

THE RELIABILITY OF PSYCHIATRIC, AND THE VALIDITY OF PSYCHOLOGICAL, DIAGNOSES

By

G. A. FOULDS, M.A., Ph.D.
Runwell Hospital, Wickford, Essex

THE PROBLEM

Criterion of Value

IN determining the diagnostic value of psychological tests the usual external criterion is the psychiatric diagnosis. This criterion has suffered a steady decline in prestige among both psychiatrists and psychologists. The following remarks by Noyes (1953) would be endorsed by many psychiatrists: "The principal value of classification is not in the categorizing of a disease entity but in quickly eliminating those considerations which will be least useful in understanding the patient and in directing attention to those which are likely to be relevant. Except in organic disorders a classificatory diagnosis is less important than a psychodynamic study of the personality. The psychiatrist should be interested in processes, not in labels . . . we should endeavour not so much to fit the symptoms into a classificatory scheme as to understand the sick person in terms of his life experience."

This is primarily the point of view of the clinician of psychoanalytic or psychobiological leanings and plenty of time to spare and illustrates the vast gap that too often exists between clinician and research worker.

When the psychiatrist has made psychodynamic studies of several personalities, he must, if he is to formulate a cohesive, teachable body of knowledge, compare one with another. With comparison comes classification; without it comes a tendency to feel things in one's bones and such rheumatoid sensitivity is notoriously difficult to convey to the tyro.

Psychologists too attack the present psychiatric nosology.

Roe (1949) writes: "I suggest that much of this research (the use of various psychological devices for making psychiatric diagnoses) is not only a waste of time, and a perpetuation of errors, but is actually preventing advance in the field. There are many reasons why this is true, but one of the most potent is that it involves clinging to a classification which has long since been outlived. I submit that using techniques which are not too precisely validated at all, to place patients in psychiatric categories, the inadequacy of which is admitted by all concerned, is a treadmill procedure guaranteed to keep us moving in circles."

Many psychiatrists, as well as psychologists, would agree with this estimate of psychiatric nosology and psychological practice and yet the curious fact is that there seems to be little or no evidence for this belief. Whether or not such diagnoses are useful, no other classificatory system has been advanced and supported by experimental evidence.

Roe states that the most potent reason for not using tests for diagnostic purposes is that psychiatric classification has long since been outlived. Apart from the fact that classification was never viable and could not, therefore, be outlived, the statement, even with the substitution of the word outmoded, is palpably false. Such classification is still in fact very widely used. The additional statement that the inadequacy of psychiatric categories is admitted

by all concerned is a rather subtle misuse of words. Inadequacy is no reason for abandoning, unless something more adequate can be substituted. One still has the possibility of making the inadequate more adequate. An adequate reason for abandoning existing categories would be their unsuitability. If our supply of Burgundy were inadequate, we would strive to make it more adequate by ordering more and better vintages; if our supply were unsuitable—for example, all the bottles were corked—we would abandon it and strive to obtain something more suitable.

The position then is that the inadequacy of psychiatric nosology is probably recognized by all concerned, but its unsuitability is not so recognized. Many citations could be made to justify this view, but two will suffice. Zubin (1953) writes: "While the reaction-pattern type of model to explain the etiology of neurosis is an attractive one, it does not lend itself as readily to experimental manipulation as the disease-entity model with the potential physiological, biochemical and genetic factors that enter into the postulation of disease models. If one regards the problem, not as a search for basic etiology, but as a search for serviceable models leading to further knowledge about etiology . . . the disease model is to be preferred." Again, Wittenborn (1951) writes: "If a well-formulated logically coherent theory for the psychoses were available, the intervening variables of this theory could provide the basis for economical description. Since no such theory is available, any phenomenal clustering tendency provides a plausible conceptualization and allows us to use some of the advantages of theory, i.e. measurement based on intervening variables . . . The manifestation of symptom clusters, regardless of the exact conditions responsible for their appearance, takes us a step nearer to our goal."

Is this psychodiagnostic procedure guaranteed to keep us moving in circles? It may do so; but it is certainly not guaranteed to do so. If it should turn out that the test categories agree in the main with reasonably reliable psychiatric categories, then the way is opened up for the description and measurement by means of these tests of the attributes of these categories. This is not an unimportant task for the psychologist. As a matter of fact Wittenborn has shown in a number of factorial studies, using a specially designed rating scale, that the factors he has been able to extract are rather congruent with the well-known psychiatric categories (Wittenborn, 1951) and that this holds even when employing psychiatrists from different institutions (Wittenborn, 1952). A study of these factors may well help us to isolate underlying processes which can be brought under experimental control. There seems then no reason for agreeing with Noyes that psychiatrists should be more interested at this stage in processes than in labels.

If psychiatric diagnoses are to be used as a criterion, obviously they must be reliable. Although psychiatric diagnosis is crucial to perhaps the majority of psychiatric, and to many psychological, researches, this vital question of reliability has scarcely been studied at all.

Ash (1949) reports agreement on major diagnostic categories between pairs of psychiatrists of 58 per cent. to 67 per cent. and on specific categories of 31½ per cent. to 43½ per cent. This is regarded as discouraging. There are, however, a number of unusual features in this study which make the results an unreliable guide for most psychiatric practice. Only 35 of the 139 cases were said to have indicated a clear-cut psychopathology, interviewing was carried out jointly and categories were used which were not mutually exclusive. Among specific categories—for example Psychopaths—are found schizoid, constitutional inferiority, drug addiction, epileptoid, etc. Clearly there is no reason why

a schizoid psychopath should not exhibit constitutional inferiority and be a drug addict. When account is taken of these facts, it is not surprising that the figures are low.

The studies by Doering (1934) and by Mehlman (1952) both utilize diagnoses made in the ordinary course of clinical work. Whilst no doubt such diagnoses should always be made with as much care as possible, it cannot be concluded from such studies that psychiatrists are incapable of making reliable assessments when put on their mettle. In addition, no preliminary discussion to clear up any semantic difficulties was possible in these studies. Masserman and Carmichael (1938) found that a major revision of diagnosis was necessary in 40 per cent. of cases within a year. These again were ordinary workaday diagnoses.

Hunt, Wittson and Hunt (1953) report on a series of Service Personnel. They estimated the agreement between psychiatrists at the Pre-Commission Station and at the Hospital to which the cases were subsequently sent. The diagnoses were virtually uncontaminated. Agreement on unsuitability for further service—a relatively easy task—was 94 per cent., about which there could be no complaint. The average agreement on the major categories: Psychosis, Psychoneurosis, Personality Disorder, was 54 per cent. and on specific categories 33 per cent. This again is treated as discouraging; but once more categories were used which were not mutually exclusive. Among Psychoneurotics are to be found the following: mixed, anxiety, unclassified, hysteria, situational, neurasthenia, psychasthenia, traumatic, hypochondria, inadequate, reactive depression. Logically this is like asking “Is this an elephant? (If so, what sort?) *or* is it from Africa? *or* from not very good stock? *or* a bit of each? *or*, if you don't know, say so”. In other words, there are at least three distinct principles of division, one denial of these principles and one expression of inability to apply these principles.

The authors went on to examine the changes in diagnosis between the two centres and these were found to be rather consistent. As they put it, “the changes were nearly all neighbourly changes which did not involve much clinical displacement and raise the question as to whether the ‘practical’ reliability of the diagnosis (in terms of the care, treatment and disposition of the cases) is not larger than our ‘pure’ statistics would indicate.” This raises another question, which does not appear to have been dealt with at all, whether the use in all these studies of a simple dichotomy of agreement—disagreement is inadequate, if not unsuitable. If one psychiatrist makes a diagnosis of Paranoid Schizophrenia, another of Paraphrenia and a third of Hysteria, there would almost certainly be general agreement that the views of the first two psychiatrists were much closer together than that of either with the third. A Scale, which takes account of partial agreement, is, therefore, necessary. In the present state of our knowledge such a scale inevitably lacks experimental support. This may be remedied as more studies of the type conducted by Eysenck (1952) and Wittenborn (1951) become available.

The Value of Psychological Tests

How far do the results of psychological tests, when administered first by the same psychologist and then by different psychologists, agree with the initial and final psychiatric diagnoses? In other words, do psychologists gain by observing the patient during the testing session? Do psychologists in fact gain by testing patients at all or could they diagnose just as well on the basis of knowledge of the type of patients admitted to a particular division and of the

age of each particular patient? Can psychologists be of any diagnostic assistance to psychiatrists?

THE PROCEDURE

In order to elicit information relevant to the above problems, it was necessary for at least two psychiatrists to diagnose the subjects independently, for tests to be administered and interpreted first by the same psychologist (without any information other than the subject's age), then by different psychologists (without any information other than age and sex), for a final diagnosis to be agreed upon by the two psychiatrists and for a scale ranging from complete agreement to complete disagreement to be employed.

Psychiatric Diagnosis

Approximately one week after admission each patient was interviewed independently by each psychiatrist. Otherwise the usual diagnostic procedure was employed. Both had available, for example, the Psychiatric Social Worker's report and any other relevant information which might have been supplied. No discussion of the patient took place until each psychiatrist had recorded his diagnosis and deposited it with the Ward Sister or Charge Nurse. Just before the patient's discharge or after approximately one to two months, the two psychiatrists discussed the case together and arrived at a final diagnosis. They remained in ignorance of the test results.

Had the main purpose of the investigation been the assessment of the reliability of psychiatric diagnoses, the final diagnosis would, of course, have been made independently by each psychiatrist. For the present purpose of ascertaining the validity of the tests it was considered that any disadvantages accruing from contamination would be offset by the probable strengthening of the final criterion.

Since it is scarcely feasible to obtain a chance score with which to compare the psychiatric and psychological diagnoses, it was decided to ask a psychiatrist and a psychologist to "guess" the diagnoses. Both of them were familiar with the type of patients admitted to each of the two Divisions and both were provided with the age and sex of the patient. Obviously such information should enable one to make better than chance guesses.

There were seven participants altogether—the Consultant Psychiatrist and the Senior Hospital Medical Officer on each Ward, the Psychiatric Research Fellow, who did the guessing, and two Psychologists. The Consultant Psychiatrists are henceforward referred to as A, the SHMOs as B, the Experimenter as E, the psychiatric "guesser" as X and the psychological "guesser" as Y. E was, in fact, a dual personality, since he was also Y. The agreement between them was, however, not impressive and it may safely be concluded that the dissociation was rather complete.

Psychological Testing

Testing was carried out on the average four days after admission. The tests, in their usual order of administration, were: The Porteus Mazes (with the method of administration and scoring described by Foulds (1951)); the Mill Hill Vocabulary Scale, four pictures from the Thematic Apperception Test (with the method of administration and scoring described by Foulds (1952)) and the Progressive Matrices. The ratio between the Matrices and Mill Hill tests has been found to be of some diagnostic value (Foulds, 1953; Himmelweit, 1945).

During the sessions in which the investigator also administered the tests, subjects were discouraged from entering into any discussion not immediately relevant to the tests. This discouragement was almost completely successful.

The Diagnostic Agreement Scale

The following diagnostic categories were used in this study:

Paranoid State
Schizophrenia
 Paranoid
 Katatonic
 Excitement
 Stupor
 Hebephrenic
 Simple
Manic-Depressive
 Manic
 Melancholia
Organic
Schizoid Psychopath
Dysthymia
 Obsessional
 Reactive Depression
 Anxiety State
Hysteria

The following diagnostic pairs were treated as identical:

Mania—Hypomania
Paranoid State—Paraphrenia
Hysteroid Psychopath—Hysteria
Manic-Depressive, Depressive State—Melancholia
Anxiety Hysteria—Hysteria with Anxiety features

The following diagnostic pairs were treated as similar:

Paranoid State—Paranoid Schizophrenia
Involutional Paranoid State—Involution Melancholia
Paranoid Schizophrenia—Katatonic Schizophrenia
Paranoid Schizophrenia—Hebephrenic Schizophrenia
Paranoid Schizophrenia—Simple Schizophrenia
Katatonic Schizophrenia—Hebephrenic Schizophrenia
Katatonic Schizophrenia—Simple Schizophrenia
Hebephrenic Schizophrenia—Simple Schizophrenia
Simple Schizophrenia—Schizoid Psychopath
Melancholia—Reactive Depression
Reactive Depression—Obsessional
Reactive Depression—Anxiety State
Obsessional—Anxiety State

The remaining 65 possible pairs were treated as different.

The method of scoring the scale was as follows:

Diagnoses Identical	..	4	Features Identical	..	2
Diagnoses Similar	..	2	Features Similar	..	1
Diagnoses Different	..	0	Features Different	..	0

If Diagnosis and Features have been reversed, both possible reversals are scored and two added if the unreversed Diagnoses are similar. Thus: Diagnosis Similar, but Diagnosis and Features reversed, otherwise Identical and Different

(or Different and Identical). E.g. Anxiety State with hysterical Features and Reactive Depressive with Anxiety Features would be scored: 2, 4, 2. The mean is then taken (to the nearest whole number) i.e. 3. Again, Diagnosis Different, Diagnosis and Features reversed, otherwise Identical and Similar—e.g. Anxiety State with Hysterical Features and Hysteria with Depressive Features would be scored: 0, 5, 4, the final score being the mean of 3.

Each of the possible score categories is illustrated below, thus:

1. *Diagnosis and Features identical* Score 6
 - A. Anxiety State with Depressive features
 - B. Anxiety State with Depressive features
2. *Diagnosis identical and no Features mentioned* Score 6
 - A. Conversion Hysteria
 - B. Conversion Hysteria
3. *Diagnosis identical and Features similar* Score 5
 - A. Obsessional with Anxiety features
 - B. Obsessional with Depressive features
4. *Diagnosis identical and Features different* Score 4
 - A. Reactive Depressive with Obsessional features
 - B. Reactive Depressive with Hysterical features
5. *Diagnosis identical and Features mentioned in one only* Score 4
 - A. Reactive Depressive with Obsessional features
 - B. Reactive Depressive
6. *Diagnosis similar with Features identical* Score 4
 - A. Paranoid Schizophrenia with Obsessional features
 - B. Paraphrenia with Obsessional features
7. *Diagnosis similar with Features not mentioned* Score 4
 - A. Paranoid Schizophrenia
 - B. Paraphrenia
8. *Diagnosis similar and Diagnosis and Features reversed, otherwise identical* Score 5
 - A. Anxiety State with Depressive features
 - B. Reactive Depressive with Anxiety features
9. *Diagnosis similar and Features similar* Score 3
 - A. Schizoid Personality with Anxiety features
 - B. Simple Schizophrenia with Depressive features
10. *Diagnosis and Features reversed, otherwise identical* Score 4
 - A. Anxiety State with Obsessional features
 - B. Obsessional with Anxiety features
11. *Diagnosis similar and Diagnosis and Features reversed, otherwise identical and different* Score 3
 - A. Anxiety State with Obsessional features
 - B. Reactive Depressive with Anxiety features
12. *Diagnosis similar and Diagnosis and Features reversed, otherwise identical and Features mentioned in one only* Score 3
 - A. Anxiety State
 - B. Reactive Depressive with Anxiety features
13. *Diagnosis similar and Features different* Score 2
 - A. Melancholia with Paranoid features
 - B. Reactive Depressive with Obsessional features
14. *Diagnosis similar and Features mentioned in one only* Score 2
 - A. Melancholia with Paranoid features
 - B. Reactive Depressive
15. *Diagnosis different, but Features identical* Score 2
 - A. Anxiety State with Obsessional features
 - B. Schizophrenia with Obsessional features
16. *Diagnosis and Features reversed, otherwise identical and similar* Score 3
 - A. Anxiety State with Hysterical features
 - B. Hysteria with Depressive features
17. *Diagnosis and Features reversed, otherwise similar* Score 2
 - A. Melancholia with Paraphrenic features
 - B. Paranoid Schizophrenia with Depressive features

18. *Diagnosis different, but Features similar* Score 2
 A. Obsessional with Anxiety features
 B. Hysteria with Depressive features
19. *Diagnosis and Features reversed, otherwise identical and different* Score 2
 A. Anxiety State with Obsessional features
 B. Hysteria with Anxiety features
20. *Diagnosis and Features reversed, otherwise identical and Features mentioned in one only* Score 2
 A. Anxiety State
 B. Hysteria with Anxiety features
21. *Diagnosis and Features reversed, otherwise similar and different* Score 1
 A. Melancholia with Paranoid features
 B. Obsessional with Depressive features
22. *Diagnosis and Features reversed, otherwise similar and Features mentioned in one only* Score 1
 A. Melancholia
 B. Obsessional with Depressive features
23. *Diagnosis and Features different* Score 0
 A. Schizophrenia with Obsessional features
 B. Melancholia with Paranoid features
24. *Diagnosis different and Features mentioned in one only* Score 0
 A. Schizophrenia with Obsessional features
 B. Melancholia
25. *Diagnosis different and no Features mentioned* Score 0
 A. Melancholia
 B. Hysteria

The Scale is clearly lacking in experimental support for the attribution of similarity to the various diagnostic pairs and for the particular weightings given to the scores. Improvements will have to wait upon further research. In the meantime the Scale does appear to give a more accurate measurement of agreement than any simple dichotomy of agreement-disagreement is likely to do.

The Subjects

The subjects were successive admissions to two of the Divisions of the Hospital, one containing both male and female Wards, the other male only.

The only cases deliberately excluded were those who had been tested at one or other of the Out-patient Departments prior to admission and those suffering from a disparate physical illness. In fact five other cases were missed because the psychiatrist decided that immediate treatment was necessary. Since these five were classical cases, their exclusion has almost certainly reduced the agreement both between psychiatrists and between psychiatrists and psychologist. One further case was omitted because he could not be induced to do any of the tests at all. Those who did something, but not everything on the tests were included. Finally one case was omitted because the Charge Nurse inadvertently asked the psychologist to test the patient on the Ward as he was an Epileptic.

RESULTS

Eighteen Cases in Which the Tests were Administered and Interpreted by the Same Psychologist

Table I shows the amount of agreement for each of the ten pairings.

TABLE I
Agreement Between Diagnosticians in 18 Cases Tested and Diagnosed by the Same Psychologist

	A:B	A:E	B:E	A:X	A:Y	B:X	B:Y	E:X	E:Y	X:Y
Mean	3.78	4.06	4.11	2.17	2.56	2.33	2.33	1.61	2.06	2.61
S.D.	1.84	1.81	1.45	1.98	1.77	2.11	2.08	1.64	2.20	2.19

A—Consultant Psychiatrist; B—SHMOs; E—Experimenter; X—Psychiatric "guesser"; Y—Psychological "guesser".

Table II shows the *t* (for correlated means) and *P* values for comparisons derived from Table I.

TABLE II
t and *P* Values for Comparisons Derived from Table I

	<i>t</i>	<i>P</i>		<i>t</i>	<i>P</i>		<i>t</i>	<i>P</i>
AB and AE	=0.62	> .1	AE and BE	=0.12	> .1	BE and AX	=3.23	< .01
and BE	=0.82	> .1	and AX	=2.59	< .02	and AY	=2.98	< .01
and AX	=2.30	< .05	and AY	=2.14	< .05	and BX	=2.78	< .02
and AY	=2.30	< .05	and BX	=2.93	< .01	and BY	=3.18	< .01
and BX	=2.42	< .05	and BY	=2.98	< .01	and EX	=5.21	< .001
and BY	=3.02	< .01	and EX	=4.30	< .001	and EY	=3.60	< .01
and EX	=3.81	< .01	and EY	=3.03	< .01	and XY	=2.94	< .01
and EY	=2.97	< .01	and XY	=2.74	< .02			
*and XY	=1.92	< .1						

There were thus 21 comparisons between two psychiatrists (or a psychiatrist and the experimenter) on the one hand and a psychiatrist (or the experimenter) and a "guesser" on the other hand. Of these 21 comparisons one (AB and XY) failed to reveal a difference significant at the 5 per cent. level of confidence; 4 were significant at the 5 per cent. level of confidence; 3 were significant at the 2 per cent. level; 11 at the 1 per cent. level and 2 at the 0.1 per cent level.

There were 3 comparisons between the psychiatrists and the experimenter and none of these was significant even at the 10 per cent. level of confidence.

The two psychiatrists who investigated the patients and the experimenter were, therefore, more or less equally successful in reaching agreement among themselves at a level well above that achieved by the "guessers".

Table III shows the amount of agreement between the experimenter and the final psychiatric diagnosis and between each of the "guessers" and the final psychiatric diagnosis.

Table IV shows the *t* and *P* values for comparisons derived from Table III.

TABLE III
Agreement Between the Final Psychiatric Diagnosis and the Experimenter and the Two "Guessers"

	E:C	X:C	Y:C	
Mean	4.50	1.94	2.22	C—Criterion diagnosis
S.D.	1.72	1.87	1.96	

TABLE IV
t and *P* Values for Comparisons Derived from Table III

	<i>t</i>	<i>P</i>
EC and XC	= 4.06	< .001
EC and YC	= 3.45	< .01
XC and YC	= 0.58	> .1

The agreement between the experimenter and the final psychiatric diagnosis differed at the 0.1 per cent. level of confidence from "guesser" X and the final psychiatric diagnosis and at the 1 per cent. level from Y and the final diagnosis.

Eighteen Cases in Which the Tests were Administered and Interpreted by Different Psychologists

Table V shows the amount of agreement for each of the ten pairings (as in Table I).

Table VI shows the *t* and *P* values for comparisons derived from Table V.

TABLE V

Agreement Between Diagnosticians in 18 Cases Tested and Diagnosed by Different Psychologists

	A:B	A:E	B:E	A:X	A:Y	B:X	B:Y	E:X	E:Y	X:Y
Mean ..	4.44	3.67	3.44	1.67	2.00	1.33	1.61	2.11	2.11	2.11
S.D. ..	1.71	1.79	2.03	1.92	1.86	1.79	1.92	1.94	1.83	1.41

TABLE VI

t and P Values for Comparisons Derived from Table V

	t	P		t	P		t	P
AB and AE	=1.60	>.1	AE and BE	=0.83	>.1	BE and AX	=2.64	<.02
and BE	=2.02	<.1	and AX	=3.08	<.01	and AY	=2.25	<.05
and AX	=3.79	<.01	and AY	=2.32	<.05	and BX	=3.40	<.01
and AY	=3.30	<.01	and BX	=3.97	<.001	and BY	=3.27	<.01
and BX	=4.78	<.001	and BY	=3.55	<.01	*and EX	=1.66	>.1
and BY	=4.35	<.001	and EX	=2.29	<.05	*and EY	=1.96	<.1
and EX	=3.15	<.01	and EY	=2.69	<.02	and XY	=2.25	<.05
and EY	=3.70	<.01	and XY	=2.94	<.01			
and XY	=4.02	<.001						

Of the 21 comparisons between two psychiatrists (or a psychiatrist and the experimenter) on the one hand and a psychiatrist (or the experimenter) and a "guesser" on the other hand, 2 failed to reveal a difference significant at the 5 per cent. level of confidence; 4 were significant at the 5 per cent. level; 2 at the 2 per cent. level; 9 at the 1 per cent. level and 4 at the 0.1 per cent. level.

Of the 3 comparisons between the psychiatrists and the experimenter none was significant at the 5 per cent. level of confidence.

The two psychiatrists, who investigated the patients, and the experimenter, who interpreted the tests without seeing the patients, were successful in reaching agreement among themselves at a level well above that achieved by the "guessers".

Table VII shows the amount of agreement between the experimenter and the final psychiatric diagnosis and between each of the "guessers" and the final psychiatric diagnosis.

Table VIII shows that the t and P values for comparisons derived from Table VII.

TABLE VII

Agreement Between the Final Psychiatric Diagnosis and the Experimenter and the Two "Guessers"

	E:C	X:C	Y:C
Mean ..	3.83	1.28	2.00
S.D. ..	1.83	1.63	2.24

TABLE VIII

t and P Values for Comparisons Derived from Table VII

	t	P
EC and XC ..	4.40	<.001
EC and YC ..	2.69	<.02
XC and YC ..	1.11	>.1

The agreement between the experimenter and the final psychiatric diagnosis differed at the 0.1 per cent. level of confidence from "guesser" X and the final psychiatric diagnosis, and at the 1 per cent. level from "guesser" Y and the final psychiatric diagnosis.

The experimenter may, therefore, be seen to have been somewhat less successful in his diagnostic predictions when he did not see the patients. The difference was not, however, statistically significant.

DISCUSSION

The unreliability of psychiatric diagnoses has been assumed without adequate experimental support. In the present investigation there was a minority of cases in which agreement by any standard would have to be regarded as poor. These are probably the cases which are discussed most frequently and which are presented at Case Conferences, thus contributing to a biased impression. The problem is not so much can psychiatrists reach agreement on diagnoses, but under what circumstances can they reach agreement and with what sort of people?

Some such scale of diagnostic agreement as the one used in this study seems to be essential. It is suggested that only those cases which have a sufficiently high mean rating score when diagnosed independently by three psychiatrists should be used in any research investigation in which the diagnosis is crucial. This is a stringent requirement; but it would have at least two important advantages: firstly, it would discourage premature publication of findings which could not be other than inconclusive and secondly, it would, by increasing the homogeneity of groups almost certainly reduce the variability of measures, thus enabling the investigator to differentiate more precisely between smaller groups. The rejected cases could, of course, be included as a separate group and their scores might be expected to fall somewhere between the extremes of any two groups which had been clearly differentiated.

In the setting of the present investigation the extent of agreement between psychiatrists, three out of four of whom had been working together for some years, did not appear to be too discouraging and could almost certainly be increased by more thorough discussion of semantic difficulties. Agreement between either psychiatrist and a psychologist using a short battery of tests (taking on the average about one and a half hours in all) was of about the same order. It would appear, therefore, that the psychologist, even when he does not make use of a Social History, etc., may be engaging in a reasonably meaningful and possibly even useful activity.

SUMMARY AND CONCLUSIONS

Criticisms of psychiatric nosology have been discussed and it has been argued that, whilst its inadequacy is generally admitted, its unsuitability has not been demonstrated.

Agreement between psychiatrists making independent diagnoses was in the region of 4 on a scale ranging from 0 to 6. Illustrations of the type of agreement represented by a score of 4 can be seen in the Appendix.

Using the agreed final psychiatric diagnosis as criterion, diagnostic predictions were made from a short battery of tests. When the psychologist interpreted and administered the tests, agreement was 4.50 on the Scale. When he made the diagnoses from the tests without seeing the patient, agreement was 3.83. The difference between the two series was not significant. Diagnoses based on knowledge of the type of patients admitted and of the age and sex of the particular patients showed a mean agreement with the criterion diagnosis of less than 2 on the scale.

In the setting of this investigation it would appear that the reliability of psychiatric, and the validity of psychological, diagnoses are not as poor as current opinion would lead one to suppose. The reader can form his own judgment from the list in the Appendix of diagnoses made on the 18 cases which were seen by two psychiatrists and the experimenter.

ACKNOWLEDGMENTS

I wish to thank Dr. R. Ström-Olsen, the Physician Superintendent, for permission to publish the results of this study; Mr. J. E. Macdonald for the testing of half of the subjects; Dr. J. D. Montagu for not being too brilliant a "guesser" and, finally, Drs. W. P. Berrington, D. W. Liddell, A. A. Robin and D. C. L. Thomas for their courageous and painstaking clinical contribution.

REFERENCES

- ASH, P., *J. abnorm. soc. Psychol.*, 1949, **44**, 272-277.
 DOERING, C. R., *Amer. J. Orthopsychiat.*, 1934, **4**, 249-257.
 EYSENCK, H. J., *The Scientific Study of Personality*, 1952. London.
 FOULDS, G. A., *Brit. J. Psychol.*, 1951, **42**, 209-217.
Idem, ibid., 1952, **43**, 33-41.
Idem, Speed of performance on mental tests among psychoneurotics, 1953. Unpublished Ph.D. thesis. University of London.
 HIMMELWEIT, H. T., *J. of Personality*, 1945, **14**, 95-105.
 HUNT, W. A., WITTON, C. L., and HUNT, E. B., in *Current Problems in Psychiatric Diagnosis* (Eds., Hoch, P. H., and Zubin, J.), 1953, 53-65. New York.
 MASSERMAN, J. H., and CARMICHAEL, H. T., *J. Ment. Sci.*, 1938, **84**, 893-946.
 MEHLMAN, B., *J. abnorm. soc. Psychol.*, 1952, **47**, 577-578.
 NOYES, A. P., *Modern Clinical Psychiatry*, 1953, p. 159. Philadelphia and London.
 ROE, A., *J. abnorm. soc. Psychol.*, 1949, **44**, 36-41.
 WITTENBORN, J. R., *J. Consult. Psychol.*, 1951, **15**, 290-302.
Idem, ibid., 1952, **16**, 107-109.
 ZUBIN, J., in *Current Problems in Psychiatric Diagnosis* (Eds., Hoch, P. H., and Zubin, J.), 1953, 112-118. New York.

APPENDIX

Cases diagnosed by a Consultant Psychiatrist (A), a Senior Hospital Medical Officer (B) and the Experimenter (E)

	Major Diagnosis	Features	Scores
1.	(A) Anxiety-Hysteria	Depressive	AB.3
	(B) Reactive Depression	Hysterical and Obsessional	AE.3
	(E) Obsessional	Hysterical	BE.4
2.	(A) Anxiety-Hysteria		AB.6
	(B) Anxiety-Hysteria		AE.3
	(E) Reactive Depression	Hysterical	BE.3
3.	(A) Reactive Depression	Anxiety	AB.5
	(B) Anxiety State	Depressive	AE.5
	(E) Anxiety State	Depressive	BE.6
4.	(A) Melancholia	Paranoid	AB.1
	(B) Hysteria	Depressive	AE.5
	(E) Paraphrenia	Melancholic	BE.2
5.	(A) Reactive Depression	Hystero-Psychopathic	AB.4
	(B) Hysteria	Depressive	AE.6
	(E) Reactive Depression	Hysterical	BE.4
6.	(A) Reactive Depression	Hysterical	AB.2
	(B) Paraphrenia	Depressive and Hysterical	AE.2
	(E) Paraphrenia	Hysterical	BE.5
7.	(A) Schizoid Personality	Obsessive	AB.4
	(B) Schizophrenia	Obsessive and Anxiety	AE.6
	(E) Schizoid Personality	Obsessive	BE.4
8.	(A) Schizophrenia Simplex		AB.4
	(B) Schizoid Psychopath		AE.6
	(E) Schizophrenia Simplex		BE.4
9.	(A) Melancholia		AB.0
	(B) Hysteria		AE.2
	(E) Reactive Depression	Hysterical	BE.2
10.	(A) Reactive Depression	Anxiety	AB.4
	(B) Reactive Depression	(Alcoholism)	AE.3
	(E) Anxiety State		BE.4
11.	(A) Anxiety-Hysteria		AB.6
	(B) Anxiety-Hysteria		AE.5
	(E) Hysteria	Obsessive	BE.5
12.	(A) Melancholia	Psychopathic	AB.2
	(B) Reactive Depression		AE.2
	(E) Reactive Depression	Anxiety	BE.4
13.	(A) Manic-manic-depressive	Depressive	AB.4
	(B) Manic-manic-depressive		AE.4
	(E) Manic-manic-depressive		BE.6
14.	(A) Manic-manic-depressive		AB.6
	(B) Manic-manic-depressive		AE.6
	(E) Manic-manic-depressive		BE.6
15.	(A) Paranoid Schizophrenia	(Homosexual)	AB.4
	(B) Paranoid Schizophrenia	Depressive	AE.6
	(E) Paranoid Schizophrenia	(Homosexual)	BE.4

	Major Diagnosis	Features	Scores
16.	(A) Hebephrenic Schizophrenia		AB.6
	(B) Hebephrenic Schizophrenia		AE.6
	(E) Hebephrenic Schizophrenia		BE.6
17.	(A) Hebephrenic Schizophrenia		AB.4
	(B) Hebephrenic Schizophrenia	Paranoid	AE.0
	(E) Psychopath	Paranoid	BE.2
18.	(A) Schizophrenia Simplex	Depressive	AB.3
	(B) Paranoid Schizophrenia	Simplex	AE.3
	(E) Paranoid Schizophrenia	Anxiety	BE.4