

## Reliability of a Procedure for Measuring and Classifying “Present Psychiatric State”

By J. K. WING, J. L. T. BIRLEY, J. E. COOPER, P. GRAHAM  
and A. D. ISAACS

### INTRODUCTION

It is generally agreed that psychiatric illnesses can fairly reliably be categorized into four broad groups—organic psychoses, functional psychoses, neuroses and personality disorders. Cases are assigned with greatest confidence to the first group and with least confidence to the fourth. The reliability of specific diagnoses within these classes is much less secure and decreases in the same order. The best recent reviews of the literature are by Kreitman (1961) and Foulds (1965).

The components of the diagnostic process include the skill and attitudes of the psychiatrist, his method of examination, the replies and attitudes of the patient, the amount of material obtained (whether the history is included, for example, or whether another informant is interviewed), the method of recording and coding data, the rules of classification and how these are applied. It is reasonable to suppose that if the contribution of each component could be fully standardized there would be no disagreement at all as to how a case should be categorized.

This paper describes a technique for structuring a clinical interview in order to assess the “present psychiatric state” of a patient. Such methods have usually only been applied to very limited sections of psychopathology and have then taken the form of check lists of symptoms. Instruments for measuring a wider spectrum of psychiatric symptomatology have been described by psychologists (Lorr *et al.*, 1963; and Wittenborn, 1955), and shown to be reliable. Check lists of items are rated on the basis of behaviour observed in unstandardized situations and the results are analysed statistically in order to produce empirical groupings. Such categoriza-

tions are equivalent to an empirical classification of skin colour. “Perceptual Distortion” (one of Lorr’s categories), for example, is a statistical concept and has no more relationship to, say, delirium tremens than pallor has to anaemia. A psychiatric diagnosis, on the other hand, is based, at least in theory, on the same concept of illness as is used in the rest of medicine. That is, the clinician selects elements from the history and examination which seem to suggest that the patient is suffering from a certain disease and, on the basis of this hypothesis, makes statements about aetiology, pathology or “natural history”. If the hypothesis is correct, these statements will be useful because they suggest means of prevention, treatment or management. At the very least the clinician should be able to give a prognosis. Thus the process of making a psychiatric diagnosis is fundamentally different to administering a Lorr scale and then applying various statistical procedures to obtain a categorization such as “Paranoid Projection”. It is, of course, possible that such statistical categories might eventually turn out to be as useful clinically as diagnoses: indeed, many of them do already look rather familiar to a clinician. In order to become so, it must be shown that they have at least as much unity of course and outcome as diagnostic groupings, and that they can as usefully indicate methods of treatment or management. None of the systems so far described has had any such claims made for it, and in any case it seems possible that such work would merely bring the psychologist full circle back to the diagnostic system which he had tried to leave behind (Katz and Lyerly, 1963).

Since the reliability and the validity of the Kraepelinian (or any other) diagnostic system

has not yet been tested in detail, and since it is used so widely and appears clinically to be so useful, it seems unwise to begin again from the beginning when constructing a categorization. The question therefore arises whether the reliability and precision and the potential for measurement achieved by instruments such as the Lorr scale cannot be transferred to the diagnostic process. If they could, categorization would become more reliable, and it would be possible to assign figures to many of the phenomena which psychiatrists commonly inquire about during the course of an interview. A diagnosis would, in fact, be based on a profile of scores measuring various aspects of psychiatric condition, obtained from the history and present state, from the patient and from other informants, and classified according to clear rules. The problem of validity could then be investigated with a greater degree of confidence.

An attempt to standardize a psychiatric interview has been described by Spitzer *et al.* (1964). In this technique, the questions to be asked by the examiner are laid down, as well as the way in which the replies are to be coded. Much use is made of non-specific probes such as, "Can you tell me more about that?", to draw out material from the subject, but there is no cross-examination on phenomenology. Many of the ratings are, in effect, dictated by the subject rather than by the examiner. The six gross scores used in the early paper were derived rationally (Behaviour, Speech, Psychosis, Non-psychosis, Sensorium, Insight) and were fairly reliable as between raters present at the same clinical interview (range 0.48 to 0.98). Reliability across two interviews with the same patient was not investigated. The technique has not been used to test the reliability of diagnoses and would not be suitable for this.

#### PROCEDURE

The work to be described in this and subsequent articles originated in a fairly successful attempt to construct, for research purposes, a simple descriptive categorization based on four leading symptoms of chronic schizophrenia—flatness of affect, poverty of speech, incoherence of speech and coherently expressed delusions

(Wing, 1961). This was extended in a second edition of the schedule, which was extensively tested at Cane Hill Hospital (J.B., A.I., J.W.). Sections dealing with neurotic symptoms were then added to make a third version, and a pilot study of reliability was undertaken which will be described later (J.C., P.G., J.W.). The fourth edition was further expanded and used to study patients with acute psychoses. The fifth edition was a reorganized version of the fourth and most of the material on reliability was collected using this schedule. The latest edition is being used by W.H.O. and the Bilateral Diagnostic Project.

There are several types of structuring in the interview. Basically, the schedule is a checklist of over 400 symptoms, which systematically covers all the phenomena likely to be considered during a present state examination and indicates how they are to be coded. Each of the symptoms is defined in greater or less detail. For most symptoms, a form of questioning is suggested, so that it should be possible to carry out the whole interview without deviating from the schedule at all. Finally there is a set of rules which governs how the material obtained during the interview is to be dealt with in order to make a preliminary diagnostic categorization.

Three main principles guide the conduct of the interview. The most important is that the interview is basically clinical. The schedule is *not* a questionnaire. It is laid out in question form to facilitate and partially standardize the interview, but each numbered question (together with suggested probes, and any other probes the examiner thinks necessary) represents a symptom and it is this symptom which is rated as present or absent. Usually the symptom is self-evident, but where there is any doubt it is specified within square brackets. In each case, what is wanted is the examiner's judgment of whether the symptom is present or not, based on the whole content of the interview. Many questions are followed by extra probes, in order to elicit further material about symptoms which the examiner suspects are present. Such probes are often suggested in brackets but these are not regarded as the only ones to be asked. For a symptom to be rated as present, the patient should describe it adequately in his own terms so that the examiner can recognize and rate it.

The suggested wording and order of questions are followed, unless there is good reason for varying them.

Thus the patient's response to the question, "Have you worried a lot during the past month?" is not rated, but used to begin a process of cross-examination which allows the examiner to make a judgment as to whether the patient worries or not, according to the definition laid down. The further questions suggested are, "What is it like when you worry? Do a lot of unpleasant thoughts go through your mind? When you get unpleasant worries, can you stop them by turning your attention to something else? Does this unpleasant worrying that you can't stop happen every day?" The examiner need not ask all these questions if the patient spontaneously gives the information. Only when the examiner is satisfied does he make the rating and, even then he may modify it on the basis of something that is said later in the interview. Altogether, nine symptoms are rated in the worrying section: each may be absent, present occasionally or in moderate degree, or present continuously or to a severely distressing extent during the month under review. Provision is also made for a rating of "not sure", which means that the examiner is not sure whether the symptom is present, even after adequate cross-questioning. The ratings are finally summed to give a score on "Worrying".

This technique of cross-examination is the one usually adopted by psychiatrists in a diagnostic interview and is completely acceptable to most patients. It differentiates the present approach from other standardized interviews, such as Spitzer's, in which, for example, only two questions are asked about worrying: "What kinds of things do you worry about?" and "How much do you worry?", and only one rating is made, necessarily accepting the *patient's* definition of a worry.

The cross-examination technique is used in each section. Symptoms are defined, questions are suggested, and on the basis of answers to these (plus further clarifying questions posed, if necessary, by the examiner) the ratings are made. In the order suggested in the sixth edition of the schedule, the sections are as follows: (1)

Attitudes, (2) Worrying, (3) Muscular Tension, (4) Anxiety, (5) Depression, (6) Irritability, (7) Hypomania, (8) Obsessional Symptoms, (9) Concentration and Interests, (10) Depersonalization, (11) Other Perceptual Disorders, (12) Hallucinations and Delusions, (13) Memory, (14) Severity and Insight, (15) Motor Abnormalities, (16) Blunting of Affect, (17) Restriction, Poverty and Incoherence of Speech. Several of these (particularly anxiety, depression hallucinations and delusions, and speech) are divided into subsections. Depression, for example, has the 8 subdivisions—*affective thought disorder, retardation, ideas of reference and relationships with others, self-opinion, depressive mood, somatic symptoms, signs of depression and depressive delusions and hallucinations.* In effect, therefore, the schedule is a miniature textbook of phenomenology. Since time and space are limited, much reliance has to be placed on the skill and knowledge of the examiner, and in this respect also the approach is a clinical one. Motor abnormalities are also assessed by a standard procedure.

The second principle is that the examiner concentrates his attention on the patient's experiences during the past month and at the time of the interview. Events which took place more than a month earlier are not considered. A companion schedule has been designed to take account of the "history" (Part II of the interview), but is not considered here.

The third principle makes use of the familiar clinical interview procedure whereby an examiner asks a number of screening questions in order to discover which areas it would be fruitful to investigate most thoroughly. Questioning on these selected topics is then more detailed and systematic. It is impossible to question a patient about every conceivable psychiatric symptom. Each section therefore begins with a few questions which are likely to demonstrate psychopathology in that area if it exists (unless the subject is concealing the truth). If there is no evidence from these answers, or from case-notes or other data, that the patient has any significant abnormality of that kind, the examiner does not proceed beyond a cut-off point, but places a tick in a "box" and continues with the screening questions in the next section.

In this way, a subject with none of the symptoms covered by the schedule can be interviewed very quickly. In spite of the rather general feeling that psychiatric patients always say "Yes" to leading questions, we have never yet conducted an interview where the cut-off procedure was not involved several times.

Instructions as to scoring, suggestions as to extra specific probes, and information about the definition of symptoms are written into the schedule. General probes are not specified.

The "present state" schedule is too large to present in its entirety (copies are available to research workers). It is organized in three parts. The first is complete in itself for non-psychotic patients. The second contains the briefing for a detailed examination of hallucinations and delusions and the third is a check-list of symptoms observable in the patient's speech and behaviour, including a standardized motor examination and a section on blunting of affect as observed in behaviour.

An interview with a patient who has very few symptoms could be completed in less than half an hour. If there are many neurotic and psychotic symptoms the interview may last one and a half hours. The average time is about three-quarters of an hour.

By way of example, three of the shorter sections from the sixth edition are presented in the Appendix.

#### PROVISIONAL CATEGORIZATION

When the interview is complete and all the ratings are made, the examiner is asked to consult a set of instructions for making a provisional categorization. This lays down a number of descriptive categories and lists the main symptoms necessary for each one. If necessary, two groups can be specified for one patient. This grouping is not, of course, a full diagnosis in the sense already discussed, which requires data from the history, from informants and from special investigations. However, if a categorization of "Present State" can be reliably made, then one important component of the diagnostic process has been standardized and there is reason to hope that the whole process can be brought under control.

A list of the provisional categories is given on page 504. Their names are always given in quotes, e.g. "Schizophrenia", to emphasize that they are not diagnoses.

#### LIMITATIONS OF THE APPROACH

It should be clearly understood that the schedule is designed to achieve a limited purpose. It has been used so far only by five experienced psychiatrists who have all been trained to Senior Registrar level at the Maudsley Hospital. Other psychiatrists may not obtain similar results. Enough information has now been incorporated into the schedule to allow other research workers to test the matter empirically.

The second limitation is that the results are given only for the "Present State" schedule (Part I of the complete interview). This allows only a provisional categorization to be made, *not* a diagnosis. It also excludes consideration of conditions which, by their nature, can only be diagnosed on the history, such as mental subnormality, personality disorder or alcoholic hallucinosis. In addition, few patients with dementias or organic psychoses have so far been interviewed: our present results relate mainly to the functional psychoses and neuroses.

Thirdly, our preoccupation up to now has been with the reliability of the provisional categories, not with their validity. The section and subsection scores are rationally based, in the sense that they represent the summed ratings of symptoms which clinically go together, but we have tested only the extent to which the profiles of scores associated with the provisional categories are the same for different raters. We have not attempted to discover whether the *right* category has been assigned, only whether two clinicians have each assigned the *same* category; whether there is a reasonable correlation between their scores and a reasonable concordance between their patterns of scores. Many other scores could be derived and the ratings could, of course, be statistically manipulated to discover whether other categories were empirically justified. In addition, we are not claiming that the size of the score necessarily represents severity of the condition, although

we suspect that there is a rough relationship at least, and that the scores could be used to measure change in one kind of severity.

A fourth limitation is that we are not concerned, in the work described here, with all the many other factors, social, psychological or biological, which clinicians must consider when dealing with patients. The approach is "static" rather than "dynamic". This is not because we think that "dynamic" material cannot be put into precise and measurable form, or that it is unimportant to try to quantify it. We insist only that such work should proceed by stages. The present work constitutes one small but necessary step.

Finally, we have not drawn random samples from the population of all those who have psychiatric conditions. At the present stage of testing this does not seem to be a major disadvantage, since a fairly wide range of conditions has been investigated and we are not attempting to make any generalizations about validity.

#### SERIES OF PATIENTS

During the course of the pilot work, several series of patients have been seen, and the results presented here concern eight of these in particular. The early work (January to March 1964) with the second edition of the schedule, in which a one in four sample of all patients in Cane Hill Hospital with an address in the former Metropolitan Borough of Camberwell were examined by two clinicians (J.B., A.I., or J.W.) and a smaller sample re-examined six weeks later, will not be presented, since major modifications were subsequently introduced into the interview procedure. The more recent editions of the schedule are, however, roughly comparable.

*Series A:* All 35 new patients referred to one out-patient clinic at the Maudsley Hospital during the period January to March 1965 were considered. Five patients were excluded after the first interview: one because he did not return for a second assessment, one homosexual man who complained of no psychiatric symptoms, one blind man, one patient who was admitted to hospital but discharged himself before the second interview could be completed, and one who was admitted from out-patients for appendectomy.

The remaining 30 patients were seen by a psychiatrist (J.C., P.G. or J.W.) who completed the third edition of the schedule after taking a history from a relative and administering the standard interview. This psychiatrist remained in clinical charge of the patient, who was asked to return one week later. On the second occasion a different psychiatrist saw the patient, this time only completing the schedule (having available to him the notes of the history taken by the first interviewer but not the results of his examination of "present state"). Half way through the exercise the order of examiners was reversed so that order effects could be studied.

*Series B:* Fifteen out-patients newly referred to the same clinic in November and December 1965 were examined according to the same procedure as Series A (J.C., P.G., or J.W.) but using the fifth edition of the schedule.

*Series C:* Twenty neurotic out-patients referred to the same clinic between September and November 1965 were examined by one psychiatrist using the fifth edition (J.C., P.G. or J.W.). The interviews were recorded on tape and independently rated by another clinician. The second rater had no information available apart from the present state interview; in particular, he knew nothing of the history, of any special investigations, or of what had been reported by other informants.

*Series D:* Twenty psychotic out-patients were examined in a similar way to Series C (J.B. and J.W.) and the interviews tape-recorded and subsequently rated independently.

*Series E:* Fifty-three long-stay in-patients in Stone House Hospital, all of whom had been given a hospital diagnosis of schizophrenia, were examined by two clinicians (J.B. and J.W.) taking turns to conduct the interview. The fifth edition of the schedule was used.

*Series F:* Ten day-patients at the Maudsley Day Hospital were seen by two psychiatrists (J.B. and J.W.) taking turns to conduct the examination. The fifth edition of the schedule was used.

*Series G:* Thirteen acutely-ill psychotic in-patients were seen within a few days of admission by two psychiatrists (J.B. and A.I.) taking turns to conduct the interview. The fourth edition of the schedule was used.

*Series H:* Eleven acutely-ill psychotic in-patients were seen within a week of admission, on two separate occasions by different psychiatrists (J.B. and A.I.). The fifth edition of the schedule was used.

#### RESULTS

##### 1. Provisional Categorization

Of the total 172 cases, the main categories allocated by the first clinician were as follows (the number of cases placed into exactly the same category by the second clinician is shown in brackets):

"Schizophrenia (Possible, Probable or Residual Condition)"	75 (69)
"Psychotic Depression (Possible or Probable)"	15 (12)
"Mania (Possible or Probable)"	4 (4)
"Other Functional Psychoses"	2 (2)
"Dementia (Possible or Probable)"	2 (2)
"Non-psychotic Depression (Retarded)"	8 (6)
"Non-psychotic Depression (Other)"	35 (28)
"Anxiety State (General, Situational or Specific)"	17 (13)
"Obsessional Neurosis (Possible or Probable)"	2 (2)
"Personality Disorder"	6 (3)
"No Abnormality in Present State"	6 (3)

The results mainly apply, therefore, to the functional psychoses and affective neuroses. The categories independently allocated by two clinicians to the patients in the various series were compared, and sorted into three groups:

- Complete agreement within the categories shown above.
- Partial agreement—the two clinicians agreed on one category but disagreed on a second.
- Disagreement—the two clinicians allocated completely different categories.

The results are summarized in Table I. There is very little difference between the series; in particular, the results are similar irrespective of the technique of making the second categorization—whether from observation of the interview, or from a tape-recording, or from a second interview on a different occasion by a

different psychiatrist. Overall, there was complete agreement in 84 per cent. of cases, partial agreement in 7 per cent. and disagreement in 9 per cent.

Ten of the 28 cases in which there was disagreement or only partial agreement involved the category "Personality Disorder".

Another 9 cases involved a discrepancy in allocation of one of the non-psychotic affective disorders—"Retarded Depression", "Other Depression" and "Anxiety State"—some of these may have been due to changes in the patient's condition. Three disagreements concerned schizo-affective conditions, and another three turned on whether the delusions present were sufficiently marked for the patient to be categorized as "Psychotic Depression". One of the remaining three discrepancies ("Possible Schizophrenia" versus "Other Functional Psychoses") was permissible in a provisional categorization. The other two were straightforward disagreements ("No Abnormality" versus "Possible Dementia", and "Probable Schizophrenia" versus "Probable Mania") in which the profiles of scores made it clear that one clinician had not followed the rules of classification laid down.

## 2. Reliability of Section Scores: Non-psychotic Symptoms

Five main scores were used to assess the reliability of the interviewing procedure in measuring non-psychotic symptomatology: "Anxiety", "Depression", "Other Specific

TABLE I  
*Degree of Agreement on Categorization Between Two Clinicians*

	Series A and B Two interviews (by different psychiatrists)	Series C and D One interview (tape-recorded)	Series E One interview (Observer sitting in)	Series F and G One interview (Observer sitting in)	Series H Two interviews (by different psychiatrists)	Total	%
Complete agreement	34	34	47	19	10	144	83.7
Partial agreement	6	3	1	1	1	12	7.0
Disagreement	5	3	5	3	—	16	9.3
Total	45	40	53	23	11	172	100.0

Symptoms", "Worrying and Muscular Tension" and "Other Associated Symptoms". "Other Specific Symptoms" included the summed scores on ideas of reference, depersonalization, obsessional phenomena and hypomania, none of which was sufficiently often rated to allow the calculation of separate correlation coefficients. "Other Associated Symptoms" included irritability, lack of concentration and lack of interest. Anxiety included "free floating" anxiety, panic attacks, situational anxiety, avoidance of anxiety provoking situations, and autonomic symptoms. Anxiety rated "on examination" was dealt with separately since it could not be assessed from a tape-recording. Depression included affective thought disorder, self-opinion, retardation, depressive mood and somatic symptoms such as anorexia, loss of weight and interrupted sleep. Depression rated "on examination" was treated separately.

The product-moment correlation coefficients for the three series A, B and C are shown in Table II.

Several points emerge from Table II. Firstly, a comparison between Series A and B (all patients seen on two occasions by different psychiatrists) shows an increase in reliability coefficients on all scores except for "worrying and tension". This presumably reflects the increasing experience of the examiners in using the schedule plus the success of modifications introduced into the fifth edition, which was easier to use than the third.

Secondly, the reliability of the scores derived from the fifth edition is satisfactory, except for one section. The low correlation (0.53) on anxiety is due to large discrepancies in the scores for three patients, two of whom had much lower scores on the second occasion while the third had a much higher score. (The coefficient for the other 12 patients is 0.87). When the records and the ratings of the three patients were examined, it became clear that they did in fact give different accounts of their symptoms to the second interviewer—this was confirmed by independent ratings of the audiotapes of both interviews. In one case, the patient improved very markedly after the first interview and it is clear that her account was affected by this improvement. In the other two cases, it did not appear that the patients had, in fact, really changed very much, but nevertheless they presented their anxiety symptoms quite differently on the two occasions. Something similar happened in three other cases—two affecting the "Worrying and Tension" score and one affecting the "Depression" score. Technically, of course, even if the patient did improve during the week between interviews, the scores should not be much affected since the whole of the previous month was covered. In practice, however, the way the patient feels at the time, and his attitude to the examination, often does affect the way he presents the development of his symptoms. These are matters which it is very difficult to standardize. The problem seems to

TABLE II  
*Reliability of Non-Psychotic Section Scores*

	Series A Two interviews Third edition N=30	Series B Two interviews Fifth edition N=15	Series C One interview (tape-recorded) Fifth edition N=20
"Anxiety" .. .. .	.484	.526	.939
"Depression" .. .. .	.719	.896	.883
"Other specific symptoms"	.649	.898	.932
"Worrying and Tension"	.810	.609	.909
"Other associated symptoms"	.416	.843	.952
"All non-psychotic symptoms"	.790	.797	.969
Anxiety on examination .. .. .	.515	.772	—
Depression on examination .. .. .	.503	.994	—

affect anxiety, worrying and tension particularly, and the other scores to a much lesser degree.

The third point to note is that the reliability coefficients are *all* high when based on two clinicians' ratings of one interview. The correlation for anxiety, for example becomes 0.94, while that for worrying and tension is 0.91.

### 3. Reliability of Section Scores: Psychotic Symptoms

Nine main scores were used to assess the reliability of the interviewing procedure in measuring psychotic symptomatology: "Auditory Hallucinations", "Delusional Experience of Disordered Thought", "Delusions of Reference, Persecution and Misinterpretation", "Other Delusions (not including depressive delusions)", "Motor Abnormalities", "Blunting of Affect", "Poverty of Content and Restriction of Quantity of Speech", "Incoherence of Speech", and "Manic Type of Speech". Each of these is a complex score—for an example of the sort of items which composed a score, see the Appendix.

The eleven patients in series H were seen on two occasions at an interval of a few days, each time by a different clinician using the fifth edition of the schedule. They were often seen in

different hospitals and several were unable to co-operate in the full interview on one or other occasion. Correlations could only be calculated for four of the main scores (range 0.2 to 0.45). The correlation coefficients for two other groups are shown in Table III.

The coefficients based on one interview are high (0.62–0.97) in both groups shown in Table III whether the second set of ratings is based on tape recordings (Series D) or observation of the interview (Series E).

### 4. Profiles of Scores

#### (a) Very Low Scores

Eight patients out of 172 were classified by one or other clinician as having "No Psychiatric Abnormality"—in three of the eight, both psychiatrists agreed on this category. In three more cases, both clinicians categorized the patient as "Personality Disorder".

All these eleven patients had very low morbidity scores on "present state"—the Grand Total of non-psychotic and psychotic scores did not exceed 12. These patients were, in fact, almost symptom-free on "present state".

TABLE III  
Reliability of Psychotic Section Scores

	Series D One interview (tape-recorded) Fifth edition N=20	Series E One interview (Observer sitting in) Fifth edition N=53
"Auditory hallucinations" .. .. .	.. .920	..919
"Delusional experience of disordered thought" ..	.. .766	..945
"Delusions of persecution, etc." .. .. .	.. .943	..883
"Other delusions" .. .. .	.. .894	..942
Total "delusions and hallucinations" .. .. .	.. .959	..933
"Motor abnormalities" .. .. .	.. —*	..930
"Blunting of affect" .. .. .	.. .822	..862
"Poverty of speech" .. .. .	.. —*	..867
"Incoherence of speech" .. .. .	.. —*	..620
"Manic type of speech" .. .. .	.. —*	.. —*
Total "behavioural and speech abnormalities" ..	.. .875	..945
All "non-psychotic symptoms" .. .. .	.. .966	..952

\* Insufficient positive readings to allow calculation of coefficient



In addition, eight other patients out of the 172 had Grand Total scores of 0-12, on both sets of ratings. Two were suffering from early dementia and scored only on the "Memory Disturbance" section which is not considered in this paper. Two others were schizophrenic patients attending the outpatient department, who had only very minor symptoms, but sufficient to allow the diagnosis to be made on "present state". The other four were long-stay patients who were almost symptom-free and should perhaps have been classified as having no abnormality.

(b) *Profiles of Affective Syndromes*

Although the section scores derived from interviews based on the third, fourth and fifth editions of the schedule are not exactly comparable, the composition of items and range of scores are similar enough to allow the summation of results in order to examine the reliability of profiles of scores. The main category allocated by the first examiner was used to define four groups of patients with "affective" syndromes: "Psychotic Depression" (N15), "Retarded Non-Psychotic Depression" (N8), "Non-Psychotic Depression, Other" (N35) and "Anxiety State" (N17). The mean section scores on "Worrying and tension", "Anxiety", "Depression", "Other Specific Symptoms" and "Other Associated Symptoms" were then calculated for each of the four groups. The profiles of mean section-scores are shown in Fig. 1, separately for the examiner and second clinician. (The score-profiles were also calculated separately for the different editions of the schedule but no discrepancy was observed).

The profiles of mean scores are almost identical for the two clinicians although the scores are lower on the second occasion of rating in the group of "Anxiety States". The three "depressive" syndromes have somewhat different profiles but the profile of "Anxiety States" shows the greatest divergence from the others.

The correlation between "Anxiety" and "Depression" scores, using 75 cases about which there was agreement on the categorization of "Affective Syndrome", is 0.07 in the first interviewer's material and 0.16 in the second.

(c) *Profiles of Schizophrenic Syndromes*

A similar exercise, using arbitrary descriptive syndromes, was carried out on the data collected from interviews with schizophrenic patients. Only those were included who scored more than 12 on "Delusions and Hallucinations", or more than 12 on "Other Psychotic Symptoms" (such as blunting of affect, motor or speech abnormalities), and for whom both clinicians made a categorization of "Probable or Possible Schizophrenia".

On inspection, there was an obvious split between paranoid and non-paranoid patients, the former all having a score on delusions and hallucinations above 12. There was no obvious way in which the non-paranoid patients could be further subdivided except by providing numerous infrequent syndromes which could not be dealt with statistically. However, there were three clear syndromes among the paranoid patients which, though they do not correspond to any clinical categorization, allow investigation of the reliability of score profiles. All patients with a score of 2 or more on "Delusional Experience of Disordered Thought", also had moderate or high scores on "Other Delusions" and "Auditory Hallucinations". Among the remaining paranoid patients, there was a group with a score of 2 or more on "Auditory Hallucinations". All these patients also had a moderate or high score on "Other Delusions". Finally, the remaining patients had "Other Delusions" only.

Thus four groups were formed and mean section scores were calculated from each clinician's ratings. These are shown in Fig. 2.

The profiles obtained from the two sets of ratings are again almost identical. It seems reasonable to assume that other groupings, whether rationally or empirically derived, would be equally reliable.

(d) *Psychotic versus Non-psychotic Profiles*

Profiles of the 10 scores used in Figs. 1 and 2 were constructed for all patients but will not be presented here since they illustrate only that the raters observed the rule to use one of the psychotic categories only when the patient was deluded, hallucinated, had severe speech dis-

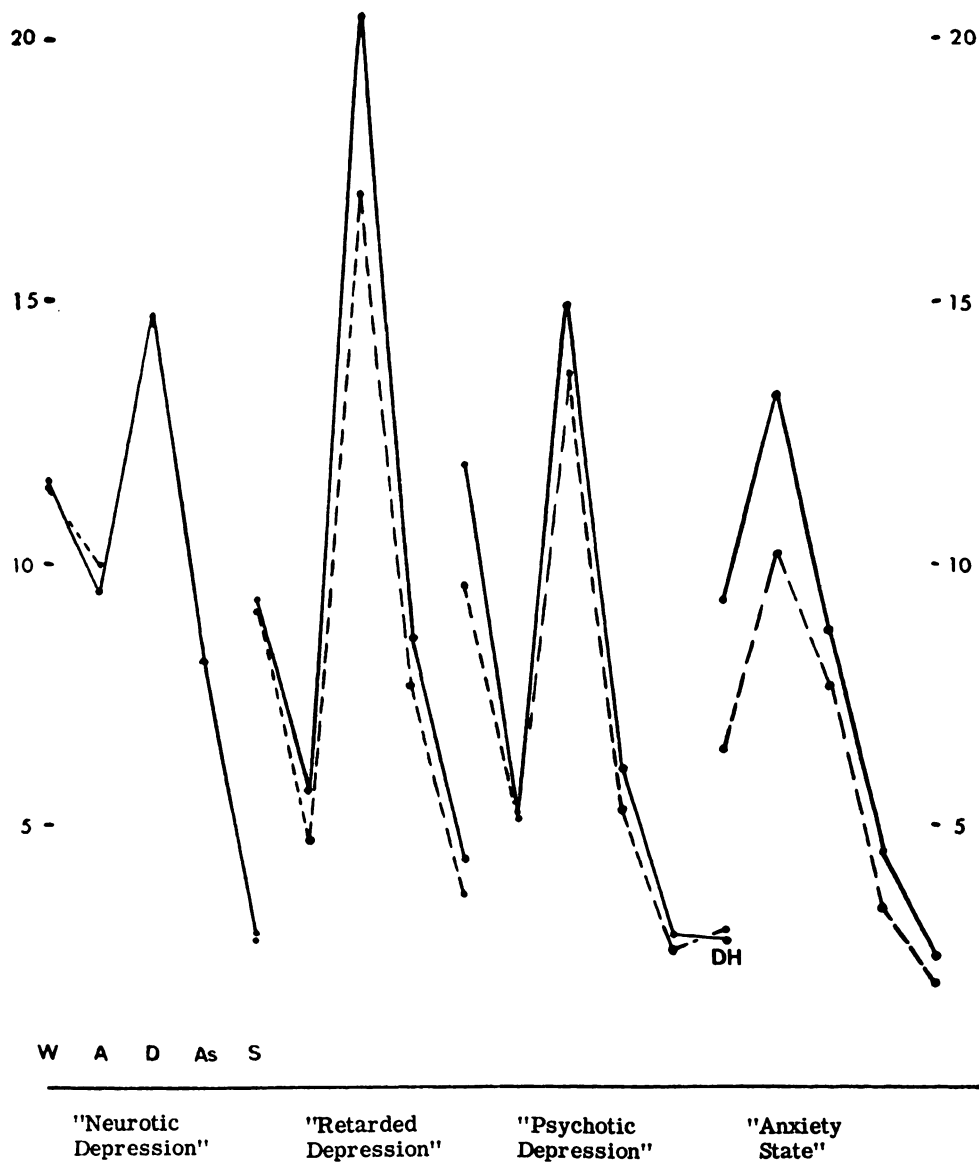


FIG. 1.—Profiles of mean scores obtained by two raters on four affective syndromes.

———— Interviewer  
 - - - - - Second Rater

Scores:  
 W Worrying and tension  
 A Anxiety  
 D Depression  
 As Other associated symptoms  
 S Other specific symptoms  
 DH Depressive delusions and hallucinations

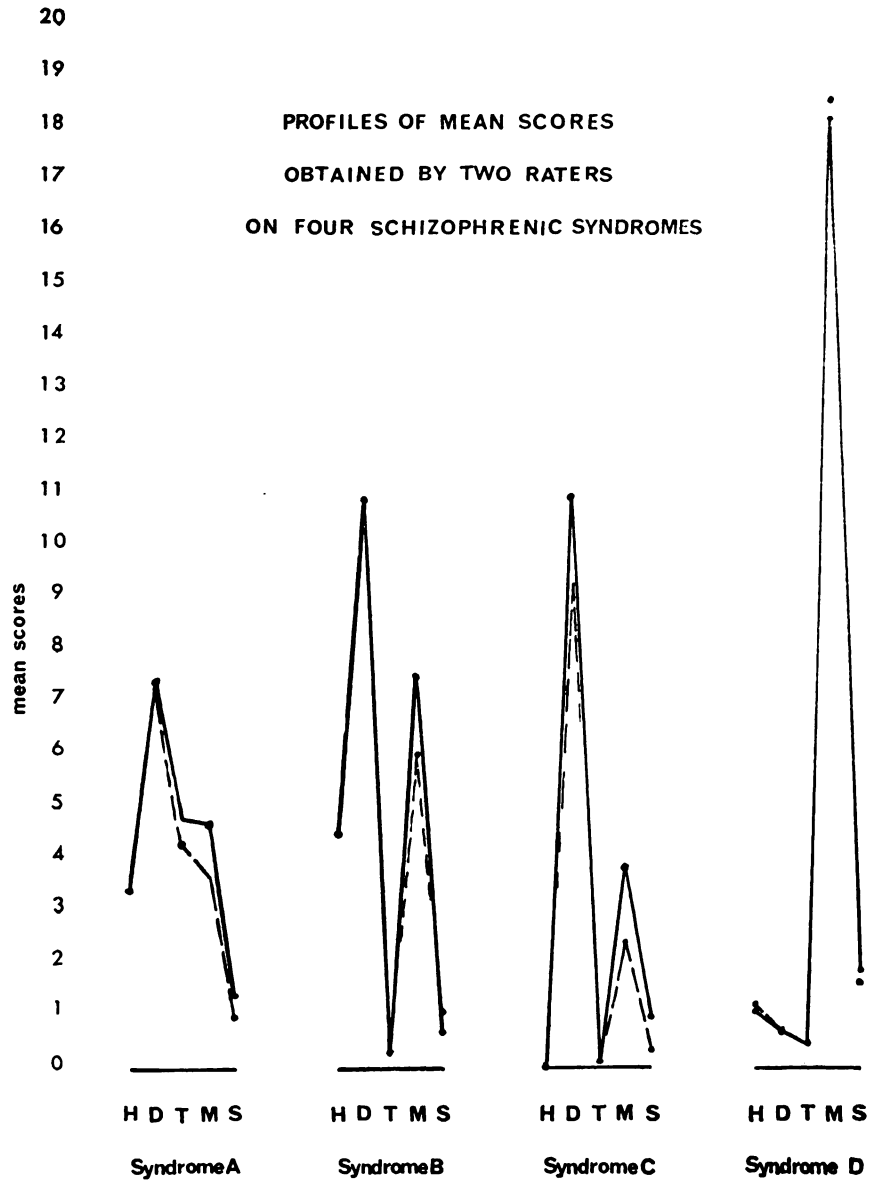


FIG. 2.

———— Interviewer  
- - - - - Second rater

H Auditory hallucinations  
D Other delusions  
T Subjectively experienced interference with thought  
M Motor symptoms, blunting of affect and poverty of speech  
S Incoherence and manic speech

order, catatonic motor abnormalities or disturbance of consciousness. The six discrepancies concerning borderline syndromes have been described in section 1.

#### 5. Reliability of Rating Individual Symptoms

The sixth edition of the schedule contains over 400 possible ratings. In order to describe, without undue tedium, the reliability of rating individual symptoms, only two subsections will be considered in detail; one from the section on anxiety which has low reliability as between two clinicians conducting separate interviews, and one from the section on depression which has a high reliability under the same conditions. Fifteen patients in series B will be used for this exercise.

##### (a) Autonomic accompaniments of anxiety

Whenever the clinician proceeds beyond the cut-off point in the section on anxiety, he is required to cross-examine the patient about situations in which he feels anxious or in which he has had a panic attack during the past month (such as buses, lifts or crowded shops), what situations he has avoided in order to prevent anxiety and what autonomic symptoms he has experienced when anxious. Ten autonomic symptoms are rated as present or absent and the degree of agreement is shown in Table IV.

Out of 150 possible ratings there was disagreement on 43, or 28.7 per cent. Both raters agreed that a symptom was absent in 52 per cent. of cases, and that it was present in 19.3 per cent. This is not a very high proportion of agreement if patients who did not have the symptom according to either rater are excluded. The reliability of several symptoms is clearly unsatisfactory (for example, blushing) and the subsection has been modified in the sixth edition. The correlation between the subtotals was 0.53.

##### (b) Somatic symptoms of depression

The degree of agreement on eight somatic symptoms of depression is shown in Table V. The two raters disagreed in only 16 per cent. of cases. Symptoms like constipation are clearly difficult to rate reliably in this kind of interview. The correlation between scores was 0.63.

#### 6. Influence of Individual Clinicians

Each of the five clinicians had his own style of rating, although discrepancies in categorization were distributed proportionately between pairs. The design of the present work was not specially directed towards examination of rater bias but it is fairly easy to demonstrate in series A, in which three clinicians conducted the interviews. Rater 1 tended consistently to rate more symptoms present than Raters 2 and 3, and Rater 2 rated more than Rater 3.

TABLE IV  
Agreement on Whether Certain Autonomic Symptoms are Present When Patient is Anxious (15 out-patients)

Autonomic Symptom	Agreement that symptom is present	Agreement that symptom is not present	Disagreement
Palpitations .. .. .	2	8	5
Dizzy, faint or giddy .. .. .	2	11	2
Shaky or trembling .. .. .	5	3	7
Butterflies in stomach .. .. .	6	5	4
Sickness or feeling of sickness .. .. .	4	8	3
Hot or cold feelings .. .. .	1	11	3
Sweating a lot .. .. .	3	6	6
Difficulty getting breath .. .. .	3	8	4
Blushing .. .. .	1	10	4
Dry mouth .. .. .	2	8	5
Total .. .. .	29	78	43

TABLE V  
*Agreement on Whether Certain Somatic Symptoms of Depression are Present (15 out-patients)*

Symptom	Agreement that symptom is present	Agreement that symptom is not present	Disagreement
Subjective restriction of energy	9	3	3
Prolonged time in bed, less housework, etc., because of lack of energy	3	8	4
Poor appetite	6	7	2
Loss of weight (3 months)	2	12	1
Interrupted sleep	7	5	3
Early morning waking	2	11	2
Constipation	—	9	6
Loss of libido	2	10	3
Total	31	65	24

7. *Do Leading Questions Produce False Positive Ratings?*

We have no doubt that, in this kind of interview, there is very little set to respond affirmatively to all questions. In any case, the fact that a patient says "yes" to a question about a symptom does *not* mean that the examiner must mark that symptom as present. He makes a judgment on all the evidence available, including any obvious response set. An opportunity to test the matter arose when a patient in the Day Hospital (diagnosis, "hysterical personality") was reputed to say that she had any symptom she was asked about. In fact, when interviewed jointly by two clinicians, she showed considerable discrimination about which symptoms she agreed to, although the overdramatic way in which she described them made it clear why the diagnosis had been made. It is obvious from Table IV that anxious patients do not say that they have every autonomic symptom they are asked about, although all the symptoms are well known as accompanying anxiety.

DISCUSSION

The flexibility of this approach, the incorporation of detailed cross-examination which allows changes in the order and wording of questions according to the way the interview is going, the freedom of the clinician to pursue some lines of

enquiry while cutting off others, the fact that the examiner and not the patient makes the judgment as to whether a symptom is present, do not seriously impair the reliability of the procedure. When two raters assess the same interview there is a very high degree of agreement on provisional categorization and very high correlation coefficients between scores representing areas of symptomatology. When there are two interviews at an interval of several days, the provisional categorizations made by the two clinicians are still very similar, but the correlations on section scores fall somewhat. However, the only non-psychotic score which does not achieve a satisfactory reliability under these conditions is that for anxiety, due to the fact that a few patients change a great deal between two interviews, either because their symptoms improve or because they give a different account of them. Thus the interview procedure appears reasonably reliable, but there will always be a problem in using two interviews to assess reliability, since the first interview may actually have an effect on the results of the second quite apart from any changes that might take place. The acutely psychotic patients who were seen twice changed very markedly and no test of reliability was possible, although the provisional diagnostic category did not change. A much larger series would be necessary to determine whether this change is a sampling effect or characteristic of most acute psychoses.

It probably depends on how soon the patient is seen after admission. The reliability of rating psychotic symptoms, based on one interview only, was satisfactory.

Individual symptom ratings show a variable reliability but, as in most scales, the reliability of the total scores is greater than that of the single items would lead one to expect.

Profiles of scores, arbitrarily or rationally derived, have a high reliability.

No statistical clustering, for example by factor analysis, was carried out, since this was not the purpose of the work, and in any case we have serious reservations about the use of any factors that might be derived. However, such an exercise is being undertaken in order to compare the clusters with the rationally derived scores.

Although no general claims can be made, one limited component of the diagnostic process does seem to have been brought under a fair degree of control, which gives ground for supposing that other components can similarly be standardized. However, much further work remains to be done. Firstly, a representative sample of patients must be studied and a detailed analysis of score patterns made in order to establish differences between the various categories. Secondly, the history needs to be standardized and its information incorporated into a full diagnostic procedure, the reliability of which should be tested. Thirdly, it is important to show whether the scores decrease as the

patient improves, and vice versa, because the value of the schedule would then be greatly enhanced. Finally, a brief form of the schedule suitable as a screening instrument for general populations needs testing, in order to discover whether it will indicate individuals who require fuller examination.

The problem of validity has not been discussed in this paper, but once the reliability of syndromes and scores is established it will become possible to investigate differences in outcome, in response to treatment, in social concomitants and other relevant variables. The uses of the procedure even as it stands are fairly clear. In particular, the method would be useful for research work which is based on clinical categories. It might also be useful for teaching phenomenology and clinical examination and for measuring change in morbidity. All these matters deserve further examination.

#### SUMMARY

A structured "present state" interview has been developed and tested. As used by five trained interviewers the provisional categorizations made are reliable and so are most of the clinical scores. The advantage of the procedure is that it is based on clinical practice and experience, but introduces a degree of standardization and precision which suggests that more of the diagnostic process might be brought under control with obvious benefit to clinical practice and research.

#### APPENDIX

##### THREE SECTIONS OF THE INTERVIEW SCHEDULE

##### SECTION 8. OBSSIONAL SYMPTOMS

[Include only events experienced as occurring against the subject's conscious resistance. Do not include the avoidance of anxiety-provoking situations].

- |     |   |   |   |                          |    |    |
|-----|---|---|---|--------------------------|----|----|
| 8:1 | Did you find that you have to keep on checking things that you know you have already done, (like gas taps, doors, things at work) during the past month? [Do not include lapse of concentration, or poor memory, so that subject is not sure whether door was closed, etc.] | o | 1 | ?                        | NR | NA |
| 8:2 | Did you find it difficult to be <i>sure</i> that you had completed a task <i>really</i> properly (so that you were satisfied it was really done), during the past month?  | o | 1 | ?                        | NR | NA |
| 8:3 | Do you spend a lot of time on personal cleanliness? (Like washing over and over when you know you are clean, or doing a lot of ironing when you know the clothes are all right).  | o | 1 | ?                        | NR | NA |
|     | CUT OFF IF NO EVIDENCE THAT REMAINING QUESTIONS ARE RELEVANT AND PROCEED TO SECTION 9   |   |   | <input type="checkbox"/> |    |    |
| 8:4 | Do you get worried about contamination by germs, or things like that? (What is it like? When did you last worry about it?)  | o | 1 | ?                        | NR | NA |

- 8:5 Do you find that unpleasant or frightening thoughts or words come into your head and won't go away even when you try to get rid of them? (Can you give me an example?) [Exclude delusions, hallucinations, worries, anxieties. Rate thoughts of knives, killing someone else, etc.] 0 1 ? NR NA
- 8:6 Do you have rituals, which you know are silly, but which you have to carry out all the same? Like saying "abracadabra" 10 times before crossing the road? [Rate only clearly obsessional phenomena.] 0 1 ? NR NA
- 8:7 Do you *have* to do certain things a particular number of times or in a special order or manner? 0 1 ? NR NA
- 8:8 [Rate obsessional rumination, e.g. prolonged preoccupation with the multiplication of numbers or with some other purposeless topic, against conscious resistance] 0 1 ? NR NA
- 8:9 [Rate other obsessional phenomena. Specify \_\_\_\_\_] 0 1 ? NR NA
- 8:10 [Overall rating of obsessional phenomena] 0 1 (mild) 2 (moderate) 3 (severe) NR NA
- 8:11, 12 Score on obsessional symptoms \_\_\_\_\_ (8:1-10, max. 12)

12C. DELUSIONS OF BODILY CONTROL

[Do not include feeling that life is planned and directed by fate, or that the future is present already in embryo, etc. Do not include voices giving the patient instructions, even if he feels he has to obey them. The symptom is that the subject's will is replaced by that of some other agency. It is rare.]

Do you feel that you are under the control of some other agency? In what way?

CUT OFF IF NO EVIDENCE THAT THE REMAINING QUESTIONS ARE RELEVANT AND PROCEED TO SECTION 12D

- 12C1 [If Yes] As though you were an automaton or robot [zombie, marionette, puppet, other . . .] Who is doing this? 0 1 ? NR NA
- 12C2 Does (this agency, person) actually speak with your voice? 0 1 ? NR NA
- 12C3 Or control your movements and actions? 0 1 ? NR NA
- 12C4 Or possess you so that you are a different person? 0 1 ? NR NA
- 12C5 What about your handwriting, is that controlled? 0 1 ? NR NA
- 12C6 Is there any other way in which your body is controlled? [Specify How does (person, agency) manage to do this? What is the mechanism?] 0 1 ? NR NA
- 12C7, 8 [Score \_\_\_\_\_ (12C1-6)]

17B, C. SOCIAL SPEECH

[Speech which fits within the social context of the interview, i.e. it is addressed to the examiner, or consists of qualifying comments on something that has been said which was part of the conversation. The fact that a comment is incoherent or irrelevant does not mean that it is non-social. If in doubt, count speech as social rather than non-social]

17B. POVERTY OF CONTENT, OR RESTRICTION OF QUANTITY, OF SPEECH

RESTRICTION OF QUANTITY (B1-11)

- 17B1 Patient is quite mute: utters no more than half a dozen recognizable words in answer to questions throughout the interview. Do not include talking out loud to self 0 1 ? NA
- 17B2 Patient is almost mute: says no more than 20 words in introductory section in answer to questions and pro rata in rest of interview. Do not include talking out loud to self 0 1 ? NA
- 17B3 Patient frequently fails to answer: questions have to be repeated. Do not rate if patient gives inaudible or incomprehensible replies 0 1 ? NA
- 17B4 Answers are restricted to the minimum necessary—Yes, No, Don't know, etc. No unnecessary words used. e.g. Q. "What work do you do?" A. "Sweep". Q. "Anything else?" A. "Scrub floor". May be occasional extra sentence or additional comment.\* 0 1 ? NA

514 PROCEDURE FOR MEASURING AND CLASSIFYING "PRESENT PSYCHIATRIC STATE"

- 17B5 No, or almost no extra sentences.\* 0 1 ? NA
  - 17B6 No, or almost no unprompted additional comments.\* 0 1 ? NA
  - 17B7 Repetition of what examiner says (echolalia)—several examples present and noted in the record 0 1 ? NA
  - 17B8 Frequent repetition of one or more stereotyped phrases, e.g. "I'm not coming the old Isaacs", or "It doesn't prove power to any patients". Do not include frequently reverting to certain themes or topics 0 1 ? NA
  - 17B9 Overall rating of restriction of quantity taking answers to B1-8 into account 0 1 2 3 ? NA
- Scoring: 1 Definite examples of one or more items present, but not a marked feature of the interview  
 2 Interview very restricted  
 3 Patient is mute or almost mute

17B10, 11 Score on restriction of speech \_\_\_\_\_ (17B1-9, max. 11)

\*Definition of "extra sentence" and "additional comment":

Q. "What work do you do?" A. "I'm a centre lathe turner. I work at Simpson's". Pause, then, without prompting: "It's a decent sort of a job really".

"I work at Simpson's," is an extra sentence.

"It's a decent sort of a job really", is an unprompted additional comment.

The sentence, "I'm a centre lathe turner" would have been sufficient to answer the question. The other two sentences give additional information and keep the conversation going.

In making these ratings, consider first the neutral part of the interview (which has been recorded verbatim) and proceed by exclusion. Rate best performance during the interview.

EXAMPLE 17B14

Vague wandering:

Q. Can you think quite clearly? Or do your thoughts seem to get muddled?

A. Yes, . . . er. When I, er, pull myself—when, er, Oh, when it's er, when there is an obstacle, I say, Ah! There's two or three meanings to that obstacle. That obstacle's got to be got out of before I can make another move.

Q. There are two or three meanings to an obstacle?

A. Yes, got two or three meanings to it. It's just the same with a woman, when they say, No, then come playing about with me, I say, Ah! turn my No into Yes and then I'm caught. So I said No to you. See what I mean? Now you turn my No into Yes, well you jolly well got it, now you can jolly well put up with it, shut your mouth up and done with it. I said No to you. See what I mean? Well, that is, er, that is where a lot of people—it doesn't evenly work out between the two parties. Does it? She's already got herself more depended to, er, jump on you and you're caught.

Q. Jump on you?

A. You know, if it's, if anything went wrong. She can blame you and she gets away with it. See, it's er, really speaking, a sense of obedience should be, is to be split up in give and take, to be balanced with the other 9 points of the law. That's how we got, how we should come to an understanding with each other, in marriage, see? To a certain extent—

Q. A sense of obedience is to be split up into give and take?

A. It's, er you know what I mean—to a certain extent, that you get an understanding clearly with the other 9 points of the law. Because those other 9 points of the law makes up the subject. They've all got to be brought—it's the machinery of putting everything in to bring about the results that's of advantage to each other. Then, of course, in the end, it might be more advantage to the one than it is to the other. You've got all that to keep a level mind, between you.

ACKNOWLEDGMENTS

We are most grateful to clinicians at the Bethlem Royal and Maudsley Hospitals, Cane Hill Hospital, St. Francis' Hospital and Stone House Hospital who allowed us to interview their patients. Dr. James Willis, in particular, was most helpful.

KATZ, M. M., and LYERLEY, S. B. (1963). "Methods for measuring adjustment and social behaviour in the community. I. Rationale, description, discriminative validity and scale development." *Psychol. Rep.*, 13, 503.

KREITMAN, N. (1961). "The reliability of psychiatric diagnosis." *J. ment. Sci.*, 107, 876.

REFERENCES

FOULDS, G. A. (1965). *Personality and Personal Illness*. London: Tavistock Publications.

LORR, M., KLETT, C. J., and McNAIR, D. M. (1963). *Syndromes of Psychosis*. Oxford: Pergamon.



- SPITZER, R. L., FLEISS, J. L., BURDOCK, E. I., and HARDESTY, A. S. (1964). "The mental status schedule: Rationale, reliability and validity." *Comprehensive Psychiat.*, 5, 384.
- WING, J. K. (1961). "A simple and reliable subclassification of chronic schizophrenia." *J. ment. Sci.*, 107, 862.
- WITTENBORN, J. R. (1955). *Wittenborn Psychiatric Rating Scales*. New York: Psychological Corporation.

J. K. Wing, M.D., Ph.D., D.P.M.

J. L. T. Birley, B.M., B.Ch., M.R.C.P., D.P.M.

*M.R.C. Social Psychiatry Research Unit, The Maudsley Hospital, London, S.E.5*

J. E. Cooper, B.M., B.Ch., M.R.C.P., D.P.M., *Bilateral Diagnostic Study, Institute of Psychiatry, The Maudsley Hospital, London, S.E.5*

P. Graham, M.B., B.Chir., M.R.C.P., D.P.M., *Institute of Psychiatry, The Maudsley Hospital, London, S.E.5*

A. D. Isaacs, M.R.C.P., D.P.M., *Consultant Psychiatrist, Cane Hill Hospital, Coulsdon, Surrey*

*(Received 11 May, 1966)*