Soc Psychiatry Psychiatr Epidemiol (1997) 32: 339–343

ORIGINAL PAPER

H. Costello · S. Moss · H. Prosser · C. Hatton

Reliability of the ICD 10 version of the Psychiatric Assessment Schedule for Adults with Developmental Disability (PAS-ADD)

Accepted: 12 September 1996

Abstract The Psychiatric Assessment Schedule for Adults with Developmental Disability (PAS-ADD) is a semi-structured clinical interview designed for use with respondents who have learning disability. The first version was based on the Present State Examination. The revised version was derived from the Schedules for Clinical Assessment in Neuropsychiatry (SCAN), and makes ICD 10 diagnoses using the SCAN diagnostic program. This current version has a 4-point scale of severity, compared with the 3-point scale of the first version. It also has a new module relating to psychotic disorders. The sample consisted of 40 individuals representing a spectrum of neurotic, depressive and psychotic disorders. Videotapes of 40 PAS-ADD interviews were re-rated by trained interviewers who had not been involved in the original study in which the videotapes were produced. The mean Kappa across all individual item codes was 0.65, ranging from 0.94 to 0.35. The mean Kappa agreement on item groups was 0.66. Correlation between total symptom scores was 0.74. Agreement on index of definition was Kappa 0.70. We concluded that, agreement was generally lower than for the ICD 9 version. This was probably due mainly to the increase in the severity categories from three to four. However, the new items (most of which related to psychosis) were of comparable reliability to other items.

Introduction

Recent years have seen the development of a number of instruments designed to improve the diagnosis and detection of psychiatric disorders in people with intellectual disability. These include purpose-designed materials such as the Psychopathology Instrument for Mentally Retarded Adults (Kazdin et al. 1983; Senatore et al. 1985) and the Reiss Screen (Reiss, 1987), as well as a number of instruments adapted for use with this population (Beck et al. 1961; Hamilton 1960; Zung 1965). In general psychiatry, much effort has been devoted to the development of structured and semi-structured clinical interviews with operationally defined diagnostic criteria (Spitzer et al. 1978). The development of these instruments has facilitated communication between investigators and has provided a method of employing the same diagnostic criteria across patient samples. The utility of these instruments with people who have intellectual disability had not until recently been demonstrated. However, our studies have shown the effectiveness and potential of an approach based on semi-structured interviewing. The Psychiatric Assessment Schedule for Adults with Developmental Disability (PAS-ADD; Moss et al. 1993), originally derived from the Present State Examination (PSE), uses parallel versions to interview both the respondent (patient) and a key informant, the final diagnosis being derived from these two sources of information. The patient interview has been designed with a multi-level structure to allow interviewing with a wide range of intellectual ability, reflecting clinical practice. This, plus the reliance on two sources of interview data, provides maximum flexibility and symptom sensitivity and makes the procedure acceptable to respondents and informants. The PAS-ADD is a relatively in-depth assessment covering a wide range of conditions. As such, it provides an assessment system offering different characteristics from brief screening questionnaires, such as the Reiss Screen, or instruments relying on informant data and observation only, or self-report questionnaires or instruments focusing on specific areas of diagnosis, such as depression.

The first version of the PAS-ADD was developed as part of the study of needs and characteristics of older people with intellectual disability (Moss et al. 1992).

H. Costello · S. Moss (✉) · H. Prosser · C. Hatton
Hester Adrian Research Centre, University of Manchester, Oxford Road, Manchester M13 9PL, UK

The core interview was a modification of the Psychiatric Assessment Schedule (Dean et al. 1983; Gask 1988), this latter instrument was based on 40 items of the PSE and was designed to elicit basically neurotic and affective symptomatology. The inter-rater reliability of the PAS-ADD (Moss et al. 1993) has been found to compare favourably with that of the PSE (Wing et al. 1977). This first version of the PAS-ADD proved successful in case detection and diagnosis with people whose developmental level was relatively low. In the Oldham study, the average IQ of those who could be adequately interviewed clinically was 39 (Patel et al. 1993).

The first version of the PAS-ADD lacked the capability of diagnosing psychotic conditions, and was superseded by the advent of the ICD 10 classification system. The revised version, the PAS-ADD 10, was derived from version 1 of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN; World Health Organisation 1992). It uses the SCAN's glossary, in an unmodified form, to provide the clinical definitions for coding. The validity of diagnoses of schizophrenia made using the PAS-ADD in relation to the clinical opinion of referring psychiatrists has recently been investigated (Moss et al. 1996). A significant change from the PSE is that the SCAN uses a 4-point scale of severity rather than a 3-point scale. It was anticipated that this would reduce the inter-rater reliability when compared with our earlier study.

The general approach to the development and use of the PAS-ADD has been to devise a core interview focusing on the more common Axis I disorders, other conditions being handled by additional modules covering important missing diagnostic categories and areas of information. These appended sections were not themselves modified. A series of screening items were added to the PAS-ADD, triggering of the screening item leading to further diagnostic interviewing in the specific area using the appropriate appended module. Thus, in the original project on ageing (Patel et al. 1993; Moss et al. 1993), additional modules covered the following: psychoses, autism, substance abuse, problem behaviours and dementia. The ICD 10 version of the PAS-ADD now includes a module on psychoses, the core interview currently covering the following ICD 10 classes of disorder:

1/F20	Schizophrenia
2/F32	Depression (severity at least F32.0)
3/F40	Phobic anxiety disorders
4/F41	Other anxiety disorders

The inter-rater reliability of the patient core interview (agreement between pairs of ratings for the same interview) was the focus of this study.

Available data for computing agreement included the following:

1. *Individual items*: the core interview consists of 145 questions adapted from the SCAN, most of which are rated on a 4-point scale of severity. (This is a change from the original version of the PAS-ADD, which was derived from the PSE and used a 3-point scale). The most stringent test of inter-rater reliability pertains to these individual items. If the interview is highly reliable, it will be possible to show that the raters agree not just about the presence or absence of each symptom, but also about the actual coding. It is, however, only possible to make stable estimates in relation to the more frequently rated items. The polydiagnostic approach of the PAS-ADD necessitates that the interview has a relatively large number of items, as a result of which most of the items in a typical interview using the PAS-ADD remain un-scored or zero scored, even for a person with florid symptoms. In the current study, the quoting of Kappas for individual items was therefore restricted to those with a prevalence of over 20%.
2. The first part of the procedure by which the SCAN program produces a diagnosis is to generate a series of *item groups* derived from clinically related items. Individual items only contribute to the item group total score if they are coded within the range that is considered clinically significant. Clinically significant items contribute to the total score for each item group. Wing et al. (1977) have used the presence/absence of item group scores greater than zero as one of their main measures of inter-rater agreement.
3. The total symptom scores were used to express level of agreement using Pearson product moment correlation.
4. The SCAN algorithm produces an "Index of Definition" between 1 and 8, which is a measure of the clinical significance of the observed set of symptoms. In the PSE, level 5 was the minimum level at which a firm diagnosis would be given, level 4 indicating a measure of morbidity, but insufficient for a firm diagnosis. While the principle remains the same, the SCAN algorithm operates differently in that diagnoses can, in some cases, be given with an Index of Definition of only 2.

Method

The sample

Forty individuals were drawn from a larger sample of adults with learning disability who were involved in another study, in the course of which they all completed a PAS-ADD interview. To generate the original sample, psychiatrists working with patients who had learning disability were contacted and asked if they would refer patients to us who (a) had sufficient verbal ability to attempt a clinical interview and (b) were thought to have a disorder within the spectrum covered by the core interview. Although all the sample members had recent contact with psychiatric services, the number of active symptoms at the time of interview varied widely. In some cases, the referring psychiatrist indicated that the individual was not currently ill. The sample of 40 was chosen to represent a wide spectrum of symptoms and non-symptoms. Part I of the Adaptive Behaviour Schedule (ABS) (Nihira et al. 1974) was completed by a key informant. These data allowed us, in conjunction with IQ and

ABS data collected in an earlier project, to derive estimates of IQ using a multiple regression technique (Hogg and Moss 1995).

Procedure

Videotapes of the original PAS-ADD interviews were re-rated by staff who had been SCAN trained and had experience of using the PAS-ADD. They had not, however, been involved in the original study, and made the ratings blind to any knowledge of the referrer's diagnosis or PAS-ADD codings made by the original interviewer.

In our reliability study of the ICD 9 version, pairs of raters observed interviews and were allowed as many re-runs of the interview or sections of the interview as necessary to achieve optimum confidence in the resulting codes. This was obviously an ideal situation in which to rate. This time, we allowed the re-rater only one view of the interview, on the basis that this would be a more accurate reflection of the real-life coding situation faced by the in vivo interviewer.

Results

Characteristics of the sample

Age, sex and number of symptoms as calculated by the SCAN program from the interviewers' codings are shown in Table 1.

Table 1 Sample characteristics

Variable	Mean	SD	Minimum	Maximum
Age	37.53	12.19	18	69
IQ	39.17	9.37	20	55
Depressive	3.84	5.06	0	20
Psychotic	5.87	7.82	0	29
Neurotic	8.58	6.88	0	24
Total symptoms	21.32	13.88	1	66

Table 2 Agreement on individual item codes

Symptom	Kappa value
Irritability	0.94
Effect of sleeping tablets	0.74
Delayed sleep	0.72
Second-/third-person auditory hallucinations	0.67
Frequency of auditory hallucinations	0.67
Weight gain	0.66
Most severe specific phobia	0.64
Specific phobia – storms	0.62
Restlessness	0.61
Subjective health	0.61
Heart pounding	0.60
Avoidance of phobia	0.60
Subjective nervous tension	0.57
Sweating	0.48
Most severe situational phobia	0.44
Trembling hands	0.35

Coding of individual items

The mean Kappa across all the individual item codes was 0.65. The earlier study on the ICD 9 version (items rated on a 3-point scale) generated a mean Kappa of 0.77. The most reliable items were irritability (0.94), the effect of sleeping tablets (0.74) and delayed sleep (0.72). In the 1993 study, the item "delayed sleep" had a low Kappa (0.30). Changes in the rating criteria that were made in the light of the earlier findings would therefore seem to have been effective. In the current study, the least reliable codings were blunting of affect (0.48), sweating (0.48), most severe situational phobia (0.44) and trembling hands (0.35).

Table 2 shows the mean Kappa values for items on the interview schedule that received at least eight (20%) non-zero responses. It can be seen that the four items that had Kappa values less than 0.6 were all associated with anxiety symptoms. Elsewhere (Moss et al. 1997) we have found that anxiety disorders represent one of the more difficult classes of disorder to diagnose, the agreement between information obtained from the PAS-ADD and clinicians' reports being lower than for other symptom areas.

The presence/absence of item groups

The SCAN algorithm derives 26 item groups from the items in the PAS-ADD. To ensure stability of the results, the computation of individual Kappas was restricted to the 13 item groups in which at least 20% of the subject interviews yielded other than 0/0 agreement. The results are shown in Table 3. The mean Kappa for agreement between interviewer and rater on the presence or absence across 13 item groups was 0.66. This compares favourably with that obtained in the study by Wing et al. (1977) which quotes an equivalent figure of 0.52 over 13 item groups. The three item groups yielding a Kappa of below 0.6 were: flat and incongruous affect, nervous tension, and autonomic anxiety and panic. Thus, as with the analysis of individual items, the reliability of coding with respect to anxiety symptoms was found to be lower. The relatively low reliability for flatness of affect probably relates to the problem of distinguishing negative symptoms in people with learning disability. This point is discussed in more detail later.

Total PSE scores

The mean product moment correlation between the pairs of raters was 0.74. This is considerably lower than in the studies by Moss et al. (1993) and Wing et al. (1977), both of which quoted a correlation of 0.96.

Table 3 Inter-rater agreement on presence/absence of syndromes

Syndrome	Kappa
Lowered bodily functioning	0.84
Depressed mood	0.83
Specific simple phobia	0.82
Non-specific auditory hallucinations	0.74
Bizarre delusions and interpretations	0.72
Positive functioning	0.69
Agoraphobia	0.66
Lowered subjective functioning	0.60
Muscular tension	0.60
Non-affective auditory hallucinations	0.60
Flat and incongruous affect	0.54
Nervous tension	0.50
Autonomic anxiety and panic	0.47

Index of Definition

In keeping with the studies by Wing et al. (1977) and Moss et al. (1993), the data were dichotomised into above and below threshold values (0–4, 5 +), reflecting the diagnostic distinction between levels four and five (see Table 3). The mean Kappa for agreement across the Index of Definition was 0.70, again lower than in the studies by Moss et al. (0.91) and Wing et al. (0.89). A clue to the lower overall reliability figure in this study came from the discovery that all the disagreements in Table 4 came from one interviewer. In each case, the interviewer had rated psychotic symptoms positively, whereas the rater had either felt that the respondent was acquiescing or that the respondent had not been probed sufficiently. This interviewer had been part of the project team for a limited period of time, and had not been deeply involved in interview development. The other interviewer, on the other hand, had had a longer involvement in the work, and was also involved in training the staff who performed the re-ratings. Other things being equal, higher quality interviewing is likely to produce better interrater reliability.

Table 4 Cross tabulation of Index of Definition

		Interviewer (Rater 1)	
		1–4	5 +
Rater 2	1–4	20	5
	5 +	1	14

Discussion

Overall, the study showed that the reliability with which individual items were coded was in most cases acceptable. Earlier problems in relation to the rating of sleep problems were clearly improved in the ICD 10 version, as indicated by an improvement in Kappa from 0.30 in the previous study to 0.72 in the present one. The greatest unreliability was in relation to symptoms of anxiety. One of the reasons for this is that it is very difficult to get information on autonomic symptoms from a person with learning disability without probing to an extent that can lead to acquiescence (or to a response where it is unclear whether the respondent is acquiescing). Another reason is that descriptions of panic attacks or phobic anxiety demand a fairly high level of verbal and intellectual ability. Elsewhere (Patel et al. 1993) we have shown that reportage of anxiety and depression symptoms has a positive correlation with IQ. The same finding applied to the sample involved in the field trials of the PAS-ADD from which the current sample was drawn. Reportage of neurotic and depressive symptoms was positively and significantly related to IQ (Moss et al. 1997).

Since the reliability of items relating to sleep difficulty showed such a marked improvement between the earlier and the subsequent study, it may be possible to make further progress in relation to these symptoms of anxiety. Certainly, attempts will be made in this direction. However, the project team's experience to date indicates that it may be very difficult to make further major improvements in the reliability of these particular items.

Generally speaking, the level of agreement reported was not as high as for our study of the ICD 9 version. The main reason for this was almost certainly the increased number of severity coding categories – from three to four. Given the same level of error, one would expect this to result in a lower overall percentage agreement. Additionally, the rating situation was purposefully chosen to be less ideal than in the earlier study, with the aim of being a more accurate reflection of the likely reliability in situations where the interview designers are not directly involved in the work. In the first study, the raters were all deeply involved in the development of the PAS-ADD, and had as a result a great familiarity with the items and the glossary items from the PSE to which they related. The raters were also given the best possible chance to make correct ratings, with as many re-runs of the tape as they wished. This time, the tape was viewed only once, and not all the personnel were so deeply involved in the PAS-ADD's development. One of the two interviewers was more experienced than the other and was also involved in training the new staff (who were themselves the re-raters of the tapes). The impact of this was shown to be reflected in the higher levels of disagreement in relation to the less experienced interviewer. The re-raters themselves were also relatively new to the PAS-ADD. However, they had been working on the interview, and were able to produce a good level of agreement on the Index of Definition when rating the tapes of an experienced interviewer. This suggests that the skills of clinically interviewing people with learning disability are one of

the most important elements in obtaining high reliability. Indeed, no one would doubt the fundamental importance of good interviewing in any project of this kind. It is clear, however, that these skills are, if anything, even more crucial in the case of people with learning disability. There are fewer chances to make a mistake since the respondent is likely to be linguistically and socially less confident, and hence less likely to correct the interviewer. Questions must thus be asked empathically, yet with the greatest possible clarity, and the interviever must always be alert to the possibility that the respondent may acquiesce.

Apart from the increased complexity of coding, the greatest change to the PAS-ADD 10 has been the addition of the psychotic module. It had been anticipated that the complexity of psychotic phenomena might make their coding considerably less reliable than for other items. In fact, this did not turn out to be the case–at least in relation to those symptoms that could be detected with sufficient frequency to make a reasonable estimate. This applied primarily to auditory hallucinations. As we have shown elsewhere (Moss et al. 1996; 1997), other symptoms of schizophrenia that are frequently reported in general population cases, i.e. thought disorder, replacement of will and primary delusions, are much more difficult to detect in people with learning disability. This very different pattern of prevalence from the general population certainly demands further attention in relation to the diagnosis of schizophrenia in this population. In the present context, however, it appears that the new psychotic module showed adequate reliability.

Overall, it is clear that people wishing to use the PAS-ADD in the future need to pay close attention to training, both of interviewing skills and of coding skills. For research projects it is clearly desirable to have regular monitoring to ensure lack of drifting in coding. All personnel involved in the PAS-ADD development projects have been SCAN trained, and this is considered essential for anyone wishing to use the instrument. Exactly what form any additional training should take has yet to be decided, although pilot training courses indicate that rating of videos, discussion of ratings in relation to the SCAN glossary definitions and role-played interviewing are all valuable components.

References

Beck A, Ward CH, Mendelson M, Mock JE, Erbaugh J (1961) An inventory for measuring depression. Arch Gen Psychiatry 4: 561–571

Dean C, Surtees P, Sashidharan SP (1983) Comparison of research diagnostic systems in an Edinburgh community sample. Br J Psychiatry 142: 247–256

Gask L (1988) The Psychiatric Assessment Schedule (revised). Dept of Psychiatry, Manchester University, Manchester

Hamilton M (1960) A rating scale for depression. J Neurol Neurosurg Psychiatry 23: 56–61

Hogg J, Moss SC (1995) The applicability of the Kaufman Assessment Battery for Children (K-ABC) with older adults (50+ years) with mental retardation. J Intellect Disabil Res 39: 167–176

Kazdin AE, Matson JL, Senatore V (1983) Assessment of depression in mentally retarded adults. Am J Psychiatry 140: 1040–1043

Moss SC, Hogg J, Horne M (1992) Individual characteristics and service support of older people with moderate, severe and profound learning disability with and without community mental handicap team support. Ment Handicap Res 6: 3–17

Moss SC, Patel P, Prosser H, Goldberg DP, Simpson N, Rowe S, Lucchino R (1993) Psychiatric morbidity in older people with moderate and severe learning disability (mental retardation). Part I. Development and reliability of the patient interview (the PAS-ADD). Br J Psychiatry 163: 471–480

Moss SC, Prosser H, Goldberg DP (1996) Validity of the schizophrenia diagnosis of the Psychiatric Assessment Schedule for Adults with Developmental Disability (PAS-ADD). Br J Psychiatry 168: 359–367

Moss S, Ibbotson B, Prosser H, Goldberg D, Patel P, Simpson N (1997) Validity of the PAS-ADD for detecting psychiatric symptoms in adults with learning disability (mental retardation). Soc Psychiatry Psychiatr Epidemiol 32: 344–354

Nihira K, Foster R, Shellhaas M, Leland H (1974) AAMD Adaptive Behavior Scale, 1974 revision. American Association on Mental Deficiency, Washington, DC

Patel P, Goldberg DP, Moss SC (1993) Psychiatric morbidity in older people with moderate and severe learning disability (mental retardation). Part II. The prevalence study. Br J Psychiatry 163: 481–491

Reiss S (1987) Reiss Screen for Maladptive Behavior. International Diagnostic Systems, Chicago

Senatore V, Matson JL, Kazdin AE (1985) An inventory to assess psychopathology of mentally retarded adults. Am J Ment Defic 89: 459–466

Spitzer RL, Endicott J, Robins E (1978) Research diagnostic criteria: rationale and reliability. Arch Gen Psychiatry 35: 773–782

Wing J, Nixon J, von-Cranach M, Strauss A (1977) Further developments of the 'Present State Examination' and CATEGO system. Arch Psychiatr Nervenkr 224: 151–160

World Health Organisation (1992) Schedules for Clinical Assessment in Neuropsychiatry, version 1. WHO, Geneva

Zung WWK (1965) A self-rating scale for depression. Arch Gen Psychiatry 12: 63–70