefore, are responsible for the low kappa attained on this diagnosis for the group as a whole (see Table I).

Table II. Differences in Mother-Child Interview – Diagnosis by Sex

	Mother-daughter (N = 110)			Mother-son (N = 197)				
	Both interviews	Mother only	Child only	Kappa	Both interviews	Mother only	Child only	Kappa
Diagnosis with kappas of .3 or higher								
Antisocial personality					4	5	2	.51
Conduct disorder					9	12	15	.32
Enuresis	13	10	3	.59	27	20	12	.52
Encopresis	3	7	3	.33				
Mixed behavior-neurotic					9	12	15	.32
Possible depression	27	11	23	.36	43	34	24	.37
Diagnoses with asymmetrical responses								
Behavior disorder	5	21 ^b	6	.15	29	46 ^a	26	.18
Neurotic disorder	9	9	28 ^b	.14				
Possible depression	27	11	23 ^a	.36				
Diagnoses with kappas below .3 and no asymmetries								
Encopresis					4	21	15	.08
Neurotic disorder					16	25	26	.22

^ap = < .05.^bp = < .01.

Table III. Differences in Mother-Child Interview — Diagnosis by Age of Child

	12-16 years (N = 114)			10-11 years (N = 84)			6-9 years (N = 109)					
	Both interviews	Mother only	Child only	Kappa	Both interviews	Mother only	Child only	Kappa	Both interviews	Mother only	Child only	Kappa
Diagnoses with kappas of .3 or higher												
Antisocial personality	5	5	2	.56								
Behavior disorder	15	17	13	.32								
Conduct disorder	8	8	11	.36								
Enuresis	8	5	6	.54	16	4	4	.73	16	21	5	.40
Mixed behavior-neurotic	8	8	11	.36								
Possible depression	35	18	21	.31	20	11	13	.39	15	16	13	.33
Diagnoses with asymmetrical responses												
Behavior disorder					10	20 ^a	7	.22	9	30 ^b	12	.09
Encopresis									3	14 ^b	7	.12
Diagnoses with kappas below .3 and no asymmetry												
Conduct disorder	2	6	5	.21	2	8	6	.13	2	5	3	.29
Encopresis												
Mixed behavior-neurotic	13	11	21	.27	10	9	16	.25	2	5	3	.29
Neurotic disorder									2	14	17	-.05

^ap = < .05.

^bp = < .01.

Boys, on the other hand, reported sufficient numbers of symptoms and showed good enough agreement with their mothers in the area of antisocial behavior to account for the relatively high kappas in diagnoses of antisocial personality, conduct disorder, and mixed behavior-neurotic disorder shown for the group as a whole in Table I.

Differences by Age

As might be expected, older children, aged 12 to 16 years, agreed with their mothers more than did the 10- to 11-year-olds and the 6- to 9-year-olds (Table III). In addition to the five diagnoses found in the total population with kappas of .30 or greater, the older children and their mothers agreed about enough symptoms to also diagnose behavior disorder. There were no significant asymmetries in the older group. In this age group only, slightly more diagnoses could be made from the children's interview (187) as compared with the mother's interviews (174).

The 10-to-11-year-old group agreed with their mothers often enough to make the diagnoses of enuresis and possible depression. The kappa for enuresis was higher than any of the others, .73, reflecting the importance of persistent enuresis in this age group. There were no diagnoses of antisocial personality, and fewer than 10 with conduct disorder and mixed behavior-neurotic disorder among the middle children. Like the total population, a sufficient number of symptoms for the diagnosis of behavior disorder was reported more often by the parents. The total number of diagnoses made from the parent and child interviews was about the same: parents, 114 and children, 110.

The youngest group of children, the 6- to 9-year-olds, also agreed with their parents on enough symptoms to warrant two fairly reliable diagnoses, enuresis and possible depression. It is interesting that a reasonable agreement could be obtained for a diagnosis of possible depression in children as young as the 6-to-9-year group.

Sufficient symptoms for two diagnoses were reported asymmetrically in the 6-to-9-year age group. These were behavior disorder and encopresis, with the parents reporting significantly more symptoms than the child in both cases. A total of 154 diagnoses could be made from the parent interviews, as compared to 109 from the children's interviews. The difference is significant at $p = < .01$.

DISCUSSION

This study shows more disagreement than agreement on diagnoses based solely on separate, but similar, interviews given to children and

their mothers. Using the kappa statistic, antisocial personality and enuresis were the most reliably diagnosed by both mother and child interviews. Boys alone accounted for the former, but both boys and girls agreed well with their mothers about bed wetting. Girls' interviews met criteria for neurotic disorder and possible depression significantly more frequently than did their mothers'. In contrast, mothers' interviews significantly more often met criteria for behavior disorder than did their sons'.

Older children (ages 12-16) were on the whole in much more agreement with their mothers than younger. However, enuresis was reliably reported from age 6 through 16, and possible depression was diagnosed with almost the same degree of mother-child agreement throughout the whole age range. The diagnosis of behavior disorder was made on mothers' interviews significantly more frequently for both boys and girls under age 12.

Further research needs to be done on the efficacy of child interviewing before ruling out or limiting the use of interviews at any specific age. When parents and children disagree, it is important to note that we cannot be sure who is the more accurate, if in fact either of them is. Clinical impressions indicate that preadolescent boys, in particular, seem to deny behavior problems about which their parents are very concerned. Further work needs to be done to see to what extent denial or different perceptions of "problem" behavior may account for differences in responses on the parent-child interview, and what other, as yet unexplained, factors may be involved in the disagreements. Many of the symptoms of behavior disorder refer to behavior at school, which could be verified by asking the teacher, thereby adding a dimension by which to check the validity of parent-child reporting.

One of the major problems with this study is the broadness of the diagnostic categories and the incompleteness of the data obtained in terms of timing, duration, and severity of groups of symptoms. We have been able to show that there is considerable agreement between mothers and children in response to similar questions about symptoms, suggesting that reliable diagnoses could be made on either mother or child interview, if the questions were specifically related to more precise criteria.

DSM III (American Psychiatric Association, 1980) offers just such an opportunity. Early field trials using case histories have shown that interrater percents of agreement on diagnosis were at least as good as that obtained with DSM II (Mattison, Cantwell, Russell, & Will, 1979). Comparisons were made of diagnostic agreement between paired clinicians independently evaluating children by means of the same sources of information. Overall rates of agreement (kappa) were .69 and .63 for two versions of DSM III. In general, there was a higher rate of agreement on conduct type disorders than on anxiety disorders. Agreement was con-

sidered to be present if the two diagnosticians placed the child in the same broad category of diagnosis. In this way the relative nonspecificity was comparable to the broad diagnostic grouping utilized here.

In recent years much progress has been made in establishing the reliability of diagnoses of adults, due to carefully constructed interview schedules, lists of diagnostic criteria, and rigorous training of raters. Although research on diagnosis of children is at an earlier stage than with adults, it seems reasonable to expect that we should be able to achieve predictable, reliable results. A structured interview for children and adolescents and their parents, based on the DSM III classification and specified diagnostic criteria, is being developed. The reliability of such an interview should exceed that of its predecessor described above, and should definitely provide a sound base for testing the reliability of parent-child reporting using the new classification system.

REFERENCES

- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, D.C.: APA, 1980.
- Andreasen, N. C., Endicott, J., Spitzer, R. L., & Winokur, G. The family history method using diagnostic criteria. Reliability and validity. *Archives of General Psychiatry*, 1977, 34, 1229-1235.
- Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Jr., Winokur, G., & Munoz, R. Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 1972, 26, 57-63.
- Helzer, J. E., Clayton, P. J., Pambakian, R., Reich, T., Woodruff, R. A., Jr., & Reveley, M. A. Reliability of psychiatric diagnosis. II. The test/retest reliability of diagnostic classification. *Archives of General Psychiatry*, 1977, 34, 136-141.
- Herjanic, B., & Campbell, W. Differentiating psychiatrically disturbed children on the basis of a structured interview. *Journal of Abnormal Child Psychology*, 1977, 5, 127-134.
- Herjanic, B., & Reich, W. Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. *Journal of Abnormal Child Psychology*, 1982, 10, 307-324.
- Mattison, R., Cantwell, D. P., Russell, A. T., & Will, L. A comparison of DSM-II and DSM-III in the diagnosis of childhood psychiatric disorders. II. Interrater agreement. *Archives of General Psychiatry*, 1979, 36, 1217-1222.
- Rimmer, J., & Chambers, D. F. Alcoholism: Methodological considerations in the study of family illness. *American Journal of Orthopsychiatry*, 1969, 39, 760-769.
- Rutter, M., Shaffer, D., & Shepherd, M. An evaluation of the proposal for a multi-axial classification of child psychiatric disorders. *Psychological Medicine*, 1973, 3, 244-250.
- Rutter, M., Tizard, J., & Whitmore, K. *Education, health and behavior*. New York: Wiley, 1970.
- Winokur, G., Clayton, P. J., & Reich, T. *Manic-depressive illness*. St. Louis: C. V. Mosby, 1969.