Chapter 2

Use of Research Instruments in Psychopathological Assessment: Some Historical Perspectives

JOSEPH ZUBIN, Ph.D.

Chapter 2

Use of Research Instruments in Psychopathological Assessment: Some Historical Perspectives

The beginnings of classification, assessment, and measurement of human behavior including psychopathological behavior are hidden in the dawn of history. There are, however, bits of cultural evidence in customs and laws indicating that at least deviant behavior was recognized and sanctioned early in the development of civilization. But even more than recognizing the presence of the deviant behavior, tests to establish the deviation were also applied quite early, e.g., the Shibboleth test (The Holy Bible) at the fords of the Jordan in which the Ephraimites were identified by their inability to pronounce Shibboleth correctly, and thereupon were slain; and the Talmudic test for mental competence consisting of the placement of a nut and a shaving of wood in front of the suspect to determine which would be selected. If the suspect chose the latter, he or she would be judged as incompetent.

Before quantitative assessment began, poets and literary phenomenologists were the assessors of human personality and

---

The preparation of this paper was supported by the Medical Research Service of The Veterans Administration.
life experience. Ovid, Cicero, St. Augustine, Shakespeare, Milton, and Freud have provided descriptions of personality that even our best instruments today cannot yet reach. The objectivity we relish today has its price in the loss of the rich content of the poet’s descriptions.

Perhaps the earliest quantitative assessment of personality, including psychopathology, was through astrology dating back 25 centuries in Mesopotamia. Palmistry began 3,000 years ago in China. Handwriting analysis came much later, and phrenology came in the middle of the 18th century in France—pioneered by Gall. These pseudosciences are still part of folklore and have left scientific impacts. Thus, astrology was probably the forerunner for the studies of Locus of Control; palmistry and handwriting analysis were the forerunners of identification of individuals through palm prints, fingerprints, and signatures; and phrenology was the forerunner of brain localization. Divining personality through animal entrail examination, cloud reading, and flame inspection were the forerunners of projective techniques.

Assessment based on handwriting analysis is another art that goes back to earliest times. Quite a literature has developed describing the meanings to be attached to various elements in a handwriting specimen, including such characteristics as length of the middle zone of the letters, width, slant, curvature, and pressure on the writing instruments, which was usually interpreted as reflecting instinctual drive.

As for handwriting, I undertook to evaluate it at the behest of Dr. Nolan D. C. Lewis, then the Director of the New York State Psychiatric Institute, because a graphologist named Thea Stein Lewinson had analyzed the handwriting of some of her patients and reported apparently to his satisfaction their personality characteristics based only on the inspection of the handwriting samples. As a result, we developed some objective measures of handwriting characteristics of high reliability, but could not establish their validity (Lewinson and Zubin 1942). How did classification, assessment, and measurement of
Use of Research Instruments

behavior get their start? Classification is the way infants get a handle on the blooming confusion that the world presents. To live in the world, they must classify, identify, and evaluate their surroundings, just as the anthropologist entering a new culture must learn how the natives perceive and organize their world. The classification of the tribal connections in the form of kinships is one of the first classifications of human beings. As for objects, Aristotle was one of the first to establish categories for the classification of objects and insisted that each category had an essence that must have been shared by the members in the category. This establishment of qualitative categories permitted comparison between objects in the form of the positive, comparative, and superlative levels of adjectives, e.g., red, redder, reddest.

But how did the categories of objects arise, and how was class membership determined? Perhaps the judgments of similarity and difference were the underpinnings of the classification system. The judgment of similarity between objects may have rested on the criterion of whether the two objects served the infant equally well for the task at hand. If the string and the long-handled fork were equally good for fetching a toy, the judgment of similarity applied. If only one of two objects suited the purpose of the individual, the objects were regarded as different. However, as development continued, the self-referred criterion of similarity gradually was replaced by more objective external criteria independent of whether the objects under scrutiny depended on self-reference as a basis. Thus, an ounce of gold and an ounce of coal are regarded as equal in weight even though self-reference would make one prefer the gold.

As a matter of interest, the psychological rationale for sorting tests, such as the Goldstein and Scheerer Sorting Test used in differentiating schizophrenic subjects from normal subjects, may have arisen on the basis of infantile self-reference, which the adult schizophrenic may regress to when the disorder develops (Zubin and Thompson 1941).
Measuring Mental Illness

But how did the categorical qualitative judgment of similarity and difference lead to measuring scales? Here again, history has no record of the development. However, we can safely assume that even before the scales for measuring such physical characteristics as warmth, weight, and length developed, they were measured subjectively. There are no data to substantiate this suggestion, but we can use our imagination in describing how it may have happened.

Let us look at the measurement of subjective warmth. Because there is no documentary evidence for initial attempts by human beings at gauging the cause and degree of subjective warmth, we shall resort to fantasy. Come with me to a cave in some prehistoric ice age, before Prometheus, and listen to a symposium on the origin of the experience of subjective warmth. One savant declares that warmth depends on the number of skins covering the body. Another claims that it depends on duration of exposure to the sun. A third postulates swift running as the source of warmth and the distance covered as a measure. The medicine man in their midst raises a controversy because of his claim that his patients often report experiencing warmth without the benefit of skins, sunshine, and running. The symposium ends without a resolution. But Prometheus later discovers fire, and it becomes possible to demonstrate that by adding faggots to the fire, the experience of warmth is raised in all the inhabitants of the cave and, by the same process, removal of faggots reduces it. The first breakthrough has occurred. Humans can manipulate an external agent to raise and lower subjective warmth at will. But this is still far from measurement.

The first known historical breakthrough occurred in ancient Egypt, where after the linguistic development of ratings in the grammatical form of positive, comparative, and superlative adjectives (warm, warmer, warmest) took place, rating scales were developed for measuring warmth in four steps, with warmth anchored at one extreme to the hottest day of summer and at the other extreme to the coldest day of winter.
Eventually, the expansion of mercury with increase of experienced warmth was noted, and the thermometer was born, and finally, humidity and air pressure were recognized as important factors, and the present-day discomfort index emerged.

What were the essential steps in the process? First, the discovery of means of inducing changes in a sensory experience by external control (fire). Second, the development of an external criterion for measurement independent of the self-referred subjective experience (mercury). The same process, no doubt, held true of the other measures such as weight and length, which have attained great objectivity. Pain is still a subjective phenomenon without an external criterion, and without measurable ways of inducing it, though recent efforts along these lines show considerable promise. Intelligence tests became objective when mental age scales were substituted for teachers' subjective impressions. Anxiety, depression, and elation are still in the rating stage. In my fantasy, I sometimes imagine that we may find a life-bearing planet, somewhere in space, where anxiety has already been measured, but where warmth or length are still on the intuitive level (Zubin 1965a).

**Early Development of Modern Rating Scales, Tests, and Interviews**

Two traditions have developed in assessment: the psychometric and clinical traditions. These two trends differ essentially in the contrast between the dimensional and categorical approaches. The dimensional approach arose from the psychometric tradition and the categorical from the clinical tradition.

It might be best to pause here for a moment and inquire as to which approach—the categorical (typological) or the dimensional—is the truest. By the dimensional approach I mean the assumption that types do not exist and instead the hyperspace of personality variables has continuous hypersurfaces without hills or valleys, or multimodal distribution, or discontinuities.
I had raised this question in 1968 (Zubin 1968) and as a result of the heated discussion which ensued, could not fall asleep that night. I turned on the TV and caught the Farmer's Hour with a program on pomology. The reporter indicated that historically apple knockers would test apples by applying their knuckles to the surface of the apple to test for maturity and water core. This typological approach of separating good from bad apples has been replaced by a conveyor belt that passes the apples under two sets of double monochromatic lights, and the amount of light transmission determines whether the apples pass as acceptable or nonacceptable. Thus, the typological test has been replaced by a dimensional test. However, when one considers that the genes of the apple may determine its acceptability we again return to a typology of alleles. But since the alleles secrete some amino acids, we can measure their quantity and return to dimensional tests again. However, the amino acids consist of patterns of molecules and these again yield a typological basis for discrimination and so on ad infinitum. Apparently, the struggle between typology and dimensionality is based on a pseudoproblem dependent on the state of the art at any given time.

Rating Scales and Personality Inventories

Following the pattern of the ancient Egyptians in providing a rating scale for the measurement of warmth, similar attempts form the basis for the ratings of human behavior. Thus, ordinal rating scales were the initial steps. The first recorded attempt was by Heymans and Wiersma (1909) who drew up an inventory of symptoms through a self-reporting instrument. The early history of the development of rating scales and personality inventories can be found in McReynolds and Ludwig (1987, 1984). These early inventories paralleled the attempts by Adolf Meyer at describing patient behavior (Zubin and Zubin 1977). This effort at describing and rating eventually led to the Woodworth Personal Data Sheet (Zubin
Use of Research Instruments

1948), the forerunner of modern personality inventories, which finally led up to the Minnesota Multiphasic Personality Inventory (MMPI).

Jung introduced measurement of emotional complexes through the Word Association Test, and Kraepelin introduced psychological testing to help in diagnosis in the following words:

As soon as our methodology has sufficiently proved itself through experience with healthy individuals, it would be possible to approach the actual ultimate goal of these efforts, the investigation of the sick personality, especially of the inborn pathological disposition... We, therefore, have first of all to investigate whether it is possible by means of psychological tests to determine individual deviations, which cannot be recognized by ordinary observation. If that succeeds, we would be in the position through the quantitative determinations at our disposal, to establish the borderline between health and disease much more precisely and more validly than has been possible so far. (Kraepelin 1896, p. 77)

Measurement of psychopathology lagged behind the measurement of psychometric traits primarily because the changes in patient behavior were not sufficiently large to observe. Patients stagnated for the most part, and static description of their behavior did not require rating or measurement. However, with the arrival of somatic therapies, changes in behavior were so striking and rapid that a demand arose for providing such measures.

The speed of the behavioral changes under the influence of drugs was so dramatic that it challenged the therapist to provide measures of the degree of change and the corresponding relation to dosage. For the first time in psychiatric history, psychotic behavior could be altered before one’s eyes, and the need for measurement and evaluation became apparent. This caught the measurement experts—mostly psychologists—unprepared, since all the available clinical tests and tools provided only static descriptive measurements like intelligence
Measuring Mental Illness

tests or projective techniques geared to constant-trait rather than varying-state measurement. Rating scales had to be pro-
duced practically overnight by such pioneers as Lorr, Witten-
born, and Malamud. There is a story, probably apocryphal, about
the physiologist Hudson Hoagland, who found some
biochemical change in Malamud’s patients at the Worcester
State Hospital and asked to see Malamud’s data on the same
patients. Malamud, who was a leading clinical psychiatrist,
handed Hoagland his voluminous case histories. “Are there
no numbers here on their behavior,” Hoagland inquired, to
which Malamud cried, “No, but if you want numbers I’ll make
them up for you.” That night, with the help of a colleague
named Sands, he converted his descriptions of behavior into
rating scales (Malamud et al. 1946; Malamud and Sands 1947).

The success of drug therapies also required better
classification of patients in order to obtain more homogeneous
groups. To fill this need, the free-floating clinical interview
had to be converted from a blunderbuss into a sharpshooting
rifle. Systematic semistructured and structured interviews re-
sulted. Their use in the United States–United Kingdom (U.S.
U.K.) Diagnostic Project and in the World Health Organiza-
tion (WHO) International Pilot Study of Schizophrenia were
the proving grounds in which these instruments were fire tested.
These projects aroused diagnosis from its academic lethargy
and finally led to the development of such instruments as the
Schedule for Affective Disorders and Schizophrenia (SADS),
the Diagnostic Interview Schedule (DIS), Research Diagnostic
Criteria (RDC), and the Diagnostic and Statistical Manual of
Mental Disorders, Third Edition (DSM-III) classification sys-
tem.

It must be pointed out, however, that the revolution in
psychopathology was not due to pharmacological intervention
alone. Even before the drug era, the opening of the doors in
European hospitals, especially in England, had brought about
better outcome results (Shepherd 1988). But, the dramatic effect
of the drugs electrified the field.
Use of Research Instruments

Tests

Perhaps the best example of measurement in human behavior is the intelligence test. Before Binet, attempts were made to rate intelligence through adjectival ratings. Eugen Bleuler, at the turn of the century, would ask his residents to assess the intelligence of a new patient. One resident began his questioning by asking a female patient where she came from and how she got to the clinic. Bleuler interrupted scoldingly with "never ask a lady about geography, she has no sense of it." (Remember that this was 1900 in Switzerland, a country that even now has denied women votes in some cantons.) Apparently Bleuler laid down the first requisite of a test: it should deal with knowledge that the patient possesses or should possess. As in my discussion of the development of temperature scales, it is important to ask, what are the criteria that the test would satisfy? There has to be some way in which we can manipulate the trait under investigation. We cannot, of course, vary intelligence through manipulation but can instead use variation in chronological age as a quasi-manipulation, since it is clear that intelligence, at least during the developmental period, varies with age. That is how the concept of mental age was born. Thus, instead of varying the subjective experience, as was the case with subjective warmth, the natural variation in intelligence with age was used to manipulate the degree of intelligence. The external criterion for intelligence corresponding to the expansion of the mercury column in the case of subjective warmth, was school progress, which presumably varied with intelligence and thus served the ends of construct validity.

The Interview

I have pointed out elsewhere that the clinical interview is the psychiatric tool par excellence for arriving at a diagnosis and the conduct of psychotherapy (Zubin et al. 1985). None-
theless, until the 1960s little research effort had been directed at investigating the reliability, validity, and usefulness of this basic tool. The interview dates back to antiquity (Zubin 1965b; Mattarazzo 1965). Interview-like conversations consisting of questions and answers are found in the Bible. The Socratic method of interviewing in the search for truth is well known. Soranus formulated rules for conversing with mental patients, recommending that laborers should be engaged in conversation about cultivation, sailors about navigation, and so on (Zilboorg and Henry 1941). Interviewing techniques have varied with the zeitgeist. The ancient Greeks regarded mental disorder as due to either perversity, possession, or drug-induced states, and verbal utterances were regarded as random effects that were of little or no interest to the physician. Interest in interviewing languished during this period. During the Moorish period in Spain, diagnosis was based on mystical and intuitive approaches, and the patient’s utterances were interpreted by seers on the basis of these criteria rather than on the basis of the patient’s needs and difficulties. Interest in the patient’s attitudes and beliefs rose in the Western Caliphate (Seville and Cordova) when the problem of personal responsibility came to the fore. During the Spanish Inquisition the accused were interrogated to establish their heresy, and the guide for enticing heretics to admit guilt was formulated as perhaps the first systematic interview (Lea 1888). With the rise of the idea of personal responsibility, implying that only those who are mentally competent can be punished, there arose a need to interview the patient in order to assess thoughts and feelings. In the early seventeenth century, Bacon inquired whether the patient knew his own age and name, thereby initiating interviewing with regard to time and orientation. Freud first introduced suggestion and hypnosis to induce patients to speak more freely before developing the psychoanalytical interview, in which all constraints on the conversation were removed and free associations were encouraged, thus permitting unconscious as well as conscious communication.
Use of Research Instruments

Certain criteria separate the interview from ordinary conversation or communication. For the purposes of this chapter, we can differentiate it from other types of human encounter as follows: 1) it consists of a meeting of at least two individuals, the interlocutors, face to face; 2) it is directed at a particular purpose by either one or all interlocutors; 3) it employs conversation and observation; and 4) it involves a hierarchical relation between the interlocutors, insofar as one is the interviewer and the other the interviewee.

For a long time the psychiatric interview remained a free-floating conversation between clinician and patient, and though this is still the vogue today, the increasing use of the interview for research purposes has led to a considerable metamorphosis of this communication technique. The impetus for improving the interview was provided by several factors, including the need for yardsticks to assess the rapid changes due to pharmacological treatment, and by individual studies like the bi-national U.S.-U.K. Diagnostic Project (Cooper et al. 1972) and the International Pilot Study of Schizophrenia (World Health Organization 1973).

The aims of the psychiatric interview are fourfold; first, to elicit (as far as possible) information about presenting symptoms and the antecedents of current episodes in order to cast light on possible causes; second, to determine the patient's feelings and attitudes about his or her current status and symptoms; and, third, to note the pertinent nonverbal behavior, which can help establish the nature of the problems. In addition, because an episode of mental illness reflects a disturbance in interpersonal relationships, the behavior sampled during the interview offers information on the nature of the disturbance. The degree to which the interview taps these main sources of information is a measure of its success, which is why psychiatric interviewing cannot be left to the novice or be based on computers alone. While sophisticated systematic interviewing can elicit factual information, the more subtle attitudinal factors, the assessment of interpersonal relationships, and the
Measuring Mental Illness

detection of abnormal nonverbal signs require a flexible give-and-take possessed only by experienced clinicians.

The proven unreliability of the early interviews prompted attempts to develop self-rating inventories and scales designed to be filled out by the patient, and checklists and rating scales to be completed by the interviewer. Self-reporting instruments have the advantage of eliminating interviewer bias and interpretation by putting the onus on the individual’s response rather than on the response of the interviewer to the patient’s behavior. This strategy does not, of course, guarantee a better result, and obvious self-reporting inventories have generally not been found to be useful for the classification of inpatients, although they have been found quite useful in dealing with minor disorders encountered primarily among student populations and in surveys of the general population. In a notable study, Dohrenwend and his colleagues (1979) found that the common denominator in such self-reports seems to be related more to Frank’s (1973) demoralization factor than to clinical psychopathology, a factor which related more to personality deviation than to clinical psychopathology.

The checklists and rating scales used by interviewers did not utilize standardized methods for obtaining the information needed for the ratings, and consequently results were not always comparable from patient to patient. To obtain a more systematic method for the collection of the information, a new clinical interview with a systematic structured format was developed. This began with an attempt to use a checklist for the Mental Status Examination used by the residents and staff of the New York State Mental Hygiene Department. This “redbook,” as it came to be called, underwent a series of revisions by each succeeding Director of the New York State Psychiatric Institute, and when Dr. Nolan D. C. Lewis became director, he invited me to help with the revision. I introduced a systematic checking schedule for recording the presence of the symptoms observed during the course of the examination. World War II interfered with its application, but soon after the end
Use of Research Instruments

of the war, Drs. Burdock and Hardesty took over this task and in 1964 with the help of two of their trainees—Robert Spitzer and Joseph Fleiss—prepared the first in the series of systematic semistructured interviews from which the SADS finally emerged. The revision of DSM and provision of RDC followed. A first step was to standardize the usual mental status examination (Spitzer et al. 1964). The questions to be asked of the patient were then specified in a proper sequence and directions for coding the responses were laid down. Thus, the loose mental status examination, formerly the psychiatrist’s mainstay, was converted into a systematic, structured interview which yielded high reliability in the scoring of items and carried some degree of validity.

This technique has undergone a series of modifications (Zubin et al. 1985). Three types of interviews were developed: 1) the nonprobing Structured Clinical Interview (SCI) (Burdock and Hardesty 1969); 2) the medium-probing schedules, the Mental Status Schedule (MSS) (Spitzer et al. 1964) and the Psychiatric Status Schedule (PSS) (Spitzer et al. 1970), and 3) a deep-probing Present State Examination (PSE) (Wing et al. 1967). These interviewing methods have demonstrated their value in the two major international studies mentioned previously.

The U.S.-U.K. project was initiated to determine why the national statistics show such disproportionate frequencies of affective disorders in the United Kingdom and of schizophrenia in the United States. When the newly developed systematic interviews (Gurland, cited in Zubin et al. 1975) were applied to samples of patients admitted to hospitals in the two countries, the cross-national differences turned out to reflect different diagnostic practices of psychiatrists rather than differing characteristics of patients in the two countries. Whereas the U.S.-U.K. project limited itself to two cultures but investigated a spectrum of mental disorders, the WHO study of schizophrenia limited itself to one disorder but examined its incidence, form, and course in nine different cultures. The WHO
Measuring Mental Illness

study found specific syndromes of schizophrenia ubiquitously distributed in both developing and advanced cultures from Ibadan, Nigeria, to Washington, D.C.

Wing (1978) has pointed out that, as a result of these two international studies, the clinical symptoms associated with schizophrenia can be categorized in two syndromes, positive and negative. The positive syndrome consists of delusions, hallucinations, incoherent speech, and other florid and productive phenomena. These symptoms tend to be responsive to treatment and to remit at the end of the episode. The negative syndrome (including the so-called clinical poverty syndrome) is composed of the symptoms of social withdrawal, emotional apathy, slowness of thought and movement, underactivity, lack of drive, and poverty of speech. These more often characterize the chronic state and are not generally affected by drugs. Strauss and co-workers (1974) have added a third syndrome category, disorders of relating, which seem to be of ontogenetic origin and may reflect such deviations as absence of intimacy in schizophrenics’ premorbid friendship patterns during adolescence (Kreisman 1970). The contrast between positive and negative symptoms reflects a striking comparison in the behavior of patients and normal individuals. Positive symptoms denote behaviors in which patients engage (hallucinations, delusions) but normal individuals do not, while the negative symptoms denote behaviors that normal individuals engage in (social interaction, emotional warmth) but patients do not (Zubin 1985a).

Personality

Following the paradigm described for the development of measurement in the physical sciences and psychological tests, attempts were made to devise techniques for measuring personality and psychopathology. Unfortunately, means for utilizing the first procedure described earlier in measuring subjective experience—manipulation of the subjective or internal phenomena—was not always available for personality as it was
Use of Research Instruments

in the case of temperature and intelligence. The second procedure, that of establishing an independent criterion (the expansion of mercury in the case of subjective warmth, chronological age for intelligence), was also rarely available. For these reasons, progress in tests for personality and in psychopathology has been slow. Perhaps the most successful test for psychopathology is the MMPI. The varying of the degree of the subjective phenomena under measurement was accomplished by studying contrasting groups of diagnostic categories including normalcy. The difficulties presented by this approach inhere in the fact that the clinical categories were based on interviewing, which at the time of the preparation of the MMPI was in a low state of reliability and validity. Perhaps a restandardization based on Research Diagnostic Criteria might help. Construct validity (corresponding to the expansion of mercury) is also problematical, and very little research has been done along these lines, because we do not have any consensus on what the constructs might be. The problems of developing tests of personality suffer from the same difficulties facing tests of psychopathology, and we are reduced to returning to interviewing as our mainstay.

One of the reasons for the failure of the Rorschach to qualify as a diagnostic instrument is the fact that in the last analysis the scoring of the responses and their clinical significance was originally based on the interview by means of which a clinical diagnosis was reached. We have already pointed out that the clinical interview in Rorschach's day was a weak rod to lean on. However, if the Rorschach protocols themselves are regarded as interviews and analyzed for their content, the resulting dimensions have been shown to correlate with diagnoses because both the Rorschach protocols and the diagnoses are based on interview content (Zubin et al. 1965).

Reliability

It is well known that the greatest advances in science come not from new theories but from new technological inventions.
The development of audiotaping and videotaping have had the most important impact on the study of reliability. I recall that I was once asked by Clarence P. Oberndorf, a noted psychoanalyst, to help him evaluate psychoanalysis. I suggested that he pick out one of his analysands and record several interviews with the patient so that we could together go over the recording and develop a content analysis of the interview to see what dimensions we could discover for evaluation. He bought an expensive Stromberg Carlson recorder and started recording his interviews. After several months he again invited me for dinner and at the end, pulled out a cigar, lit it, and puffed on it and then gingerly pushed the button to start the recording. No sooner had the record started when I noticed that Oberndorf grew uneasy and as time went on his face reddened and finally he rose and turned off the sound, saying: "This never happened—it's impossible." That was the end of our attempt at evaluation.

I had better luck in studying the reliability of the ratings on interviews with the aid of videotapes in our U.S.-U.K. project. Instead of having a second interviewer sit with the main interviewer to get reliability ratings, we offered the videotapes as a basis for reliability studies. Furthermore, groups of raters could be obtained either individually by mailing them the videotapes or we could gather raters at a convention and obtain their ratings simultaneously. In one such experiment, which Dr. Katz and I conducted, we found large discrepancies in the judgment of whether the patient was neurotic or psychotic. An analysis of the ratings indicated that the differential was based on the rating of apathy. Those who gave a high rating on apathy judged the patient to be psychotic, while those who rated the patient low in apathy judged him to be neurotic.

Validity

The history of the search for validity began only when the reliability of our measuring instruments was established. Until the etiology of mental disorders is discovered, our valid-
Use of Research Instruments

ity estimates will have to depend on such criteria as response to treatment or follow-up outcomes, insofar as diagnostic instruments are concerned. The search for “markers” identifying the individuals who have already developed episodes of illness or who are at high risk of developing such episodes is one hopeful way of tackling the validity problem. Several indicators have now been found that are potential markers for vulnerability to schizophrenia in both probands and their unaffected siblings (Zubin 1985b), and there are also some markers for affective disorders and other disorders. These markers may eventually lead us to the discovery of etiology, which will provide the validity criteria we need.

Summary

We have traced the development of research instruments for measurement from primitive times to their current status. The progress in measurement in psychopathology has not been as rapid in psychopathology as in the physical sciences, but, nevertheless, considerable advances have been made. The history of the development of the interview technique was traced, and its current advances with regard to reliability but not validity seem promising. The newly developing search for markers may prove to be of help in providing a firmer basis for validity.

References

The Holy Bible, Judges, chapter 12
Use of Research Instruments


Zubin J: Recent advances in screening the emotionally maladjusted. J Pers 17:141–145, 1948


41
Measuring Mental Illness


Zubin J: Scientific models for psychopathology in the '70s. Seminars in Psychiatry 4:283–296, 1972

Zubin J: Negative symptoms: are they indigenous to schizophrenia? Schizophr Bull 11:461–470, 1985a

