AN EVALUATION OF THE PREDICTIVE POWER OF A TEST KNOWN TO DIFFERENTIATE BETWEEN ELDERLY "FUNCTIONAL" AND "ORGANIC" PSYCHIATRIC PATIENTS

By

JAMES INGLIS, Ph.D.
Assistant Professor of Psychology
Queen's University, Kingston, Ontario

CATHERINE COLWELL
Psychiatric Social Worker

and

FELIX POST, M.B., M.R.C.P., D.P.M.
Consultant Psychiatrist
The Maudsley Hospital and The Bethlem Royal Hospital, London

FEW studies of psychological tests which have been shown to differentiate between psychiatric groups have attempted to follow-up the patients originally tested so as to determine what the predictive power of the test might be in the light of subsequent events. One of these few studies has been reported by Walton (1958) and commented upon by Inglis (1959) who was able to show that the test Walton described was better able to predict the outcome of illness after a follow-up period of two years, than was the original diagnostic label. The importance of this kind of study has been pointed out most clearly by Payne (1958). His argument is that, "Description is only one implication of the diagnostic label. Can the test score aid the doctor in making a prognosis? This need not be the case. Let us consider the original validation of the test again. The doctor who diagnosed the standardization group of patients might well be able to give a more or less accurate prognosis for this group of patients. In fact there might be a significant relationship between the presence or absence of the label he assigns, and prognosis. We also know that there is a significant correlation between presence or absence of his label and the *test* scores. *This does not prove, however, that there is any relationship whatsoever between the diagnostic test score and prognosis.* Two things which correlate with the same thing do *not* necessarily correlate with each other, unless the correlations concerned are greater than ·7" (1958, p. 27).

The present paper reports the investigation of the predictive power of an adaptation of the Bender Visual Motor Gestalt Test which had already been shown to discriminate significantly between "functional" and "organic" groups of elderly psychiatric patients in a study reported by Shapiro, Post, Lofving and Inglis (1956). The criteria used for the initial diagnostic classification are fully discussed in the latter paper. The differentiation between these groups on the test was achieved through the use of a relatively subjective scoring system which involved a simple rating of each of the patient's drawings.

Shapiro, Field and Post (1957) were able to show that a more objective measure derived from the same test also differentiated the criterion groups at the same high level of significance. This was an angles-measure score which involved the summing, regardless of sign, of the deviations from a right angle of each of the eight angles of two designs. The predictive power of this score is examined in the present study. In each case the angles measure was secured by Dr. Jack Field (1958) in a series of investigations which he carried out on elderly psychiatric patients.

METHOD

Subjects. The individuals followed up had all been patients in the Geriatric Unit of the Bethlem Royal Hospital. Altogether 59 patients were so investigated, although in the numerical data to be reported the overall number is not constant for each variable since some data were not available for all patients. There were 18 men and 41 women, with a mean age of about 69 ($\overline{m}=68\cdot66$; s.d.$=6\cdot23$). Twenty-one of these had originally been diagnosed as "organic" and 38 as "functional".

Follow-up Procedure. The psychiatric social worker tried to follow up each patient individually, as near as possible two years after the patient had been given the psychological test. She succeeded in contacting each patient's relatives and in collecting information under the separate headings given below. This information was then rated by the psychiatrist and by the psychologist, working quite independently, in terms of the criteria and scores listed under each heading, as follows:

A. *The patient's daily routine during his "best" phase of the follow-up period since in-patient stay:*

General Description	*Score*
Inactive	0
Pottering about and "helping"	1
Employed, doing efficient housework, etc.	2

B. *Evidence for psychological disorders and symptoms, other than cognitive, since period in hospital:*

General Description:	*Score*
Continuously severely disabling	0
Continuously mildly disabling	1
Intermittently severely disabling	2
Intermittently mildly disabling	3
No psychological disturbance at any time	4

C. *Evidence of cognitive disturbances and intellectual changes since in-patient period:*

General Description:	*Score*
Definite	0
Dubious	1
None	2

D. *Evidence of personality changes since in-patient period:*

General Description:	*Score*
Never anything like his old self, or continuously ill	0
Mild invalidism	1
As well as, or "better than" before illness	2

In addition to these items other evidence was gathered by the psychiatrist on several topics. A note was made, for example, of any changes in diagnosis. This item did not, in fact, prove to be of any importance since, after a space of two years, only one patient had been re-diagnosed; this was a patient originally labelled organic who was later re-categorized as functional. Account was also taken of the number of months the patient had spent in any hospital from the date of testing, expressed as a percentage of the follow-up period. Finally it was noted if the patient were still alive or not.

Results

The results obtained may be conveniently considered in terms of three problems.

1. To determine the degree of reliability of the independent ratings made by the psychiatrist and the psychologist of the data gathered by the social worker.

2. To examine the prognostic power of the first diagnosis in terms of its correlation with the various estimates made of the outcome of illness over the two-year period.

3. To examine the predictive power of the test score in the same terms.

1. *The Reliability of the Ratings*

The reliability of the independent ratings made of the social worker's notes by the psychiatrist and the psychologist was simply assessed by running product-moment correlations (McNemar, 1949) between the two sets of ratings. The results are shown in Table I.

Table I

Product-moment Correlations Between Ratings Made Independently by the Psychiatrist and the Psychologist

		r
A.	Best daily behaviour	0·87*
B.	Non-cognitive disturbances	0·66*
C.	Intellectual changes	0·85*
D.	Personality changes	0·90*
	Sum of ratings (A+B+C+D)	0·92*

* Significant at or beyond the 1 per cent. level.

It can be seen that the correlations obtained reflect a large amount of agreement between the two raters. This shows that the criteria used could be regarded as reliable indicators of the patients' state as described in the social worker's notes.

As most of the patients had been known to the psychiatrist since their first admission to hospital his ratings were accepted as being probably the more valid and these were therefore used in the attempts to elucidate the further problems examined.

2. *The Prognostic Power of the First Diagnosis*

In order to discover what indications the first diagnostic label might hold for the subsequent developments, as assessed by the various estimates secured, bi-serial correlations (McNemar, 1949) were run between the dichotomous diagnostic variable (i.e. functional or organic) and these estimates are shown in Table II. It should be noted that since the first diagnosis and the alive-dead classifications are both dichotomous a contingency coefficient (McNemar, 1949) was taken as the appropriate measure of association in this instance.

TABLE II

Bi-serial Correlations (etc.) Between First Diagnosis and the Variables Estimated in the Follow-up Study

	r bis.
First diagnosis × Best daily behaviour	0·59*
First diagnosis × Non-cognitive disturbances	0·48*
First diagnosis × Intellectual changes	0·57*
First diagnosis × Personalty changes	0·54*
First diagnosis × Sum of 4 ratings made by the psychiatrist	0·76*
First diagnosis × Percentage time in hospital	−0·21 N.S.
	c.c.
First diagnosis × Alive or dead	0·48*

* Significant at or beyond the 1 per cent. level.
N.S. = Not significant.

Table II shows that the first diagnosis has quite a significant relationship with the effects of the illness over the subsequent 2 years on a variety of estimates.

Patients diagnosed "functional" were, as one would expect, on follow-up significantly better adjusted to life, showing fewer psychiatric symptoms, fewer intellectual and personality changes, and were more often still alive, than those who at the time of testing had been clinically diagnosed as suffering from psychiatric illnesses associated with cerebral pathology. Functional patients had also spent less time in hospitals, though not to a significant extent. This, too, is not surprising for the following reason: though organic mental patients are more likely to become permanent hospital inmates, it has been shown recently (Colwell and Post, 1959) that "functionals" also tend to spend relatively long periods in hospital during the first two years after admission on account of their unfortunate tendency to recur at an early date.

It remains to be seen how the prognostic power of the clinical diagnosis compares with the relation between the same estimates and the test score.

3. *The Predictive Power of the Test*

(a) *The Group as a Whole.* For all the patients studied, simply considered as one group, the relation between the angles-measure score and most of the other variables could be estimated by product-moment correlations. The only exceptions were the relations between the test score and the dichotomous variables (i.e. first diagnosis, second diagnosis and alive-dead) and in these cases a bi-serial correlation was used. The results for the group as a whole are shown in Table III.

The predominance of negative correlations is simply due to the fact that a high score on the angles measure was characteristic of greater pathology whereas the reverse was true of most of the follow-up variables.

It is only to be expected that there should be significant relations between the test score and first and second diagnosis. This is merely another indication

TABLE III

Product-moment Correlations (etc.) Between the Angles-Measure Score and the Variables Estimated in the Follow-up Study for the Group Considered as a Whole

	r
Angles measure × Best daily behaviour	−0·26 N.S.
Angles measure × Non-cognitive disturbances	−0·16 N.S.
Angles measure × Intellectual changes	−0·46*
Angles measure × Personality changes	−0·20 N.S.
Angles measure × Sum of 4 ratings made by the psychiatrist	−0·28 N.S.
Angles measure × Percentage time in hospital	−0·23 N.S.
	r bis.
Angles measure × Alive or dead	0·01 N.S.
Angles measure × First diagnosis	−0·38†
Angles measure × Second diagnosis	−0·41*

* Significant at or beyond the 1 per cent. level.
† Significant at or beyond the 2 per cent. level.
N.S. = Not significant.

of the fact that the test successfully discriminated between the criterion groups.

It is, however, disappointing to find that for most of the follow-up variables the test appears to lack any predictive significances whatever. An interesting exception is provided by the significant correlation between the angles measure and follow-up evidence of intellectual decline. Even this, however, is certainly no better than the prognostic significance of the first diagnosis in relation to these same functions. (see Table II)

(b) *The Diagnostic Groups Considered Separately.* It would be possible to argue that even if the test had little predictive power for the group taken as a whole it might be more successful for one diagnostic group than for the other. Product-moment correlations were therefore run between test score and the other variables for the two groups separately. Since the alive-dead dichotomy entered into the calculations a bi-serial correlation was calculated in this case. The results are shown in Table IV.

TABLE IV

Product-moment Correlations (etc.) Between the Angles-Measure Score and the Variables Estimated in the Follow-up Study for the Groups Considered Separately

	Functional	Organic
	r	r
Angles measure × Best daily behaviour	0·02 N.S.	−0·05 N.S.
Angles measure × Non-cognitive disturbances	−0·03 N.S.	0·13 N.S.
Angles measure × Intellectual changes	−0·11 N.S.	−0·29 N.S.
Angles measure × Personality changes	0·06 N.S.	0·06 N.S.
Angles measure × Sum of 4 ratings made by the psychiatrist	0·11 N.S.	−0·06 N.S.
Angles measure × Percentage time in hospital	−0·16 N.S.	−0·10 N.S.
	r bis.	r bis.
Angles measure × Alive or dead	−0·001 N.S.	0·12 N.S.

N.S. = Not significant.

Again it can be seen that there is little or no relationship between the test score and the estimates made of the patient's state during the follow-up period. The test cannot, therefore, be used to predict either a malignant or a relatively benign course of the illness in patients with organic psychoses, nor will it

indicate the amount of future disabilities in the case of functional psychiatric disorders.

Discussion

The results of this investigation confirm most strongly the contention made by Payne (1958) that psychological tests must be directly validated for the purpose to which they are being put. The present investigation has demonstrated that even a test which has been validated and cross-validated in terms of its classificatory power does not necessarily have the predictive or prognostic power of the classificatory system against which it was initially standardized.

It is possible that this failure may be, in part at least, due to the indirect nature of the relation between the functions which are probably measured by the test and the psychiatric disturbances. After all, the inability to copy drawings is not a symptom of which elderly patients commonly complain; nor has it usually been regarded as a cardinal sign of disorder in those patients by the clinician. It may be that a more profitable approach to the discovery of those psychological variables which *do* carry some of the implications of valid psychiatric diagnosis will be made rather in the direct study of the principal descriptive characteristics of psychiatric disturbances. Such a method has, for example, been attempted by Inglis (1957) in the experimental investigation of memory disorder in elderly psychiatric patients.

The present study, on the other hand, successfully reconfirms the usefulness of psychiatric classification of elderly patients in terms of "functional" and "organic" syndromes, and is in line with other recent follow-up studies (Post, 1951; Roth and Morissey, 1952; Norris and Post, 1954).

Summary

This study reports on the investigation of the predictive power of a test known to discriminate between certain psychiatric categories. It was shown that reliable ratings could be made of the content of follow-up notes made by the psychiatric social worker. When, however, the test was assessed against these and other objective criteria it was shown that it did not have the prognostic power demonstrated by the original psychiatric diagnosis.

Acknowledgments

The authors would like to thank Dr. Jack Field for his kindness in providing the test data and Miss Elizabeth Carter for her help in computing the correlations.

References

Colwell, C., and Post, F., "The Community Needs of Elderly Psychiatric Patients", *Brit. med. J.*, 1959, *ii*, 214–217.

Field, J., "An investigation of the reproduction of simple designs by elderly psychiatric patients", Unpublished Ph.D. thesis. 1958. University of London Library.

Inglis, J., "An experimental study of learning and 'memory function' in elderly psychiatric patients", *J. Ment. Sci.*, 1957, **103**, 796–803.

Idem, "On the prognostic value of the modified word learning test in psychiatric patients over 65", *ibid*, 1959, **105**, 1100–1101.

McNemar, Q., *Psychological Statistics*, 1949. London: Chapman & Hall.

Norris, Vera and Post, F., "Treatment of elderly psychiatric patients; use of a diagnostic classification", *Brit. med. J.*, 1954, *i*, 675–679.

Payne, R. W., "Diagnostic and personality testing in clinical psychology", *Amer. J. Psychiat.*, 1958, **115**, 25–29.

POST, F., "The outcome of mental breakdown in old age", *Brit. med. J.*, 1951, *i*, 436–448.
ROTH, M., and MORRISEY, J. D., "Problems in the Diagnosis and Classification of Mental Disorders of Old Age", *J. Ment. Sci.*, 1952, **89**, 56–59.
SHAPIRO, M. B., FIELD, J., and POST, F., "An enquiry into the determinants of a differentiation between 'organic' and 'non-organic' psychiatric patients on the Bender Gestalt test", *ibid.*, 1957, **103**, 364–374.
Idem, POST, F., LOFVING, BARBRO, and INGLIS, J., " 'Memory function' in psychiatric patients over 60. Some methodological and diagnostic implications", *ibid.*, 1956, **102**, 233–246.
WALTON, D., "The diagnostic and predictive accuracy of the modified word learning test in psychiatric patients over 65", *ibid.*, 1958, **104**, 1119–1122.