

# BJPpsych

The British Journal of Psychiatry

## Psychiatric morbidity in older people with moderate and severe learning disability. I: Development and reliability of the patient interview (PAS-ADD).

S Moss, P Patel, H Prosser, D Goldberg, N Simpson, S Rowe and R Lucchino  
*BJP* 1993, 163:471-480.

Access the most recent version at doi: [10.1192/bjp.163.4.471](https://doi.org/10.1192/bjp.163.4.471)

---

**Reprints/  
permissions**

To obtain reprints or permission to reproduce material from this paper, please write to [permissions@rcpsych.ac.uk](mailto:permissions@rcpsych.ac.uk)

**You can respond  
to this article at**

<http://bjp.rcpsych.org/cgi/eletter-submit/163/4/471>

**Downloaded  
from**

<http://bjp.rcpsych.org/> on September 26, 2011  
Published by The Royal College of Psychiatrists

---

## Psychiatric Morbidity in Older People with Moderate and Severe Learning Disability

### I: Development and Reliability of the Patient Interview (PAS-ADD)

STEVE MOSS, PRADIP PATEL, HELEN PROSSER, DAVID GOLDBERG,  
NEILL SIMPSON, STEVE ROWE and RON LUCCHINO

This paper describes the development of the PAS-ADD, a semistructured clinical interview for use specifically with patients with learning disabilities, based on items drawn from the PSE. The PAS-ADD includes a number of novel features including: parallel interviewing of patient and informant; a three-tier structure to provide a flexible interview appropriate to the patient's intellectual level; use of a memorable 'anchor event' in the patient's life to improve time focus; and simplified wording, improved organisation and lay out. Inter-rater reliability was investigated using an experimental design in which two raters viewed and re-rated videotaped PAS-ADD interviews which had been conducted by an experienced clinician. Reliability results compared favourably with those obtained in a major study of PSE reliability with a sample drawn from non-learning disabled individuals. Mean kappa for all items was 0.72. Other indexes of reliability were also good. In the current phase of development, the PAS-ADD is to be expanded to include further diagnostic categories, including schizophrenia and autism. The new version will be updated for use with ICD-10 criteria.

This paper discusses the problems of psychiatric interviewing and diagnosis of people with learning disabilities (LD) (mental retardation), and describes the development of a new psychiatric semistructured clinical interview specifically for use with LD individuals – the Psychiatric Assessment Schedule for Adults with a Developmental Disability (PAS-ADD; Moss & Goldberg, 1991). The following paper (this issue) presents the findings of a prevalence study of psychiatric morbidity in older people with LD. This latter study used the PAS-ADD as the primary assessment of mental health, and a variety of other assessment techniques relating to IQ, functional ability, and dementia; it was part of a larger demographic study of people with LD over 50 years of age, commissioned by the Department of Health and conducted in the Oldham Metropolitan Borough (Moss & Hogg, 1989; Moss *et al.*, 1992a,b; 1993).

#### The need for a new clinical interview

One of the main problems to be tackled in studies of the mental health of people with LD is the relative lack of suitable techniques for detection and diagnosis in this population. In this respect, the problem of interviewing people with LD about their mental state is clearly a major obstacle. People with LD are likely to find it difficult to express their emotions verbally, as a result of which many studies have relied on third-party reports for information

on which to make a diagnosis (Sturmey *et al.*, 1991). Although information from key informants is certainly of paramount importance in the detection and assessment of psychiatric morbidity in this population, we would argue that the confidence in a diagnosis based solely on third-party reports cannot be high. In addition, the existing instruments for psychiatric evaluation of people with LD tend to be brief questionnaires rather than detailed instruments, for example the Psychopathology Instrument for Mentally Retarded Adults (PIMRA; Matson *et al.*, 1984), the Reiss screen (Reiss, 1987) and the Diagnostic Assessment for the Severely Handicapped (DASH) scale (Matson *et al.*, 1991). Although valuable for screening purposes, such evaluations cannot provide the depth and level of detail necessary to make an accurate diagnosis. These screening instruments are often designed for use by interviewers who are not clinically trained. In comparison, the PAS-ADD is a semistructured interview allowing a considerable degree of flexibility in use. It thus requires appropriate training in mental health assessment, but capitalises on that experience to improve the overall sensitivity and validity of diagnosis.

The main body of the interview concentrates on the more common neurotic conditions found in community samples, other conditions being handled by a number of additional modules referenced by 'skip-out' items within the body of the interview.

Within the context of the Oldham psychiatric study (following paper), no modification of the additional modules for use specifically with LD patients was attempted. Rather, development work focused on the main body of the interview, and it is on this section which the present reliability study was conducted.

### Overview of the PAS-ADD

The current version of the PAS-ADD is a modification of the Psychiatric Assessment Schedule (PAS; Dean *et al*, 1983; Gask, 1988), this latter instrument being based on the 40 items of the Present State Examination (PSE; Wing *et al*, 1974) designed to elicit basically neurotic symptoms, with the addition of extra items in order to make it capable of making DSM-III-R diagnoses (American Psychiatric Association, 1987) relating to major depression, generalised anxiety, dysthymia, panic disorder, and agoraphobia.

Apart from simplification of wording, PAS-ADD has a number of novel features:

- (a) parallel patient and informant interviewing, these two sets of data being combined to increase sensitivity
- (b) a three-tier structure, designed to make the interview flexible for use with a wide range of linguistic abilities
- (c) the use of a memorable 'anchor' event in the subject's life, which helps focus discussion on the four weeks immediately before the interview
- (d) reorganisation of the items to maximise ease of interviewing with LD individuals
- (e) clear visual cues for the interviewer to the conditional jumps which are to be made if a previous question or series of questions has indicated that a line of questioning can be terminated - this increases the focus and efficiency of interviewing, minimising the risk of loss of attention by the patient.

As we show in the following paper, the PAS-ADD has proved successful in case detection and diagnosis with people whose developmental level is relatively low. In people with LD over 50 years of age we have achieved adequate clinical interviewing with a group whose average IQ was only 39. Seventy-five per cent of the cases we detected were unknown to psychiatric services. Using ICD-9 (World Health Organization, 1978) and DSM-III-R algorithms, our findings have indicated that the PAS-ADD is successful in detecting cases of major depression, generalised anxiety, panic disorder, and agoraphobia. The interview is currently being further developed for use

with ICD-10 (World Health Organization, 1992a), covering a broader range of conditions, including schizophrenia and pervasive developmental disorders (Moss & Goldberg, 1991).

### Development of the PAS-ADD

The general strategy was to select a suitable existing interview for modification, rather than to devise a new instrument from scratch. This had the advantage of providing a framework based on an existing research and clinical foundation, which would additionally provide comparative data on reliability. The extent of modification was to be limited to item organisation and wording, thus preserving the validity of the original interview in relation to the spread of symptoms to be elicited and the constitution of the diagnostic algorithms. This approach naturally made the assumption that symptoms and syndromes occurring in LD people are the same as in the general population. While future research may prove this assumption to be false, the current evidence suggests that psychiatric symptoms in LD and non-LD individuals have similar clinical significance (Philips, 1967; Menolascino, 1970; Reid, 1972a,b; Eaton & Menolascino, 1982; Reiss, 1982).

In making the choice of starting instrument, the following features were considered important:

- (a) if it asks patients about presenting symptoms, their duration and historical development
- (b) if it examines mental state
- (c) if it uses informant data to corroborate history and additional information
- (d) if it uses information from case notes and other relevant medical records
- (e) if it is standardised and repeatable
- (f) if it allows standardised research diagnoses using both ICD-9 and DSM-III-R systems
- (g) if it is of the simplest possible linguistic structure commensurate with an appropriate degree of sensitivity to, and discrimination between, symptoms.

In consideration of the above criteria we selected four possible instruments:

- (a) the Composite International Diagnostic Interview, revised (CIDI-R) (Robins, 1985)
- (b) the Structured Clinical Interview for use with DSM-III-R (SCID-II) (Spitzer *et al*, 1987)
- (c) the Schedule for Clinical Assessment in Neuropsychiatry (SCAN) (World Health Organization, 1992b)
- (d) the Psychiatric Assessment Schedule (PAS) (Gask, 1988).

While being less comprehensive than any of the other three instruments, the PAS possessed a number of qualities which recommended it to our purpose:

- (a) the language was specifically chosen to be straightforward – an important starting point when considering the needs of LD individuals
- (b) it was designed for use by an experienced practitioner
- (c) its clinical ratings were on a three-point scale – the rating of symptoms as not present, present to a moderate degree, or present to a severe degree, was considered the maximum practical level of precision which we could hope to achieve with subjects who have reduced linguistic and intellectual levels
- (d) it was specifically designed to give both DSM-III-R and ICD-9 diagnoses.

The obvious shortcomings of the PAS lay in its omission of psychotic symptoms, behavioural disorders, and the optional modules available within the SCAN. To circumvent these limitations, additional instruments or sections of instruments were appended to the modified PAS to cover 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, triggering of the screening item leading to further diagnostic interviewing in the specific area using the appropriate appended module.

#### Modification of the PAS

In considering the modification which might make the interview more usable with LD individuals, it was apparent that only a certain amount of improvement could be achieved by changes in the working, since some of the interview areas are basically more conceptually difficult than others. Experience of interviewing people with LD in a wide range of circumstances indicated to us the importance of minimising the impact of failure. Inability to answer questions can lead rapidly to a sense of failure or helplessness, and a wandering of attention (Sigelman *et al.*, 1981). Such effects are likely to lead to an effective termination of the interview. An initial modification, namely organising the items in terms of difficulty, was rejected because it would lead to incoherent lines of questioning. This would cause the clinician constantly to refer back to previous answers.

Our interview was based on a design in which the clinician is required to make one of three decisions about the possibility of interviewing a person with LD: (a) that the result of an initial opening section would indicate that successful clinical

interviewing is impossible; (b) that intellectual level would permit *some* items to be answered; or (c) that the full interview could be successfully completed. A three-tier interview was thus developed on the following basis. The first tier is an introductory 'open' section at the start of the interview, designed to put the client at ease and to give the psychiatrist an opportunity to gauge the linguistic ability and time concept of the subject. For the second tier, conceptually simpler questions were identified by one of the project team (NS) on the basis of his extensive experience of psychiatric interviewing with this population. To these simpler items were added a few more questions to produce a set of 'core items', the aim being to constitute the minimum item set necessary to identify possible cases. These additions were necessary to make it responsive to diagnostic concepts such as anxiety and depression. It was hoped that, with this second tier of core items, the interview could be used successfully to identify possible cases, even where the person's linguistic ability is limited. The third tier is the full set of items.

Second- and third-tier items are clearly distinguished within the schedule. When presented with a linguistically less able patient, the clinician could thus skip easily over the more difficult items, yet retain the flow of the interview.

The introductory section is crucial because it gives the clinician the opportunity to judge the adequacy of the subject's account of symptoms and of the subject's time perspective. Accurate information on time-course is essential to reach a correct diagnosis, yet this can be difficult to obtain from patients with LD. In the PAS-ADD, the patient's perception of time is focused by means of an 'anchor event' – an event which, in prior discussion with the informant, fixes a period of around four weeks before the interview, and which is memorable to the patient. The anchor event is introduced in the introductory section, and regularly referred back to during the interview. During the introductory section, discussion of the anchor event provides a context within which the clinician codes the adequacy of time-concept. Coding of adequacy of account of symptoms is performed immediately at the end of the interview.

The interviewer is encouraged not to make an early decision that the subject can only complete the core items. In the case of reduced linguistic ability, the clinician should certainly ensure that these items are completed, but should also try to obtain answers to as many other relevant questions as possible.

The loss of clearly answered items is bound to have a detrimental effect both on the confidence of clinical decisions and on the ability of algorithms to compute

a positive diagnosis. To minimise the extent of this loss, the range of core items was constructed to give a degree of information within each diagnostic category. In some cases, the core items are sufficient to give the potential of a firm diagnosis. Moss *et al* (1991) describe the loss of diagnostic power of ICD-9 and DSM-III-R algorithms where the clinician could complete only the core items.

In addition to the partitioning of items described above, further efficiency was introduced into the interview by making explicit the conditional jumps which the interviewer can make if a previous question or series of questions has indicated that line of questioning can be terminated. This is also designed to aid the flow of the discussion, minimising further the risk of wandering attention.

#### **The additional modules**

These are briefly mentioned to complete the description of the current PAS-ADD. At the present time they have not been modified specifically for use with LD individuals, and may be substituted or changed in the current phase of development.

#### *Psychoses*

If the subject showed evidence of expansive mood (question 54 on the PAS-ADD), section eight of the SCAN was invoked. If there was evidence of hallucinations or delusions (questions 65 and 66 on the PAS-ADD), part two of the SCAN was completed. This was chosen as being the most up-to-date version of the PSE available, that is the same source as that from which the PAS was generated.

#### *Autism*

During the main clinical interview using the PAS-ADD, screening for autism was performed using 14 behavioural observation items based on DSM-III-R criteria. In cases where there was sufficient evidence to suspect autistic disorder, an additional 37-item informant questionnaire was completed. This latter questionnaire was devised by Tantam (1986, 1988) in a study of adults with Asperger's syndrome and autistic disorder. The author provided us with a key for the diagnosis of DSM-III-R autistic disorder, and Gillberg & Gillberg's (1989) criteria of Asperger's syndrome. The screening items from the PAS-ADD were also incorporated in this key so that the final diagnosis would incorporate both informant data and clinical observation.

Tantam's questionnaire was chosen in preference to the more well known Handicap, Behaviour and Skills questionnaire (HBS; Wing & Gould, 1978) as

being more suited to the requirements of the prevalence study described in the following paper. The HBS is a detailed informant interview taking up to 45 minutes to complete and needing specialised training. In addition, it contains items assessing functional behaviour - information which was already available from other assessments we had made. The items in Tantam's questionnaire, on the other hand, related specifically to the clinical conditions under consideration.

#### *Alcohol*

AUDIT (Babor *et al*, 1989) is a 10-item questionnaire used to identify people whose alcohol consumption has become excessive or dangerous. Each item can be scored from 0 to 4, a score of 8 or more signifying a positive case. In addition, high scores on specific combinations of items indicate 'hazardous alcohol use', 'alcohol dependence', and 'harmful alcohol use'. The questionnaire has been validated in primary health care settings in six countries (Babor *et al*, 1989), and found to have good sensitivity and specificity.

#### *Behaviour problems*

Unlike the modules relating to psychoses, autism, and alcohol, there was no specific filter item in the PAS-ADD relating to behaviour problems. This is an area which can only be investigated by observation and functional analysis, coupled with informant interviewing. The latter is more likely to yield more valid information than the clinical interview, since a normal clinical interview is too short to give an overall impression and since behaviour problems are often related to settings and individuals. The informant questionnaire we used for this purpose was drawn from the instrumentation developed by Qureshi *et al* (1990).

#### **Assessment of reliability**

This initial reliability study focuses on the 66 questions in the main body of the interview. It is concerned with the reproducibility (reliability of comparisons between sets of ratings of the same interview) rather than repeatability (reliability of comparisons between sets of ratings derived from interviews of the same person conducted at different times) (Wing *et al*, 1977). This latter rating will be explored during the current phase of development. Ratings of repeatability are naturally lower than reproducibility, involving such factors as changes in the individual's mental state, effects of repeated interviewing, interviewer training, amalgamation of information from patients and

informants, and the use of data from medical records.

Since the PAS and PAS-ADD are derived from the PSE, we were able to draw on reliability data from this latter interview for comparison purposes (Wing *et al*, 1977).

#### Available data for computing agreement

Within the PSE (and PAS-ADD), symptoms are rated on a three-point scale of severity: absent (0); present in a moderate degree (1); present in a severe degree (2). Additional ratings are also available (a) if the appropriate questions were asked and answered with evasion and incoherence, but the interviewer is nevertheless unsure whether the symptom is present or not (rated 8), (b) if it is inappropriate to make a rating because some aspect of the examination is missing due to language disorder, refusal to answer, and so on (rated 9).

The CATEGO computer algorithm derives various measures from the PSE data based on ICD-9 decision rules (Wing *et al*, 1974). The program sorts symptoms into clusters called syndromes, which can be expressed in terms of scores based on the summed constituent symptoms, and also in terms of presence/absence of the syndrome. Wing *et al* used the presence/absence of individual syndromes as one of their main measures of inter-rater agreement.

The total of the syndrome scores gives rise to a total PSE score, giving an indication of the overall presence of psychiatric symptoms. They have been used by Wing *et al* to express level of agreement between two raters using Pearson's product moment correlation.

CATEGO produces an 'index of definition' for the subject of the interview, which incorporates rules for deciding eight levels at which the functional psychoses and neuroses can be recognised as 'psychiatric disorders'. The rules operationalised within the programme are an embodiment of clinical experience concerning the confidence with which a diagnosis can be made. Level 5 is regarded as a minimum basis for classifying a disorder, level 4 indicating a measure of morbidity, but insufficient for a firm diagnosis. Levels 6, 7, and 8 provide an increasing degree of certainty that the symptoms present can be classified into one of the conventional categories of the functional psychoses and neuroses (Wing *et al*, 1977).

The most stringent test of inter-rater reliability pertains to the individual items, that is the symptoms. 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 (0, 1, 2). Wing *et al* presented only limited data on the presence or absence of certain symptoms. In the present study, we calculated reliabilities in relation to all symptoms. We present mean inter-rater reliabilities for all questions, and give details of those symptoms on which it was more difficult to gain good agreement.

#### Indices of agreement

The original study of Wing *et al* (1977) expressed the view that simple percentage agreements for presence and absence of symptoms was preferable to the calculation of Cohen's kappa, because the latter does not "reflect the degree of agreement that the characteristic is absent" (Wing *et al*, 1977, p. 508). This does not, however, seem a convincing argument. Kappa does not minimise the importance of frequently used coding categories, but merely takes account of the increased probability of chance agreement when the responses are not evenly distributed. We are therefore of the opinion that kappa is the safer statistic. If good agreement levels can be obtained using kappa, then one can be confident of an overall high reliability for both presence and absence of the symptom. Wing *et al* presented the kappa values for their own data, so we used these for comparison with the present study. Kappa continues to be the preferred statistic for the measurement of agreement in studies of psychiatric interviewing and diagnosis (e.g. Cottler *et al*, 1989; Wittchen *et al*, 1991; Watson *et al*, 1991).

For continuous and semi-continuous data such as the PSE total score, product moment correlation still seems to be regularly used in current studies (e.g. Folgeson *et al*, 1991; Watson *et al*, 1991). Wing *et al* also used this measure for their own total score data. We present it here for comparison, although it must be remembered that correlation does not take account of a possible *numerical* bias between the raters. Since, from a diagnostic point of view, absolute magnitude of the PSE score is highly important, this is therefore not a true measure of reliability.

The index of definition was dichotomised by Wing *et al* into a 2x2 contingency table (0-4, 5+), this categorisation being based on the importance of the distinction between level 4 and level 5. For comparison purposes, we present a similar analysis for our own data. However, a more stringent test of reliability in this respect is absolute agreement with respect to the *exact* index of definition for each subject. We also present an analysis of this form.

### Method

The sample comprised a mixture of 25 people who, on the basis of expert clinical opinion, were probable or definite cases, and some who were clearly not suffering from mental illness. In selecting the former group, the aim was to include people whose range of symptoms would mainly be covered by the existing version of the PAS-ADD. It was hoped to avoid including subjects whose symptoms necessitated use of one of the modules, since these were not being tested at this time. In fact, one subject did show evidence of schizophrenia, which caused the interviewer to use also the relevant section of the SCAN.

The criterion for inclusion on the basis of learning disability was that subjects were either (a) receiving services from a Community Mental Handicap Team, or (b) were resident in a long-stay mental handicap hospital.

Each subject was clinically interviewed by one of two psychiatrists, both of whom were PSE trained, specialising in the psychiatry of learning disability, and experienced in using the PAS-ADD. All interviews were videotaped, informed consent having been obtained from the subject, key informant, and family.

Videotapes were subsequently rated by two independent raters, that is neither of whom had been involved in the clinical interview itself. We adopted this approach because initial pilot work had indicated a potential problem in comparing ratings made by the clinician conducting the interview with re-ratings made from the videotape by a second observer. Clinicians working with LD patients often have to rely heavily on information from third parties, and tend to score the PAS-ADD with this background information in mind. Even if a strategy of talking to the patient *before* the information is consciously adopted, it is difficult not to pick up some clues during the (essential) period of familiarisation with the patient. In addition, people with LD sometimes have speech which is difficult to understand, but which may become easier for the interviewer with practice. Overall, the effect is to make the rating tasks unequal for the interviewer and re-rater.

Two lay raters (i.e. non-psychiatrists) were employed to conduct the rating. One was a research psychologist with PSE training, who had considerable experience of interviewing people with LD. The second rater was a gerontologist, with experience of people with LD but new to the field of psychiatric interviewing.

Our pilot work had shown the potential difficulties of obtaining reliable codings, particularly in relation to certain symptoms. The prior period of rater training was therefore crucial to the obtaining of good reliabilities. During this period the raters spent several days rating, using practice tapes, after which the whole project team spent a number of sessions carefully going over each item, clarifying coding rules and agreeing criteria. (This information, apart from providing guidance on the next version of the PAS-ADD, will eventually be incorporated into the clinical glossary.) When it was felt that the maximum possible level of concordance between the raters had been achieved, rating of the experimental tapes began.

Each of the raters coded the interviews independently, with as many re-runs of the interview or sections of the

interview as necessary to achieve optimum confidence in the resulting codes. The opportunity to re-run is obviously not possible in the interviewer's *in vivo* situation; on the other hand, the interviewer is in control of the interview, and can choose to go over material more intensively if the coding seems to be uncertain. Our choice was therefore considered a reasonable basis on which to estimate reliability in the video-rating mode.

Codes for the two raters were analysed by CATEGO to provide the three measures discussed above, that is syndrome presence/absence, PSE total symptom score, and index of definition. In addition, Cohen's kappa was calculated for all interview questions from their 3×3 contingency tables based on the 3-point rating system. For this purpose, all ratings 8 and 9 were counted as 0.

### Results

In the case of two of the subjects, both raters coded all items as 9: in one case this was because the individual was deemed to have given answers which could not be confidently rated; in the other case the problem was a combination of poor speech production and bad sound-recording quality. These two interviews were therefore excluded from the current analysis. It is important to note, however, that in normal clinical usage the PAS-ADD is designed for parallel interviewing of both patient and informant, the final diagnosis being reached by a combination of the information from both these sources. Thus, a patient who cannot be interviewed can nevertheless receive a diagnosis on the basis of the informant interview alone. This approach is demonstrated and evaluated in the following paper.

#### Reliability of individual interview items

Apart from the symptom items, both interviews have a rating of the adequacy of the patient's account of symptoms, rated on a 4-point scale:

- 0 - subject responds adequately
- 1 - account somewhat inadequate, but interview can proceed
- 2 - account seriously inadequate, but interview proceeds in an attempt to rate some subjective responses, as well as behaviour, affect, and speech
- 3 - impossible to continue with interview - only behaviour, affect, and speech sections rated.

In addition, the PAS-ADD requires the interviewer to rate the subject's ability to recall the anchor event on a 2-point scale. These two initial items will be considered first.

#### Recall of anchor event

Agreement between the raters on recall of the anchor event was perfect (kappa ( $k$ ) = 1). Twenty subjects were rated as good, and three as poor. The use of an anchor event is an important aspect of PAS-ADD interviewing. Overall, it was found to be of great benefit to the interviewer, both in focusing the discussion and in evaluating the subject's grasp of time concept.

Table 1  
Cross-tabulation: adequacy of account of symptoms

Rater 2	Rater 1			Row total
	0	1	2	
0	14	0	0	14
1	3	4	0	7
2	0	1	1	2
Column total	17	5	1	23

$k = 0.64$ .

#### Adequacy of account

Table 1 shows the extent of agreement for adequacy of account of symptoms. It can be seen that there appears to be some bias, rater 2 being more prepared to rate the account as inadequate. Since we use clinically non-trained raters one would not necessarily expect them to be able to make a good assessment of this dimension. For future development of the interview, however, this is a potentially important issue; if two clinicians disagree markedly about the adequacy of a clinical interview, the one who regards the interview as less adequate may place more diagnostic significance in the evidence of the informant. If the informant's account is at odds with that of the patient, the two clinicians may draw quite different conclusions.

#### Questions rated on a 3-point scale

Inter-rater agreement ( $k$ ) was calculated for the 57 questions rated on a 3-point scale. These included the 40 PSE items. Mean kappa for all items was 0.77. However, some of the symptoms showed only low prevalence in the sample. To give a more stable estimate, mean kappa was therefore recalculated for the 24 questions receiving at least five (22%) non-zero responses. In this case, the mean kappa was 0.72.

Wing *et al* (1977) do not quote equivalent figures for comparison. However, Kendell *et al* (1968) found a mean kappa of 0.76 for simultaneous ratings of one interview, a figure similar to ours.

For the majority of questions, kappa was greater than 0.6. Three items were, however, below this value: restlessness (0.46), delayed sleep (0.30), and loss of interest (0.58). Delayed sleep seems to be particularly difficult to rate, several of the disagreements being extreme (Table 2). Our subsequent discussions confirmed the difficulty the raters had with this item. Sometimes, for instance, the

Table 2  
Frequency of cases cross-tabulation showing poor agreement on rating of delayed sleep

Rater 2	Rater 1			Row total
	0	1	2	
0	14	0	2	16
1	3	1	1	5
2	1	0	1	2
Column total	18	1	4	23

$k = 0.30$ .

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

Syndrome	Kappa	
	Moss <i>et al</i> (this study)	Wing <i>et al</i> (1977)
Depressed mood	0.83	0.64
Situational anxiety	1.0	0.79
Ideas of reference	0.7	—
Tension	0.65	0.57
Worrying	0.91	0.56
Irritability	0.81	0.30
Social unease	0.90	0.66
Loss of interest	0.62	0.63
Other symptoms of depression	0.79	0.85
Mean kappa	0.80	0.63

subject would report having difficulty getting off to sleep, but then reported sleeping late on a regular basis, or sleeping during the day. Clearly, the exact criteria for rating will have to be further clarified to improve reliability.

#### Reliability of syndrome identification

Of the 38 possible syndromes generated by CATEGO, 13 are available with the PAS-ADD item-set. While all 13 of these syndromes are manifested to some degree by the sample, some received insufficient non-zero ratings to give a stable estimate of reliability. Statistical processing of syndrome data is therefore restricted to those syndromes in which at least five (22%) of the subject interviews yielded other than 0/0 agreement.

Table 3 shows kappas for the nine syndromes fulfilling this criterion, and the corresponding values obtained by Wing *et al* (1977). Generally speaking, our values were considerably higher than Wing *et al*'s, with the exception of the code for 'other symptoms of depression' (OD).

#### Total PSE score

The product moment correlation for the two sets of scores was high - 0.96, a figure identical to that reported by Wing *et al*. Since a score of 11+ is one of the criteria for reaching index of definition level 4, Wing *et al* also measured the level of agreement on two dichotomised categories: 0-10 and 11+. On this basis our figures yielded a kappa of 0.83 (Table 4), compared with Wing *et al*'s figure of 0.89 for their re-rated tapes.

Table 4  
Frequency of cases cross-tabulation: total PSE score

Rater 2	Rater 1		Row total
	0-10	11+	
0-10	12	0	12
11+	2	9	11
Column total	14	9	23

$k = 0.83$ .



Table 5  
Frequency of cases cross-tabulation: index of definition

Index, rater 2	Index, rater 1						
	1	2	3	4	5	6	7
1	4	1					
2		3	1	1			
3			2				
4				2	1		
5					3	1	
6						2	
7							2

#### Index of definition

Table 5 shows the cross-tabulation for index of definition. It can be seen that agreement was good, 18 of the 23 showing a perfect match. Only one of the disagreements was a 2-point discrepancy. Weighted kappa was the statistic chosen to measure agreement, since this takes account of the magnitude of disagreement. The most straightforward weighting criterion is to increase the weighting by one for each cell position away from the diagonal. This yields a kappa of 0.88.

Wing *et al* dichotomised their data into above- and below-threshold values (0-4, 5+), and found a kappa of 0.89. Our value for kappa on this basis was 0.91.

#### Discussion

The present study has shown that, given adequate training, it is possible to achieve levels of inter-rater reliability for re-rated interviews comparable to those achieved by Wing *et al* (1977) for the general population. Particularly encouraging is the high level of reliability which was achieved for nearly all the individual items, not just the global measures of PSE total score and index of definition.

The establishing of good inter-rater reliability is an essential foundation for the PAS-ADD. Here, we have shown good reliabilities obtained by lay raters. In future studies, however, the interview must be shown to have reliability in the wider clinical setting, that is that independent clinicians, having interviewed patient and informant, and in possession of information on medical history, will identify the same individuals as cases and give the same diagnoses. This is a much broader problem, involving not just the properties of the PAS-ADD, but also issues relating to interviewer training, amalgamation of information from patients and informants, and the use of data from medical records. While all these issues are being explored in the current phase of development (described shortly), the amalgamation of information from patients and informants is also a central aspect of the prevalence study presented in the following paper.

Psychiatric interviewing of people with LD is a specialised task demanding skill and experience. Although our clinical interviewers were already experienced in working with this population, it was clear that proficiency increased markedly over two years. This, coupled with the detailed training of the lay raters which was necessary to achieve good inter-rater agreement, highlights the importance of good interviewing and rating skills when working with LD subjects. In this respect, analysis of the interactions between psychiatrist and patient could lead to a fuller understanding of the qualities necessary for successful interviewing with this population. A particularly important consideration when interviewing a person with LD is the likelihood of a communication breakdown due to a lack of understanding on the part of the patient. In such cases, the means by which the psychiatrist 'repairs' the conversation may be crucial in determining the overall success of the interview.

There is a continuing question about the reliability of current verbal diagnostic criteria in non-verbal people, particularly in relation to more complex aspects of mental state. Some would argue (Brugha *et al*, 1988) that even people with borderline and mild LD cannot reliably report on phenomena such as hallucinations and delusions. Further research certainly needs to be done to define the limits of applicability of current predominantly verbal criteria for defining mental illnesses. In our judgement, about half of the current sample could not respond sufficiently to the PAS-ADD for meaningful diagnoses to be made. In these cases, a confident diagnosis would necessitate longitudinal observations of reliably defined behavioural correlates of mental disorder. Future research thus needs to give priority to the identification of clinically relevant, non-verbal criteria of mental disorder in individuals with severe/profound LD.

In a yet wider perspective, it is necessary to explore in more detail the assertion that general psychiatric assessment principles can be successfully used within this population (e.g. Sovner & Hurley, 1983). While it is clear that diagnostic algorithms can produce diagnoses from the interviews of people with LD, it may be that the significance of specific behaviour patterns are different, or of differing magnitude, compared with non-handicapped individuals. Thus, the resulting diagnoses may not be fully valid. This argues for fundamental research into the appropriateness of DSM-IV and ICD-10 algorithms for use with LD patients, perhaps developing new algorithms 'from the bottom up', rather than starting from this unsupported assumption.

### Future developments of the PAS-ADD

Ongoing developments are discussed more fully in the following paper. Briefly, continued development of the PAS-ADD is being funded by the Department of Health, the ultimate aim being to provide a comprehensive ICD-10 clinical interview and accompanying glossary. In the current phase of development (Moss & Goldberg, 1991), the range of the PAS-ADD will be expanded to include the principal diagnostic categories of people with LD seen by psychiatric services. The expanded PAS-ADD will permit ICD-10 diagnoses of: F20 schizophrenia; F32 depression (severity at least F32.0); F40 phobic anxiety disorders; F41 other anxiety disorders; and F84 pervasive developmental disorders.

Longer-term plans involve use of the library of videotaped interviews collected during the project to produce a package of teaching materials designed to develop successful interviewing with the PAS-ADD, and to help care staff develop their awareness of the significance of diagnostically significant behaviour patterns in their LD clients.

### References

- AMERICAN PSYCHIATRIC ASSOCIATION (1987) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised) (DSM-III-R). Washington, DC: APA.
- BABOR, T. F., DE LA FUENTE, J. R., SAUNDERS, J., *et al* (1989) *Audit - Guidelines for use in Primary Health Care*. Geneva: World Health Organization, Division of Mental Health.
- BRUGHA, T. S., COLLACOT, R., WARRINGTON, J., *et al* (1988) Eliciting and reliably rating delusions and hallucinations in the mildly mentally retarded. Paper presented at the 8th IASSMD Congress, Dublin.
- COTTLER, L. B., ROBINS, L. N. & HELZER, J. E. (1989) The reliability of the CIDI-SAM: a comprehensive substance abuse interview. *British Journal of Addiction*, **84**, 801-814.
- DEAN, C., SURTEES, P. & SASHIDHARAN, S. P. (1983) Comparison of research diagnostic systems in an Edinburgh community sample. *British Journal of Psychiatry*, **142**, 247-256.
- EATON, L. F. & MENOLASCINO, F. J. (1982) Psychiatric disorders in the mentally retarded: types, problems, and challenges. *American Journal of Psychiatry*, **139**, 1297-1303.
- FOLGESON, D. L., NEUCHTERLEIN, K. H., ASARNOW, R. F., *et al* (1991) Interrater reliability of the Structured Clinical Interview for DSM-III-R, Axis II: schizophrenia spectrum and affective spectrum disorders. *Psychiatry Research*, **39**, 55-63.
- GASK, L. (1988) *The Psychiatric Assessment Schedule (PAS)* (revised). Manchester: Department of Psychiatry, Manchester University.
- GILLBERG, L. C. & GILLBERG, C. (1989) Asperger's syndrome - some epidemiological considerations: a research note. *Journal of Child Psychology & Psychiatry*, **30**, 631-638.
- KENDELL, R. E., EVERITT, B., COOPER, J. E., *et al* (1968) The reliability of the 'Present State Examination'. *Social Psychiatry*, **3**, 123-129.
- MATSON, J. L., KADZIN, A. E. & SENATORE, V. (1984) Psychometric properties of the Psychopathology Instrument for Mentally Retarded Adults (PIMRA). *Applied Research in Mental Retardation*, **5**, 881-889.
- , GARDNER, W. I., COE, D. A., *et al* (1991) A scale for evaluating emotional disorders in severely and profoundly mentally retarded persons. *British Journal of Psychiatry*, **159**, 404-409.
- MENOLASCINO, F. J. (ed.) (1970) *Psychiatric Approaches to Mental Retardation*. New York: Basic Books.
- MOSS, S. C. & HOGG, J. (1989) A cluster analysis of support networks of older people with severe intellectual impairment. *Australian and New Zealand Journal of Developmental Disabilities*, **15**, 169-188.
- & GOLDBERG, D. P. (1991) *The Psychiatric Assessment Schedule for Adults with a Developmental Disability (the PAS-ADD): Interview development and compilation of the clinical glossary* (Project proposal to the DOH). Manchester: HARC, University of Manchester.
- , —— & PATEL, P. (1991) *Psychiatric and Physical Morbidity in Older People with Severe Mental Handicap*. Manchester: HARC, University of Manchester.
- , ——, ——, *et al* (1992a) Physical morbidity in older people with moderate, severe and profound mental handicap, and its relation to psychiatric morbidity. *Social Psychiatry and Psychiatric Epidemiology*, **28**, 32-39.
- , HOGG, J. & HORNE, M. (1992b) Demographic characteristics of a population of people with moderate, severe and profound intellectual disability (mental handicap) over 50 years of age: age structure, IQ and adaptive skills. *Journal of Intellectual Disability Research*, **36**, 387-401.
- , —— & —— (1993) Individual characteristics and service support of older people with moderate, severe and profound learning disability with and without Community Mental Handicap Team support. *Mental Handicap Research*, **6**, 3-17.
- PHILIPS, I. (1967) Psychopathology and mental retardation. *American Journal of Psychiatry*, **124**, 29-35.
- QURESHI, H., ALBORZ, A. & KIERNAN, C. (1990) *Prevalence of Individuals with Mental Handicap who Show Severe Problem Behaviour: Update of Preliminary Results*. Manchester: HARC, University of Manchester.
- REID, A. H. (1972a) Psychoses in adult mental defectives: I. Manic-depressive psychoses. *British Journal of Psychiatry*, **120**, 205-212.
- (1972b) Psychoses in adult mental defectives: II. Schizophrenic and paranoid psychoses. *British Journal of Psychiatry*, **120**, 213-218.
- REISS, S. (1982) Psychopathology and mental retardation: survey of a developmental disabilities mental health program. *Mental Retardation*, **20**, 128-132.
- (1987) *Reiss Screen for Maladaptive Behavior*. Chicago: International Diagnostic Systems, Inc.
- ROBINS, L. N. (1985) *The Composite International Diagnostic Interview (CIDI) 1-2*, St Louis, Missouri: DIS Training Faculty and Staff, Washington University School of Medicine.
- SIGELMAN, C. K., BUDD, E. C., SPANHEL, C. L., *et al* (1981) When in doubt say yes: acquiescence in interviews with mentally retarded persons. *Mental Retardation*, **19**, 53-58.
- SOVNER, R. & HURLEY, D. A. (1983) Do the mentally retarded suffer from affective illness? *Archives of General Psychiatry*, **40**, 61-67.
- SPITZER, R. L., WILLIAMS, J. & GIBBON, M. (1987) *Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II)*. New York: New York State Psychiatric Institute.
- STURMEY, P., REED, J. & CORBETT, J. (1991) Psychometric assessment of psychiatric disorders in people with learning difficulties (mental handicap): a review of measure. *Psychological Medicine*, **21**, 143-155.
- TANTAM, D. (1986) *PhD Thesis*. London: University of London.
- (1988) Lifelong eccentricity & social isolation. II: Asperger's syndrome or schizoid personality disorder? *British Journal of Psychiatry*, **153**, 783-791.
- WATSON, C. G., JUBA, M. P., MANIFOLD, V., *et al* (1991) The PTSD interview: rationale, description, reliability, and concurrent

- validity of a DSM-III-based technique. *Journal of Clinical Psychology*, **47**, 179-188.
- WING, J. K., COOPER, J. E. & SARTORIUS, N. (1974) *Measurement and Classification of Psychiatric Symptoms: An Instruction Manual for the PSE and CATEGO Program*. Cambridge: Cambridge University Press.
- , NIXON, J. M., MANN, S. A., *et al* (1977) Reliability of the PSE (ninth edition) used in a population study. *Psychological Medicine*, **7**, 505-516.
- & GOULD, J. (1978) Systematic recording of behaviours and skills of retarded and psychotic children. *Journal of Autism and Childhood Schizophrenia*, **8**, 79-97.
- WITTCHEN, H. U., ROBINS, L. N., COTTLER, L. B., *et al* (1991) Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). *British Journal of Psychiatry*, **159**, 645-653.
- WORLD HEALTH ORGANIZATION (1978) *Mental Disorders: Glossary and Guide to their Classification in Accordance with the Ninth Revision of the International Classification of Diseases (ICD-9)*. Geneva: WHO.
- (1992a) *The ICD-10 Classification of Mental and Behavioural Disorders*. Geneva: WHO.
- (1992b) *Schedules for Clinical Assessment in Neuropsychiatry*. Geneva: WHO.

\*Steve Moss, BSc, MSc, PhD, *Senior Research Fellow*; Pradip Patel, MB, ChB, MRCPsych, *Consultant Psychiatrist*; Helen Prosser, BA, *Research Associate, Hester Adrian Research Centre, University of Manchester, Oxford Road, Manchester M13 9PL*; David Goldberg, BM, BCh, MA, DM, MSc, FRCPsych, FRCPhys, *Director of Research and Development, Institute of Psychiatry, London*; Neill Simpson, MB, ChB, MSc, MRCPsych, *Consultant Psychiatrist, Central Manchester Health Authority*; Steve Rowe, MB, ChB, MRCPsych, *Consultant Psychiatrist, Olive Mount Hospital, Liverpool*; Ron Lucchino, BA, MS, PhD, *Director, Institute of Gerontology, Utica College of Syracuse University, Utica, New York*

\*Correspondence

(Received September 1992, accepted December 1992)