Psychiatry and Clinical Neurosciences (1999), **53**, 57–61

Regular Article

Reliability of childhood mental disorder: Diagnoses by Japanese psychologists

CHIEKO HASUI, MD,[1] TOMOKO SUGIURA, MD,[1] ERIKO TANAKA, MD,[1]
SHINJI SAKAMOTO, MD,[1] MASUMI SUGAWARA, MD,[1]
TOSHINORI KITAMURA, MD[1] AND YUKO AOKI, MD[2]
[1]*Department of Sociocultural Environmental Research, National Institute of Mental Health, Ichikawa, Chiba and* [2]*Tsurumidai Hospital, Beppu, Oita, Japan*

Abstract The recent entry of psychologists into psychiatric practices and schools in Japan calls for diagnostic skills, because psychiatrists are less available in such settings. We examined the reliability of the diagnoses of 10 DSM-IV childhood mental disorders by 11 Japanese psychologists, using 20 case vignettes. Most categories had good reliability (κ), except for attention deficit hyperactivity disorder. Japanese clinical psychologists may be able to use DSM-IV reliably as a tool for diagnosis of childhood psychiatric disorders, if sufficient training is provided.

Key words childhood mental disorder, diagnosis, psychologist, reliability.

INTRODUCTION

In Japan, diagnosis of mental illness has long been a function only for psychiatrists. Non-medical psychiatric professionals such as nurses, clinical psychologists, psychiatric caseworkers, and occupational therapists have never been involved in diagnosis. This is partly due to the law, which exclusively empowers qualified medical graduates to diagnose and prescribe medication for patients in all branches of medicine, and also partly to the fact that few paramedical staff have been employed in psychiatric institutions. Recently, however, more and more psychology graduates have entered clinical fields. Thus, the Japanese Certification Board for Clinical Psychologists Inc. was established in 1990 to license officially a clinical psychology qualification. More clinical psychologists, be they licensed or not, are being seen in different parts of the mental healthcare system: mental hospitals, psychiatric clinics, child guidance clinics, schools, and work places. This is in concordance with the fact that the number of psychiatrists are fewer than needed. Prompted by a series of 'scandals' in schools (e.g. bullying, suicides, school refusal, and learning difficulties), the Japanese government has dispatched psychologists to both elementary schools and junior high schools to meet the needs of support-seeking children. Furthermore, most schools are reluctant to invite psychiatrists as a consultant.

These recent changes in the roles of clinical psychologists in Japan require them to have more diagnostic skills, because psychiatrists are rarely available in school settings. School psychologists are now taking the role of the primary diagnostician for childhood mental disorders. Yet both the graduate and postgraduate teaching and training of clinical psychologists have traditionally been focused on the understanding of mental mechanisms, such as defense styles and psychotherapeutic skills. For example, the DSM-III/DSM-III-R[1,2] is rarely taught in psychology courses.

The recent introduction of behavior therapy and cognitive therapy to Japan has not been influential in shifting the emphasis of clinical psychology education to the diagnostic aspect of psychopathology. Yet Japanese psychologists' diagnostic skills have rarely been questioned. Fujihara *et al*. examined the interrater reliability of diagnoses of adult mental disorders made by psychologists as well as psychiatrists, using case vignettes.[3] Psychologists achieved good interrater reliability on major depressive

Correspondence address: Chieko Hasui, MD, Department of Sociocultural Environmental Research, National Institute of Mental Health, 1-7-3 Konodai, Ichikawa, Chiba 272-0827, Japan.

Received 13 May 1998; revised 31 August 1998; accepted 7 September 1998.

disorder and manic disorder, as defined by the Research Diagnostic Criteria (RDC).[4] Kato *et al.* reported that Japanese postgraduate psychologists obtained high interrater agreement on the diagnosis of mood disorders.[5] These studies showed that Japanese psychologists could, to some extent, diagnose psychopathology reliably and accurately. However, the cases described in these studies were limited to adult subjects. Some psychiatric conditions such as conduct disorder and attention deficit disorder are seen exclusively among children, while psychiatric disorders seen among adults, such as depression, may present with different clinical pictures among children. Therefore there is a need to examine Japanese clinical psychologists' skills in diagnosing psychopathology among children.

We report here the reliability and accuracy of diagnoses of childhood mental disorders by Japanese psychologists. To our knowledge, this is the first study of its kind. It is of note here that we use the term 'reliability' rather than 'validity' throughout this paper. One may argue that the agreement of novice diagnosticians with a specialist in terms of psychiatric diagnosis is a reflection of validity. However, the diagnosis made by specialists also needs validation. Therefore we think that the agreement between the two should be interpreted as accuracy of novice diagnosticians.

SUBJECTS AND METHODS

A total of 11 psychologists were recruited for a larger study on the epidemiology of family mental health and mental illness. They were all women, with an age-range from 22 to 36 years. Of these psychologists, eight had already had clinical experience. However, none had been officially trained in the use of the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV).[6] For the present study, they were given a series of lectures on DSM-IV and on the use of a structured interview to diagnose both adult and child mental disorders. The training consisted of 12 one-day-a-week courses in which half a day was spent on a classroom lecture on the psychiatric diagnosis of DSM-IV, and another half a day was spent for the role-play of a structured diagnostic interview.

The participant psychologists independently diagnosed each of 20 case vignettes. The vignettes were taken from the *DSM-III-R Case Book*, which had been translated into Japanese.[7] For the purpose of the present study, the vignettes were slightly modified to allow DSM-IV diagnoses. This is because some DSM-IV categories have additional items necessary for definite diagnoses. All the selected cases were diagnosed by one of us (TK) and his diagnoses were perfectly in concordance with those given in the *Case Book*. The vignettes chosen covered a broad range of childhood mental disorders: two cases of attention deficit hyperactivity disorder (ADHD), four cases of conduct disorder, three cases of separation disorder, three cases of generalized anxiety disorder, one case of oppositional disorder, four cases of enuresis, two cases of encopresis, three cases of major depression, one case of obsessive–compulsive disorder, and three cases of disorder of infancy childhood or adolescence not otherwise specified. The total number exceeds that of the case vignettes because of the multiple diagnostic policy used in the DSM-IV.

The reliability for each diagnostic category was calculated using the κ coefficient.[8] Although there are many methods to calculate interrater agreement, the κ coefficient is the best known coefficient. Diagnostic categories with a base rate of less than 10% were excluded from further analyses. After each participant psychologist had made the DSM-IV diagnoses independently, they were compared with the diagnoses made by one of us (TK), who had at least 25 years' experience in clinical and research work using DSM, as well as working as an editor of several international journals. These were thus used as a specialist's judgment, against which the psychologists' diagnoses were validated.

After ranking all the participants by the mean κ coefficients of the concordance with the specialist's diagnoses, they were divided into three groups: high ($n=4$); middle ($n=4$); and low ($n=3$), representing the top, middle, and bottom thirds of the participants, respectively. We then calculated the interrater reliability of each diagnostic category for the three groups separately. It was hypothesized that if the participants showing discordance with the specialist's diagnosis had a bias towards the same direction, they would still show a high interrater reliability among themselves. We also examined the reliability of the judgement of the diagnostic items of categories with low reliability, to explore the causes of low reliability.

RESULTS

Of the DSM-IV diagnostic categories, seven had base rates of 10% or more (Table 1). Of these seven categories, conduct disorder, generalized anxiety disorder, major depression, separation anxiety disorder, encopresis, and enuresis obtained mean κ values of more than 0.6 between each rater and the specialist, while the κ coefficient of ADHD failed to exceed 0.6.

Table 1. Base rate and mean kappa between each of raters and the specialist

DSM-IV diagnosis	Base rate (%)	κ coefficient of raters Mean (range)
Conduct disorder	50	0.74 (0.34–1.00)
Major depression	15	0.91 (0.69–1.00)
Generalized anxiety disorder	15	0.64 (0.00–0.82)
Separation anxiety disorder	10	0.66 (0.44–1.00)
Enuresis	10	0.60 (0.00–1.00)
Encopresis	10	0.81 (0.40–1.00)
Attention deficit hyperactivity disorder	10	0.54 (–0.07–1.00)

We found no significant correlations between the psychologist's κ coefficients and the length of their clinical experiences.

When the concordance of the diagnoses of the 20 cases between each rater and the specialist was examined among the high, middle and low groups separately, different pictures emerged (Table 2). As expected, all the diagnostic categories showed good concordance (expressed by κ) among the high group. Among the middle group, the concordance of diagnosis of major depression was good, while those of the remaining categories were fair. Among the low group, it was good for major depression; fair for conduct disorder, separation anxiety disorder, and enuresis; and poor for generalized anxiety disorder, encopresis and ADHD. Thus, major depression retained a high κ coefficient, regardless of the group membership.

Then, we calculated the means of the κ coefficients of each diagnostic category between raters (excluding the specialist) in each group. As expected, the concordance of diagnosis between the psychologists was generally high among the high group (whose diagnosis was in better concordance with the specialist). However, it was worse in the middle and low groups, with a few exceptions. In most categories and groups, the concordance between the raters was lower than that between the raters and the specialist. Exceptions were encopresis in the high and middle groups and separation anxiety disorder in the middle group.

Because the concordance of diagnosis between the psychologists and the specialist failed to exceed 0.6 for ADHD, we examined the reliability of each diagnostic item of ADHD. Of the 18 diagnostic items of ADHD, nine items are inattention and the others are hyperactivity–impulsivity. Of diagnostic criteria for inattention, the criteria of 'often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities', and 'often has difficulty organizing tasks and activities' showed a low κ (0.498, 0.280, respectively). In the hyperactivity–impulsivity domain, the criterion of 'is often "on the go" or often acts as if driven by a motor' showed a low κ (0.278).

DISCUSSION

Our data suggest that DSM-novice psychologists may handle the DSM-IV with good reliability. Diagnoses of the DSM-IV categories other than ADHD showed high agreement between each of the raters and the specialist. High agreement did not associate with the duration of their clinical experiences. There appeared no evidence that the participant psychologists misunderstood the DSM-IV diagnostic rules with a bias towards the same direction. However, this issue needs further consideration because the number of the psychologists was small, and because we divided the sample into three rather than taking other subdivision methods.

However, ADHD failed to achieve satisfactory reliability. In addition to its low base rate, this category may be difficult for novice diagnosticians to understand and use. Of the diagnoses made by the 11 psychologists for the two cases of ADHD (the specialist's diagnosis), 82% (18/22) were correctly diagnosed. Thus, cases of ADHD were recognized as such by the participants. However, seven other cases for which the specialist's diagnosis was not ADHD were diagnosed as ADHD by at least one participant. These seven cases included four of conduct disorder and one of oppositional disorder. It may be that the participant psychologists found it difficult to differentiate ADHD from the group of 'attention deficit and disruptive behavior disorder' of the DSM-IV, particularly conduct disorder. Symptomatic overlap of ADHD and disruptive behavior disorders was indicated by Faraone *et al.*[9] Some investigators have reported differences in the concept of ADHD between different countries.[10–13] For example, in England, children with hyperactivity are more likely to be diagnosed as having conduct disorder, while in North America, such children are diagnosed as having hyperactivity or attention deficit disorders. Thus, UK-trained diagnosticians prefer the category of conduct disorder, while US-trained prefer ADHD; there are large discrepancies in diagnostic practice between the two countries. Examining the classification of child psychopathology in DSM-III, Achenbach stated that 'a child could have most of the problems of the hyperactive syndrome without specifically showing

Table 2. Interrater reliability and the agreement between each of the raters and the specialist for seven diagnostic criteria

DSM-IV diagnosis	Base rate (%)	Interrater reliability between the raters Mean (range)	Agreement with the specialist Mean (range)
High (top third of the range of κ)			
Conduct disorder	50	0.79 (0.69–0.85)	0.91 (0.82–1.00)
Major depression	15	0.68 (0.50–0.77)	0.87 (0.69–1.00)
Generalized anxiety disorder	15	0.66 (0.61–0.77)	0.78 (0.77–0.82)
Separation anxiety disorder	10	0.73 (0.60–0.82)	0.73 (0.61–0.77)
Encopresis	10	0.88 (0.64–1.00)	0.73 (0.64–1.00)
Enuresis	10	0.88 (0.64–1.00)	0.91 (0.64–1.00)
Attention deficit hyperactivity disorder	10	0.44 (0.34–0.64)	0.71 (0.611–1.00)
Middle (middle third of the range of κ)			
Conduct disorder	50	0.40 (0.40–0.71)	0.65 (0.68–0.82)
Major depression	15	0.94 (0.82–1.00)	0.95 (0.82–1.00)
Generalized anxiety disorder	15	0.48 (0.23–0.61)	0.63 (0.31–0.82)
Separation anxiety disorder	10	0.66 (0.50–0.85)	0.61 (0.20–0.61)
Encopresis	10	1.00 (1.00)	0.64 (0.64)
Enuresis	10	0.59 (0.44–0.77)	0.79 (0.40–1.00)
Attention deficit hyperactivity disorder	10	0.28 (0.18–0.45)	0.61 (0.20–1.00)
Low (low third of the range of κ)			
Conduct disorder	50	0.40 (0.34–0.45)	0.61 (0.34–0.82)
Major depression	15	0.89 (0.77–1.00)	0.92 (0.77–1.00)
Generalized anxiety disorder	15	0.39 (0.00–0.77)	0.45 (0.00–0.77)
Separation anxiety disorder	10	0.61 (0.61)	0.68 (0.44–1.00)
Encopresis	10	0.29 (–0.07–0.64)	0.36 (0.00–0.64)
Enuresis	10	0.29 (–0.07–0.64)	0.69 (0.44–1.00)
Attention deficit hyperactivity disorder	10	0.26 (0.17–0.34)	0.23 (–0.07–0.44)

hyperactive behavior'.[14] The diagnostic category of ADHD may be less clear than others, and therefore difficult to diagnose. Further clarification of the concept of the attention deficit and disruptive behavior disorder may be needed for the better understanding of Japanese psychologists.

Of another interest was the finding that major depression maintained a high agreement with the specialist's diagnosis across the three groups. Thus the κ coefficients of major depression were 0.87, 0.95, and 0.92 for the top, middle, and bottom groups, respectively. This is in contrast with other diagnosis such as ADHD. This may mean that graduate and postgraduate education in psychology has a strong emphasis on emotional disorders, and less attention is paid to behavior disorders. If this is the case, our data may be helpful for rethinking Japanese psychology education.

There are a few drawbacks to the present study. First, the participants applied the diagnostic rules of DSM-IV for case descriptions, which contain symptoms, and behavior of the category. Therefore, this did not examine the participants' capacity to elicit symptoms. Second, since all the cases were translated from those in the US, one should be cautious about generalizing it into Japanese clinical settings. Third, it is unsatisfactory to calculate the reliability of psychologists' diagnoses on the basis of 20 cases. We have no knowledge of their diagnostic capacity for categories other than those examined in this study. Our study should therefore be regarded as a preliminary one.

Furthermore, the number of the participant psychologists was small, some of them having little clinical experiences. It can hardly be claimed that they are representative of Japanese clinical psychologists. Reliability studies are necessary in a large (e.g. more than 30 people) population of experienced (e.g. clinical involvement of 3 years or longer) psychologists. Such studies may give us a succinct perspective of clinical psychologists' capacity to take diagnostic responsibility in clinical settings. However, the present study suggests the possibility that even novice psychologists can understand diagnostic rules proposed by the DSM-IV. Therefore it encourages further use of operationalized diagnostic criteria by psychologists in Japan.

Despite these limitations, the present study showed, although tentatively, that Japanese psychologists

could use the DSM-IV diagnostic rules for childhood mental disorders. Considering the ever-increasing demands on school counsellors and school psychologists in Japan, further work is urgently needed to explore the diagnostic capacity of psychologists for that setting.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edn. American Psychiatric Association, Washington DC, 1980.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edn (revised). American Psychiatric Association, Washington DC, 1987.
3. Fujihara S, Kitamura T, Ikuta N *et al*. A comparison of diagnostic reliability of clinical psychologists and psychiatrists by a case vignette design. *Jpn J. Clin. Psychol*. 1988; **17**: 1081–1087 (in Japanese).
4. Spitzer RL, Endicott J, Robins E. Research Diagnostic Criteria (RDC) for a Selected Group of Functional Disorders. Biometric Research, New York State Psychiatric Institute, New York, 1977.
5. Kato C, Watanabe S, Nagata T, Kitamura T. Reliability of psychiatric diagnosis: Application of operationalized diagnostic criteria. *Jpn J. Educ. Psychol.* 1990; **38**: 413–417 (in Japanese).
6. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychiatric Association, Washington DC, 1994.
7. Takahashi S. *DSM-III-R Case Book. A Learning Companion to the Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised). Igaku Shoin, Japan, 1990 (in Japanese).
8. Cohen J. A coefficient of agreement for nominal scales. *Educ. Psychol. Measure.* 1960; **20**: 37–46.
9. Faraone SV, Biederman J, Keenan K, Tsuang MT. Separation of DSM-III attention deficit disorder and conduct disorder: Evidence from a family-genetic study of American child psychiatric patients. *Psychol. Med.* 1991; **21**: 109–121.
10. Prendergast M, Taylor E, Rapoport JL *et al.* The diagnosis of childhood hyperactivity: a U.S.–U.K. cross-national study of DSM-III and ICD-9. *J. Child Psychol. Psychiatry* 1988; **29**: 289–300.
11. Taylor E, Sandberg S. Hyperactive behavior in English schoolchildren: A questionnaire survey. *J. Abnorm. Child Psychol.* 1983; **12**: 143–156.
12. Thorley G. Hyperkinetic syndrome of childhood: Clinical characteristics. *Br. J. Psychiatry* 1984; **144**: 16–24.
13. Werry JS, Methven RJ, Fitzpatrick J, Dixon H. The interrater reliability of DSM III in children. *J. Abnorm. Child Psychol.* 1983; **11**: 341–354.
14. Achenbach TM. DSM-III in light of empirical research on the classification of child psychopathology. *J. Am. Acad. Child Adolesc. Psychiatry* 1980; **19**: 395–412.