Critical Issues in Psychiatric Diagnosis,
edited by Robert L. Spitzer and Donald F. Klein.

Discussion: Role of Psychological Testing
in Psychiatric Diagnosis

Joseph Zubin

V.A. Hospital, Pittsburgh, Pennsylvania; Department of Psychiatry, University of
Pittsburgh; and New York State Psychiatric Institute, New York, New York 10036

About a quarter of a century ago our association first tackled the problem
of the relation of psychological test to diagnosis. That symposium held in
1950 can serve as a base line from which to note the progress made in this
field. But before delving into history, let us survey this present volume.

All the contributors seem to agree that at the present time the usefulness
of psychological tests, especially of projective techniques in diagnosis, is
rather limited. They also seem to agree that the use of individual signs of
psychopathology in the form of specific Rorschach, TAT, and Draw-A-Man
responses is not a valid procedure for making inferences regarding the presence
of psychopathology. If there is any value in the use of projective tech-
niques, it inheres in regarding the protocols not as tests but as interviews
which, when analyzed for their content, yield the kind of information, perhaps less directly, that the usual clinical interview offers. This is the reason
why scaling Rorschach responses for their content rather than for the alleged
stimuli in the cards which evoked them may be a valuable approach. Such
content scaling yields dimensions which are also tapped by the clinical inter-
viewer, and hence one would expect considerable correlation between them.
I had put forth this idea for the usefulness of the Rorschach in the early
fifties. The last quarter of a century seems to have vindicated my contention.

By the way, Dr. Singer indicates that I suggested that “atomistic” ap-
proaches to the validation of the Rorschach be abandoned and that instead
global impressions be used. I am equally opposed to both of these methods.
Instead I proposed that careful scaling for content analysis, not global im-
pressions, be used.

It is clear that the difficulties in relating the results of psychological testing
to diagnosis are not inherent in the tests alone since the diagnostic schema is
also faulty. It is fortunate, however, that as a result of the efforts of our
Biometrics Research Unit, through the cooperation of my colleagues Bur-
dock, Endicott, Fleiss, Gurland, Hardesty, Sharpe, and Spitzer, combined
with the efforts of the Medical Research Council Social Psychiatry Research
Unit at the Institute of Psychiatry, the Maudsley Hospital in London headed
by John K. Wing, a group of systematic structured interviews of high reliability and some validity were developed.

It should be noted that systematic structured interviews represent a confluence of two disciplines—the psychological and the psychiatric, the former stressing the psychometric tradition of objectivity, reliability, and validity, the latter stressing the clinical tradition of understanding, insightfulness, and dynamic interaction. As is well known, the two disciplines had lived side by side in the halls of psychopathology—in hospitals and clinics—with but limited interaction. Until recently, the psychologist had limited himself to testing, eschewing the free-wheeling clinical interview, while the psychiatrist had limited himself to the interview in the search for a diagnosis. Only where psychologists and psychiatrists attained an equal footing could mutual interaction take place, and one of the places affording such free coalitions is the Biometrics Research Unit at the Psychiatric Institute. Here a series of structured interviews developed which combined the virtues of the two disciplines, mixing clinical relevance with psychometric precision. These interviews occupy the middle ground between psychological tests and clinical interviews, and in fact one of them—the Structured Clinical Interview of Burdock and Hardesty—actually regards itself as a test rather than an interview and provides standard norms for scoring the resulting dimensions and their profiles not unlike the way the Wechsler Bellevue Intelligence Test is scored. We undertook the development of the structured interviews when it was realized that the clinical interview, although unreliable and invalid, nevertheless was the only criterion available for determining the validity of tests in the clinical field. If it were to serve as a criterion, it had to be objectified and made reliable and valid.

The usefulness of these interviews has been fully demonstrated in the U.S.—U.K. Project on the Diagnosis of the Mental Disorders conducted by the Section on Diagnosis and Psychopathology of the Biometrics Research Unit headed by Barry Gurland (now Chief of the Department of Geriatrics at P.I.), with the help of Larry Sharpe in the United States and John E. Cooper, Robert E. Kendell, and John Copeland in the United Kingdom. A further evidence of their usefulness was afforded by the WHO International Pilot Study of Schizophrenia under the general guidance of Tsung Yi Lin, John K. Wing, and Norman Sartorius, and represented in the United States by Lyman Wynne, John Strauss, and William Carpenter.

But reliable interview results are not enough for arriving at a diagnosis, since the dimensional scores provided by these interviews have to be integrated into a typological schema before a diagnosis can be arrived at.

Now, with the help of the St. Louis group (Eli Robins, Sam Guze, and George Winokur), Drs. Spitzer and Endicott have succeeded in developing a set of operational definitions for the diagnostic categories in the form of Research Diagnostic Criteria which promise to provide a high level of reliability for the diagnostic labels themselves.
It is now possible to relate the dimensional profiles provided by the systematic structured interviews to the requirements demanded by the Research Diagnostic Criteria. It is even conceivable that a computer program could be developed for interdigitating the information from both sources, but it would probably be premature since the final diagnosis still requires some clinical flair.

We now have a much better base against which to judge the value of psychological tests in relation to diagnosis.

But why is diagnosis in the mental disorders fraught with so much difficulty? A comparison with diagnoses in the medical and social fields does not indicate that we are much worse off, but the troubles of others are only half a consolation. Our diagnoses depend primarily on overt behavior, and not on any external criteria independent of the gross deviant behavior in the form of neurophysiological or neurochemical or biological tests. We must realize that the behavior of the patient consists of three components: (a) his systematic way of behaving which characterized him before the episode developed, i.e., his personality, (b) his systematic way of behaving after the episode began—his psychopathology, and (c) incidental unsystematic happenstance aspects of his behavior. We can dismiss the third source, since it should not be systematic enough to interfere with our judgment of the first two sources, but it often causes difficulties in the unwary clinician.

Since the patient had a personality long before he became a patient, it is important to try to assess its premorbid character in any diagnostic procedure. In fact, this premorbid personality complicates our diagnoses. If we ever succeed in getting external criteria independent of gross behavior, we would have to detect the presence of the focal disorder which the patient has without much difficulty, even as it can be done in such conditions as TB or PKU. However, as long as we depend on behavior, we will have an admixture of the effects of the focal disorder impinging on the premorbid personality to produce an episode of illness. I say illness advisedly because it differs from person to person even though they may have the same focal disorder. The diagnosis of mental disorders is so difficult because we are faced with a complex illness from which we have to dissect the underlying cause—the focal disorder.

Perhaps the failure of the MMPI to relate to diagnosis is at least partially a reflection of the fact that the MMPI is basically a test of personality and not of psychopathology.

What is the relation between the premorbid personality and psychopathology?

I have elsewhere discussed this question at length (4,5). Here let me briefly indicate that premorbid personality may be regarded as identical with the psychopathology of the patient, or as totally independent of it, with the middle ground held by an interaction between the two. The evidence in the literature is rather evenly divided between the tenability of any of these
formulations, and to keep an open mind perhaps the second hypothesis—the null hypothesis—might be adopted tentatively.

Once we recognize the distinction between illness and focal disorder, we realize that we must also distinguish between the traits of the individuals and his states. A trait is a characteristic which persists throughout life with such modifications as the environment produces in it. A state is transitory, and the characteristic behavior during a state such as an episode disappears when the episode ends. This introduces another complication in diagnosis. Can we separate the traits which characterize the person both before and after the episode from the states—those behavioral characteristics which represent the psychopathology of the episode?

Thus far, this volume has dealt with traits that are highly subject to the modification of experience, and I wonder whether the scales of anhedonia and body image aberration developed by the Chapmans are trait or state related. Certainly many Rorschach responses are definitely state related, as I will argue later on the basis of recent evidence. Furthermore, even the trait-related behaviors are not uninfluenced by experience. That is why in our own studies of vulnerability we have stressed the use of techniques which involve responses that are relatively free of past experiences. We believe that neurophysiological responses and psychophysiological responses which occur within 1,000 msec following stimulation are relatively freer of the effects of prior environmental experiences and hence would show less variability and be more directly related to psychopathology.

Despite the influence of experience on such traits as anhedonia and perhaps even body image aberration, it would be well to consider whether individuals prone to schizophrenia have a lower threshold for these traits. If so, they could be used in detecting vulnerable individuals with a view to prevention as the Chapmans apparently are aiming at.

In general, in both the psychological tests as well as in the diagnostic arena, there has occurred a revolution which has broken through the mould of previous research and has provided reliable techniques. How valid they are requires another breakthrough. Content validity is relatively easy to get at, and has already been achieved in the interviewing area. Concurrent validity is really a form of reliability, and this type of validity has also been achieved. Predictive validity in psychopathology will have to wait for the development of good criteria for outcome which are now begging to be developed. How to integrate outcome as viewed by the patient, by his family, by the therapist, by changes observed in testing and interviewing, at the end of the episode, by the milieu he returns to, and by the system of health care is indeed a challenging question.

Even with today's fallible criteria of outcome, we nevertheless can point to what amounts to a law of prediction in psychopathology. Good premorbidss tend to have a good outcome whereas poor premorbidss have a poor outcome. This has held true since prognostic studies began. A survey of all prognostic
studies in the world literature as of 1960 (8) indicated that of the approximately 300 studies (296 to be exact) dealing with premorbid personality, good premorbs had a good outcome in all but 2 studies, and these were in the area of psychosexual development, and 10 studies were neutral with regard to outcome. If those with sudden onset and those with shorter duration of illness are added, the number of favorable-to-good outcomes rises to 708 out of 740, with only 21 studies with unfavorable outcome and 11 studies with neutral outcome. One explanation that can be provided for this law is that schizophrenia is not a continuing but an episodic disorder occurring in vulnerable individuals when they are subjected to internal or external stress-producing life events. Only their vulnerability persists, like an allergy. But all patients recover from their episodes. When they do, they revert back more or less to their premorbid status. If they had a good premorbid status, and return to it, they are regarded as having a good outcome. If they had a poor premorbid status, they also return to it, but are regarded as having a poor outcome because they still can not cope adequately (5).

Others are so vulnerable that they recover only briefly before entering another episode, the brief interlude of recovery going unnoticed. There is probably another group of patients who have actually recovered but because of iatrogenic and hospitalization effects still appear to be in their episode. Finally, there may be a minority who remain continuously chronic, but according to recent studies (7), they are a rapidly vanishing group. This may be why, as Dr. Singer reports, the Rorschach Prognostic Rating Scale correlates negatively with severity of personality disorder and positively with educational level and good premorbid history. Here too, the good premorbs tend to have the good prognosis.

The great need now is to develop construct validity for our tests. The attempts at providing measures of anhedonia and of body image aberration are steps in this direction. We need more tests of this variety perhaps in the area of speech analysis, neurophysiological behavior, and other behaviors to see if some of the concepts abounding in psychopathology regarding diagnostic groupings can be validated. Could one provide, for example, an objective validation of Schneider’s first rank symptoms? Look what REM did for the subjective clinical concepts regarding sleep. Can we provide objective indicators accompanying thought broadcasting, or feeling of being controlled? It is a challenge well worth considering.

The innovative approach of Dr. Singer to the consensus Rorschach is a good example of a productive approach to the construct validation of deficiency in communication in schizophrenics and their families. I wonder whether the recent critique of the success with which Dr. Singer can identify by their Rorschach protocol the parents of schizophrenic children could not be resolved by analyzing a consensus Rorschach in which both parent and child participated for a series consisting of parents with normal children and parents with schizophrenic children.
Perhaps the most insightful evaluation of the role of the Rorschach in detecting schizophrenia comes from the Bürgholzli where Rorschach himself worked. In a follow-up of 208 schizophrenic probands and their relatives, Manfred Bleuler tried to use the Rorschach to determine whether some of the clinically normal relatives of schizophrenics were latent schizophrenics and would yield protocols which best resembled the protocols of schizophrenics themselves and whether the resemblance to schizophrenic protocols was not found in a control group consisting of relatives of brain-disordered patients (organic brain syndrome cases who had no blood relatives with schizophrenia). This study is reported in a German volume entitled The Schizophrenic Mental Disturbances in the Light of Long Term Patient and Family Histories (1), which is now being translated at Yale University. The present report is based on a translation by David Zubin of a section written by A. Uchtenhagen entitled “Schizophrenia-like Rorschach results with the blood relatives of schizophrenics.”

Uchtenhagen, who conducted this study for Manfred Bleuler, found the following:

1. In agreement with earlier studies, about a quarter of the blood relatives of schizophrenics react like schizophrenics on the Rorschach even though they are normal clinically.

2. Clinically normal relatives of brain-disordered patients who showed psychotic behavior without being schizophrenic also showed about 25% of their protocols to be schizophrenic-like, no different than the protocols of blood relatives of schizophrenics.

3. It is concluded that the presence of a schizophrenia-like Rorschach protocol does not mean the presence of schizophrenia in the subject. The Rorschach test is not a definitive diagnostic device for the detection of a schizophrenic psychosis.

4. Since about a quarter of the blood relatives showed abnormal protocols and at best only 6 to 12% of blood relatives develop schizophrenia phenotypically, a prognosis based on the abnormal Rorschach would be wrong in from 19 to 13% of the cases (or 14 to 8% when abnormal but nonschizophrenic personalities are removed from the sample).

5. The descendants of schizophrenia-like reacting subjects did not develop schizophrenia more often than the general population—hence, such parents can not be regarded as providing a schizophrenogenic milieu, nor should they be counseled not to propagate.

In view of these negative findings regarding the usefulness of the Rorschach in detecting latent schizophrenia, Uchtenhagen asks, Does the presence of a schizophrenic-like test result mean anything at all?

His answer is rather illuminating. He dismisses the possibility that the
schizophrenia-like pattern is an unspecific reaction to the test situation by both schizophrenics and nonschizophrenics because of the high frequency with which such patterns are found in genuine schizophrenics and because there are many points of contact between the indicators in the schizophrenic-like response pattern and individual clinical symptoms such as stupor, unpredictable variation in control of reality, flightiness, and perception of cruel, sinister, cold world. He also dismisses the possibility that the deviant pattern results from the fear of being found mentally ill like the sick relative because although the fear is detectable and often directly expressed, its presence did not differentiate between those who showed schizophrenic-like patterns and those who did not. He similarly dismisses entanglement in a guilt-ridden fashion with the fate of the sick relative.

Uchtenhagen offers the following interesting hypothesis:

“The presence of a schizophrenia-like test result can mean that the subject in question carries the possibility of schizophrenic experience and behavior in him, but that these do not show up in his accustomed niche in life. This possibility manifests itself in him under a specific key stress, namely, the confrontation with the Rorschach test. Just as patients with a so-called schizophrenic reaction to an unbearable conflict specific for them decompensate psychotically (the conflict is more clearly recognizable and the frequency of psychoses in the family clearly smaller than in schizophrenia proper), so the "schizophrenic reaction to the Rorschach test" is the psychosis-like breakdown in face of the specific unbearable conflicts activated in the projection test (here the conflicts are clearer and the familiar concentration of psychoses is even smaller than with the schizophrenic reactions). The unconscious conflicts activated in the projection test mobilize fears and awaken defense mechanisms, which otherwise remain hidden under a more or less well-adapted behavior, and do not result in delusions and flightiness, as in the clinically ill.

The unconscious conflicts, the activation of which can lead to the picture of a schizophrenic psychosis or a schizophrenia-like attitude on a test, are by no means specific for the illness schizophrenia. They do not correlate with clinical pictures of illness, and often demonstrate to the contrary the closest similarity with the condition of normals. It is, therefore, to be recommended that the interpretative attitude of the subjects be correlated less with their social behavior than with their inner state of conflict. The apparently equal interpretative attitude of psychotics and the clinically normal clearly attests to this. Rorschach research will profit more from this than from an endeavor toward the diagnosis of illness pictures, which are clinically defined in terms of symptoms. Rorschach diagnosis will be able to ascertain its place in clinical practice and in the theory of schizophrenia all the more as it contributes to the understanding of inner states of conflict, rather than engaging in the prognosis of behavior.”

In other words, at least some of the unaffected relatives of schizophrenics
are vulnerable to schizophrenia to a sufficient degree so as to react to the
Rorschach situation stressfully in a schizophrenic-like manner. Perhaps their
vulnerability is of such a low degree that ordinary life exigencies and chal-
lenges are insufficient to elicit a full-blown episode, but their vulnerability
is revealed by the miniepisode evoked by the Rorschach challenge. Further-
more, the schizophrenic-like pattern of responses seems to reflect a state-
rather than a trait-related behavior. Perhaps retesting under states in which the
Rorschach is not such a stressful challenge such as would be provided by the
more relaxed situation of vocational guidance or school counseling may yield
a more normal Rorschach. That the role of the contingency in which the
Rorschach (or any other test) is given is important has been demonstrated
in many instances, but Bleuler’s demonstration of the dependence of the re-
sults on the situation is most striking.

Dr. Gittelman’s finding that the Rorschach discriminates between normal
children and those with “emotional problems” may be similarly explained
as reflecting a state rather than a trait which is elicited by the Rorschach
situation. Her doubts about the utility of this finding may reflect this possi-
ability. Her generalization about the failure of tests in prognosis may reflect
the possibility that state-induced test results may be good for noting clinical
change, i.e., serve as markers for the presence or absence of episodes but
not good for traits related to prognosis.

Regarding her discussion of the continuity relationship between person-
ality and psychopathology, it may hold true only for disorders in children
whose personality has not yet developed fully so that it is difficult to separate
the interaction effect between the two. It probably does not hold true for
adults, as I indicated earlier.

SUMMARY

A quarter of a century ago, in June, 1950, the 40th annual meeting of
this association was devoted to the topic: Relation of Psychological Tests to
Psychiatry. The foreword to the 1950 symposium published in 1952 (3)
reads as follows:

The use of psychological tests in psychiatry is so widespread today that
evaluation of their actual and potential contribution to diagnosis, prognosis
and treatment is long overdue. With this in mind, the American Psycholog-
pathological Association invited a select number of experts in this field to
discuss critically the role of tests in psychiatry. This volume summarizes the
results of these discussions.

A review of these contributions indicates that there is a great diversity of
opinion regarding not only the clinical usefulness of the tests, but their
nature, purpose and scientific value. Nearly all the contributors would like
to see these tests improved. Some are hopeful that the quantification of these
technics would lead to better understanding of the mind of the normal as
well as that of the abnormal. Others do not hold out so much hope for them,
and regard them at best as ancillary tools. That the intelligence tests have provided a scientific basis for the measurement of mental function in school children and in the feebleminded is generally accepted. That they have served well as screening techniques and guides in selection of men for military, vocational and scholastic purposes is also generally accepted. That they can prove to be as useful in the field of mental disease and epilepsy has not yet been fully demonstrated.

Personality tests of both the inventory as well as of the projective type arouse the greatest diversity of opinion. According to some investigators they are on a much lower level of scientific development than the intelligence tests. While their clinical usefulness is undeniable, their scientific accuracy and precision leaves much to be desired. Others regard them as far superior to any of the other psychological technics and, although they are aware of their present shortcomings, hold out a bright future for their development.

Heretofore, the chief use of these tests has been for diagnostic purposes. Their use in the evaluation of therapy has only now begun and their use for prognostic purposes is in its early stages.

The encouraging progress made in the treatment of mental disorders by the somatotherapies as well as the psychotherapies, and the rather rapid alteration in patient behavior observed under and after these therapies has provided proving grounds for a crucial investigation of the reliability, validity and prognostic values of these tests. Some of these tests have proved their worth; others have been found wanting. Since the changes which these therapies bring about are chiefly in the sphere of affect, it is highly desirable for test makers to turn their attention to the development of this neglected area in psychological test construction. While waiting for such tests to be developed, it might be well to sharpen the interview technic, provide it with more scientific checks and counterchecks and standardize its procedures and the evaluation of its contents. Armed with the probe of the interview, the new tests that are being further developed will have a ready touchstone for their evaluation.

No apology need be made for the limited number of tests discussed in these papers. To have included all would have strained to the limit the publisher as well as the reader. It is hoped, however, that a sufficient sampling has been presented to give a fair cross-section of present day practice. (3)

What progress has been made since then?

1. One advance projected in 1950 has already taken place. We have sharpened the interview and converted it from a blunderbus to a sharp shooting rifle. The hope that this interview can become a touchstone for validating psychological tests can now be fulfilled.

2. Progress in personality tests of the inventory type and the projective type has been only modest in comparison. The provision of special diagnostic profiles for the MMPI is a notable advance, but they have not yet been generally accepted in diagnostic practice. In projective techniques, the tendency to regard them not as tests but as interviews is proving to be useful since the content analysis of the protocols correlates with diagnostic interview results.
3. The promise of prognostic tests has not yet paid off, but here again the improvement in diagnosis may lay the foundation for better prognosis.

4. The use of tests in evaluating outcome of treatment has turned out so poorly that rating scales had to be substituted.

5. The hope that tests of affect would be developed has not come to pass, and here again, rating scales have taken over.

6. The distinction between trait-related and state-related behavior has come to the fore, and this powerful distinction helps to explain why it is important to know the conditions under which test results were obtained. The impact of the test situation on the results obtained with the Rorschach by Manfred Bleuler on blood relatives of schizophrenics is a case in point.

7. The role that the premorbid personality plays in directing the expression of the focal disorder into a unique illness and the role that it plays in outcome are important developments. This helps explain why the cross-sectional test alone, without knowledge of the premorbid personality, may be of little help in either diagnosis or prognosis.

8. The discovery of test markers or other types of markers that will indicate the beginnings and ends of episodes, as well as markers that persist regardless of the presence or absence of episodes (markers characteristic of vulnerable individuals), is the next development to be expected. These markers can serve to validate diagnosis.

9. With regard to diagnosis, it seems that we have licked the problem of reliability with regard to the interview and the diagnostic categories. With regard to the validity of the interviews we have not been so fortunate. It is true that the response to lithium can serve as a confirmation of a diagnosis of manic-depressive psychoses and that similarly the effectiveness of some other drug treatments bolsters our belief in our diagnostic accuracy, but diagnosis by outcome has proved dangerous in the past.

10. Construct validity studies is the path we must take to consolidate our gains, and the Chapmans have shown us a good example in their attempt to establish the constructs of anhedonia and body image aberration for schizophrenia. At the present time there seems to be more activity and progress in the development of psychophysiological and neurochemical approaches to construct validity of our diagnoses than in the personality inventory and projective techniques approach. What we need to find are objective indicators of the presence of a given disorder by establishing the presence of markers which reflect the constructs on which the diagnoses rest.

11. Much of the literature on the role of psychological tests in diag-
nosis can be faulted on the ground that most of the tests indicated poorer performance in mental patients and few if any of the studies controlled for differences in motivation and interest between patients and normals. In order to eliminate the effect of these spurious influences, techniques in which patients excel or in which motivation is cancelled out need to be found (6).

12. The use of forced choice methods and of signal detection theory techniques in the investigation of psychopathology has indicated that many of the differences between patients, especially schizophrenics, and normal controls reflect not differences in sensitivity to incoming stimuli, but differences in the criterion which the patient adopts in his decisions regarding reporting or not reporting what he perceives. Thus, differences in critical flicker function between schizophrenics and normals have been found to be due not to differences in sensitivity to flicker, but to the reluctance of schizophrenics to report the transition from flicker to fusion until they are very sure of their judgment (2). For this reason many of our previous findings are suspect and need to be replicated with forced choice or signal detection methods before they can be of use in the construct validity of diagnosis.

REFERENCES


