Mental Status Schedule: Potential Use as a Criterion Measure of Change in Psychotherapy Research*

ROBERT L. SPITZER, M.D.† | New York, N. Y.

The greatest obstacle to the evaluation of the results of psychotherapy is no longer the reluctance of psychotherapists to have their work evaluated or the absence of suitable research designs and controls; it is, I believe, the lack of suitable instruments for measuring change in the kinds of psychopathology at which psychotherapy is directed.

I would like to describe the Mental Status Schedule (MSS), an instrument designed to improve the research value of clinical judgments based on data collected during a psychiatric interview. I believe this instrument has considerable potential as a criterion measure of change in psychotherapy research. It is the result of nearly three years of work in the Biometrics Research Unit of the New York State Department of Mental Hygiene at the New York Psychiatric Institute and the Department of Psychiatry at Columbia University. The instrument was developed with the collaboration of Eugene I. Burdock and Mrs. Anne Hardesty, and is one of many instruments developed by our department under the direction of Joseph Zubin for the purpose of objectively evaluating psychopathologic behavior of mental patients.

I shall first describe the instrument and report some preliminary findings on its reliability and validity. I will then indicate some of the potentials and limitations in its use as a criterion measure of change in psychotherapy research.

The MSS is an instrument which provides an easily learned technique for examining, recording, and evaluating the mental status of a psychiatric patient. The patient is examined by means of an interview schedule, which the interviewer uses to systematically survey and probe a wide range of current behavior and symptomatology. During the course of the interview, the interviewer records his judgments of 248 short statements or items describing the patient’s verbal and nonverbal behavior by marking each item as either true or false on an answer sheet (1). The interview schedule is printed on

* Presented at the Third National Meeting of the Association for the Advancement of Psychotherapy, Los Angeles, Calif., May 3, 1964.
† Biometrics Research, New York State Psychiatric Institute and Columbia University.

the left side of each page. The items of the inventory are grouped on the right side opposite the corresponding questions and statements (Fig. 1).

1. Refuses to shake hands.

2. Perspires profusely or his hand is either wet or clammy when shaken.

3. *Gives a description of his behavior which is inadequate or insufficient to account for admission (e.g., just says he got "nervous" or "depressed").

4. Says he has felt elated or "high."

5. Says he feels nothing, has no feelings or feels dead.

6. Says he has no worries or that nothing bothers him.

7. Mentions he worries a lot or that he can’t stop worrying.

8. Admits to three or more different fears or says that he keeps feeling afraid of different things.

9. Mentions a fear of being abandoned or left all alone.

10. Indicates he is fearful of losing his mind or losing control of his emotions.

11. Indicates a morbid fear that something terrible will happen to him.

12. Indicates he has an irrational fear of a particular object or situation (e.g., crowds, heights) [phobia].

13. Says he gets attacks of sudden fear or panic.

14. Indicates his fear prevents him from participating in some activity.

15. Admits that he is bothered by feelings of anxiety.

16. Admits he feels anxious practically all the time.

17. Says he has felt restless or unable to stay still.

18. Admits he is bothered by feelings of sadness or depression.

19. Admits he feels depressed practically all the time.

20. Admits he feels like crying.

21. Mentions he loves himself or that he thinks he is perfect.

22. Accuses himself of being unworthy, sinful or evil.

23. Indicates he is bothered by feelings of inadequacy or says he doesn’t like himself.

24. Indicates he is bothered by feelings of guilt.

25. Indicates he feels hurt or overwhelmed when criticized.

*This section should only be used with patients who have been in a psychiatric ward or clinic; it may be omitted in repeated testing. M9S-1

Figure 1.
The questions and statements of the Schedule are ordered according to a
definite sequence, designed to provide a natural progression of topics with
smooth transitions. The evaluation generally takes from 20 to 50 minutes,
depending upon the patient's verbal productivity and cooperation. After
the standard evaluation, the examiner may continue the interview in any
manner he wishes to provide additional data which may be of special interest.

The patient is first asked some identifying questions, and how long it
has been since he came to the hospital or clinic. Questions are then directed
at the following areas or topics in this order: original complaint, present
problems, mood, worries, fears, anxiety, restlessness, depression, crying, self-
appraisal, response to criticism, interpersonal relations, irritability, anger,
ideation, physical complaints, body image, rituals, sex, fatigue, psychomotor
retardation, appetite, olfactory hallucinations, sleep, thinking, memory,
imagination, perceptual distortions, bodily sensations, plans and outlook for
the future, humor, enjoyment and interest, feelings of punishment, death,
suicidal trends, decision-making, referential thinking, dreams, auditory and
visual hallucinations, delusional ideation, antisocial attitudes or actions, con-
ception of illness, attitude toward treatment, recognition of need to change,
remote and recent memory, orientation, and immediate recall. The patient
is then asked to wait while the interviewer marks a group of 76 items which
describe behaviors that could have occurred at any point in the interview
and which are not associated with specific questions. This group of items
covers the following areas: grooming; attention; quantity, quality, and rate
of speech; emotion; and spontaneous physical behavior. The interview con-
cludes when the interviewer asks the patient, "How do you feel answering
these questions?" and then marks four more items which relate to the pa-
tient's attitude toward the interview.

Many questions are open-ended. For example, "Tell me about your
physical condition?" or "What kind of sex life do you have now?" Some
questions are phrased so that the burden of denial is upon the patient, as in
the question, "What kind of trouble do you have with people?" or "What
kind of disturbing thoughts do you get?" Some questions are worded to
make it easier for patients to admit the behavior in question. For example,
"When a person feels depressed or hopeless he may think about his own
death. Do you?" Other questions are designed so that the patient does not
know what the interviewer is really trying to determine. For example, when
the patient is asked, "What do you enjoy doing?" the interviewer is actually
trying to determine whether the patient feels there is nothing that he enjoys
doing. Similarly, the patient is asked, "What are your plans for the future?"
when it is actually the absence of any plans for the future that the inter-
viewer is trying to determine.

Alternate location for many questions is provided so that the examiner
can ask the question in the form or tense most appropriate to the patient's circumstances or previous responses. For example, several questions allow the interviewer to refer to the patient's circumstances as "illness," "trouble," or "situation." When these questions are reached, the examiner will usually know whether the patient regards himself as ill or merely in trouble or in a difficult situation. The examiner selects the term which seems to correspond with the patient's conception of his difficulties.

There are numerous specific but optional follow-up questions, which are in parentheses. They are used to clarify or probe into areas where the patient seems to have further information related to certain items of the inventory. Where a group of follow-up questions appear together, the examiner chooses those questions which are appropriate to the patient's previous response. For example, if the patient acknowledges thoughts of his own death, the examiner then goes on to the follow-up question, "Do you think of killing yourself?" If the patient responds affirmatively, the interviewer then goes on to the next follow-up question, "Have you thought how you would kill yourself?" Some of the follow-up questions are designed to incorporate words from the patient's previous verbalizations, for example, the follow-up question on page 1: "Does this fear of . . . . . . prevent you from doing something you want to do?" If the patient spoke of a fear of crowds, the follow-up question would be, "Does this fear of crowds prevent you from doing something you want to do?"

In addition to the alternate locution and follow-up questions, the interviewer is always free to make a general request for more information, such as "Could you tell me more," or "I don't understand that." However, he should avoid asking specific questions not contained in the schedule, as this would impair the comparability of results with other interviewers.

The 248 items of the inventory were developed by surveying the standard psychiatric texts and by interviewing several hundred psychiatric patients with preliminary forms of the instrument. Whenever a pathologic behavior was observed that was not described in an item, consideration was given to construction of a new item. As a result, the universe of 248 items describes an extremely wide variety of pathologic behavior, ranging from ubiquitous symptoms such as restlessness and occasional anxiety to symptoms indicative of severe pathology such as paranoid delusions and incoherence.

Most items were reworded many times before they reached a form which seemed to reduce ambiguity to a minimum and to achieve maximum agreement among different observers. The items focus on observable behavior only and do not make any reference to unconscious motivation of processes. Technical terms have been avoided in the body of the items but, where appropriate, appear in italics at the end of the item, for example, item 12: "Indicates that he has an irrational fear of a particular object or situation,
for example, crowds or heights (phobia).” The items describe the elementary observations upon which complex clinical judgments are built. Thus, no single item is reserved for depression, but various items describe behavior which, taken together, comprise the syndrome. All of the items are in a dichotomous form. However, the intensity of some dimensions, such as incoherence, can be indicated by appropriately marking several items, for example, item 190: “Speech is at times disorganized and incoherent;” 191, “Speech is frequently disorganized and incoherent;” and 192, “Speech is for the most part disorganized and incoherent.”

An item is judged true if (1) the behavior described in the item is observed during the interview; (2) if the patient describes the behavior as having occurred at any time during the previous week; or (3) if the behavior, although not occurring within one week, is characteristic of the patient’s mode of functioning or is still a problem for him.

Scoring

One advantage of expressing clinical judgments in dichotomous items is that many different methods of summarizing the primary data contained in the individual items into quantitative indices or scores are possible. The first scoring system used was the total score, which is merely the number of items judged true. Despite the obvious limitations of such a system, which ignores the patterning of behavior, we have found it to be a reliable and valid, although crude, measure of severity of illness.

We have conducted several reliability studies on a wide range of patients, including hospitalized patients as well as patients attending outpatient psychiatric clinics (2). When using well-trained and motivated observers, the reliability of the total scores has consistently been over .90, the mean being .945, using the intraclass correlation as a coefficient of reliability.

The total score has also been found to discriminate between groups of patients known to differ in severity and extent of psychopathology. A preliminary study compared three groups of patients: (1) a random sample of newly admitted patients to the New York State Psychiatric Institute; (2) a sample of patients accepted for outpatient psychotherapy at the Columbia-Presbyterian Medical Center; and (3) a sample of hospitalized patients on the orthopedic service of Presbyterian Hospital who, for the purposes of this study, constituted a hospitalized control group. There were 25 patients in each group. It was anticipated that hospitalized mental patients would tend to have the highest scores, outpatients the next highest, and medical patients the lowest scores. The results confirmed these expectations. The mean score for the mental patients was 31, for outpatients 18, and for the hospitalized control group 10, these being significantly different at the .001 level.
To determine the extent of agreement between the total score and a clinical evaluation of severity of illness, 20 formerly hospitalized patients were each interviewed separately by two psychiatrists; one always administered the Mental Status Schedule, while the other, an experienced clinical psychiatrist, conducted an unstructured interview. Neither psychiatrist had knowledge of the results of the other's examination. The psychiatrist who conducted the unstructured interview ranked the 20 patients on the basis of his over-all clinical impressions of the severity of illness. His ranking was compared with the ordering of patients determined by total score on the MSS and the Spearman coefficient of rank correlation was found to be .81. Thus, the magnitude of total score is highly related to the severity of illness as defined by expert clinical judgment when a wide range of psychopathology is present.

To test the sensitivity of the instrument to changes in the status of patients, 45 adult patients newly admitted to the New York Psychiatric Institute were each administered the instrument upon admission and again ten weeks later. The measure of change was taken to be the difference between the patient's initial and follow-up total score. An independent clinical evaluation of change was then obtained by having the various resident psychiatrists who were treating the patients—the mode of treatment was predominantly psychotherapy—rate on a five-point scale their clinical judgment of the degree of improvement exhibited by their patients over this period. Patients judged clinically to have improved greatly also had the greatest change in score on the average; patients who improved only slightly changed less; and patients judged not to have improved had the smallest mean change—a difference which was significant at the .05 level. However, there was considerable overlap between the groups, which was at least in part due to the different clinical criteria of improvement employed by the different residents.

Duane Denney at the University of Oregon Medical School studied 59 hospitalized medical patients, administering both the Mental Status Schedule and the Cornell Medical Index. Considerable evidence shows that a score of above 30 on the Cornell Medical Index indicates psychiatric disturbance. The average Mental Status Schedule score for those patients with a Cornell Medical Index score of above 30 was 20, whereas the average score for those patients with a Cornell Medical Index score below 30 was only 10. This difference is highly significant. Using the Cornell Medical Index cutting point score of 30 and a Mental Status Schedule score of 10, there was an agreement as to the patient's being psychiatrically ill or well in 72 per cent of the cases.

Further evidence of the validity of the total score has been offered by a study conducted by Bruce Muller of Fairfield State Hospital in Connecticut.
He administered the Mental Status Schedule to 40 hospitalized schizophrenic women and compared their total scores with various subscales of the Minnesota Multiphasic Personality Inventory. The MSS total score correlated highly with several scales, namely the F, Paranoia, Psychasthenia, and Schizophrenia scales, with correlations ranging from .67 to .79. Patients treated by drugs showed a significantly greater drop in MSS total score as compared with a matched group treated with a placebo.

Currently we are working on the development of scales which convert the behaviors expressed in the items of the inventory into scaled judgments of clinical dimensions, which we believe will be far more useful and will supplant the use of the total score as a meaningful index of severity of illness. Twenty scales have been tentatively developed. They are: Decreased Communication, Denial of Illness, Grooming, Perceptual Distortions, Memory-Orientiation, Delusional Ideation, Suspiciousness, Abnormal Physical Behavior, Somatic Complaints, Elation-Grandiosity, Guilt, Depressed Mood, Diminished Self-Esteem, Apathy and Retardation, Suicidal Trends, Conceptual Disorganization, Obsessive-Compulsive Behavior, Anxiety, Anger, and Flatness of Affect.

We have constructed these scales by first deciding which items of the MSS are related to the different clinical dimensions. A few scales have items in common. For example, the items describing persecutory delusions are part of both the delusional ideation and the suspiciousness scale. Next we decided for each group of items what seemed to be the most appropriate operation for converting the small units of behavior described in the items into scale values of the clinical dimension.

Certain behaviors by themselves indicate a lower limit of pathology. For example, if a patient keeps expressing hatred or contempt, this by itself indicates at least a moderate degree of anger. On the other hand, any patient who assaults the interviewer, even if he shows no other expression of anger, is among the most angry of patients and is assigned the highest scale value. For such dimensions, every item is associated with a corresponding scale value and the final scale value for a patient on the dimension is defined as the highest scale value corresponding to the items judged true. In a few instances, a scale value is defined by two particular items judged true. For example, the highest scale value of the dimension for delusional ideation is attained when a patient both keeps talking about his delusions and indicates that he takes action on the basis of them. For a relatively few dimensions, the scaling operation is merely adding the number of items judged true. For example, the scale for Flatness of Affect contains five items, each of which seems of equal importance. The scale value thus ranges from zero, if none of these items is true, to five, if all of the items are true. The reliability of these scales, as determined by joint observation of a large series of inpatients,
ranges from .53 for Flatness of Affect to better than .90 for such scales as Depression and Denial of Illness, and as high as .99 for Suicide. We hope that by revising the operations for defining the unreliable scales we can increase their reliability.‡ The use of these scales in contrasting a sample of 100 rural Kentucky and 100 urban New York schizophrenics is described elsewhere (3).

The qualifications for the examiner administering the MSS are skill in observation, familiarity with psychopathology, and the ability to conduct a standard interview. Although primarily designed for psychiatrists, we have found that psychologists and sociologists with interviewing experience and a knowledge of psychopathology can effectively administer the instrument. A manual of instructions explains in detail the variations in the administration of the interview schedule which are permitted, criteria for evaluating the items of the inventory, and two transcribed practice interviews. We also have a small library of tape-recorded interviews which we have used for training purposes as well as for estimating the accuracy of the observer's judgment. The use of recorded MSS interviews to estimate accuracy of psychiatric judgment is described elsewhere (4).

Research Design and the MSS as a Criterion Measure of Change in Psychotherapy

For the MSS to be used as a measure of change, it must be administered before or at the onset of treatment and then readministered after a designated period of time or at the conclusion of treatment. If at all possible, the person administering the instrument to the patient should not be the patient's psychotherapist. The therapist could administer the instrument on his first contact with the patient by introducing it with a statement such as, "I would like to ask you some questions so I can understand what problems are bothering you. I'll touch on many different topics, some of which will naturally be of more concern to you than others. After this we can go over in more detail some of the things that seem to be especially important in understanding your difficulties." However, most therapists, I believe, would feel quite uncomfortable in beginning their initial therapeutic contact with a patient in such a manner, and I believe that for many patients this kind of an initial contact with the therapist might also be the last.

‡ October 1965: Successive revisions of the 20 scales have been made, based upon studies of their correlations with global clinical judgments and the intercorrelations of the items. Surprisingly, with the exception of Anger and Decreased Communication, the scales were found to be more valid and reliable if scored by merely adding the number of items true rather than calculating scale levels. In addition to these clinical scales, scales are being developed based on factor analyses of two matched samples of 1,000 cases each.
Another reason why an independent assessment is preferable is that the therapist, were he to administer the instrument, might be biased in the direction of detecting more pathology on the initial contact so as to maximize the amount of change that would be possible due to his psychotherapeutic efforts. Similarly, the follow-up evaluation should also be performed by an individual with no investment in the outcome of the treatment, but it could be the same individual who performed the initial research evaluation.

In administering the MSS two or more times to over 100 patients, we have been pleasantly surprised to find that patients rarely remember being asked the same questions, and usually have only a feeling that the same general areas are being covered in the interview. When the interval between examinations is long, such as a full year, patients rarely even remember the areas discussed. Our experience in administering the instrument to a few individuals three or four times indicates that it should be possible to administer the instrument at the initiation of treatment, at the conclusion of treatment, and after a follow-up period.

The MSS could be used in a research design that calls for evaluating a control group which either received a different form of treatment, no treatment, or was on a waiting list. The experimental and control groups could be compared in terms of over-all severity of illness, the severity of various clinical syndromes or dimensions, and by an individual item analysis. Information could be obtained on the interaction of various specific therapies with specific forms of symptomatology. It should be possible to identify certain patterns of behavior which indicate poor or good prognosis with the various therapeutic procedures.

The instrument is not suitable as a criterion measure for all patients treated by psychotherapy. Because of the universe of items in the instrument, it is largely unsuitable for patients whose primary difficulties are alcoholism, psychopathic behavior, drug addiction, or homosexuality, and would be of doubtful value for extremely well-integrated patients undergoing psychoanalytic reconstructive therapy. It would be most suitable for moderately to severely ill neurotic or psychotic patients manifesting disturbances which are reflected in subjective complaints and overt disturbances in behavior manifested during a single interview. Thus, the instrument would, I believe, be fairly useful as a criterion measure for most patients attending outpatient clinics.

As the instrument does not attempt to measure the underlying structural aspects of the personality, an individual with a well-defended flight into health would with any scoring system appear as having minimal psychopathology. If, however, he exhibited much denial of illness, this would be picked up in numerous items which are constructed to capture this dimension, for example, “Says he has no worries or that nothing bothers him,”
or "indicates that he does not recognize that he needs to change his attitude in some specific way."

The instrument does not evaluate what has been termed "positive mental health." We are working on the development of another instrument, the Psychiatric Status Schedule, which is designed to evaluate the areas of positive mental health which are not covered in the MSS,§ for example, how does the individual use his leisure time? What is the quality of his friendships and interpersonal relations? How does he get along on his job? To what extent is he involved in his community? Does he have hobbies or special interests which are particular sources of satisfaction to him?

A frequent criticism made of the Mental Status Schedule is that it does not conform to any theory of psychopathology. I myself have had psychoanalytic training and believe that the psychoanalytic theory of the dynamic unconscious and the structural hypothesis is indispensable for an understanding of psychopathology. However, the items of the MSS were constructed so as to record a universe of behavior which psychiatrists over the years have recognized as being important in evaluating mentally disturbed individuals. As such, they would seem to be suitable for evaluating patients by clinicians who, although committed to various theories of psychopathology, can at least agree that these items describe the overt psychopathology of a large proportion of the mentally disturbed.

What advantages does this instrument have over a questionnaire or self-report inventory, such as the MMPI? The questionnaire or self-report methods are mainly limited to detecting psychopathology of which the patient is aware and able to report. A patient generally cannot, for example, report that he is delusional or incoherent. The detection of such behavior, as well as psychomotor evidence of pathology, requires the judgment of a trained observer conducting some kind of a psychiatric interview.

What advantages does the instrument have over the usual clinical interview as a source of data for research evaluation? In the usual clinical interview, the interviewer is free to explore whatever areas of behavior he considers most significant and to phrase his questions in whatever manner he deems most appropriate for the individual patient. In contrast, the inter-

§ October 1965: A preliminary version of the Psychiatric Status Schedule is now available. In contrast to the MSS which was primarily designed for use with psychiatric patients, the PSS is for use with both patients, former patients and nonpatients. Like the Mental Status Schedule (MSS), the PSS consists of an interview schedule which is used by the interviewer to elicit the information needed to judge a matching inventory of precoded items. The 248 items taken from the MSS are supplemented by items covering performance in the following areas: living arrangements, travel, eating, personal hygiene, drugs, alcohol, sex, illegal acts, leisure time, interpersonal relations, and self-injury. In addition, depending upon the roles of the individual being interviewed, the work, housework, mate, parent, or student roles are covered.
viewer administering the MSS follows a specific schedule of questions and statements. The use of the same interview schedule for all patients has the research advantage that differences observed among patients tend to reflect actual differences rather than artifacts caused by differences in areas of psychopathology explored, or interviewing technique. Furthermore, knowing what questions are actually asked gives a framework within which the patient's responses can be understood by persons not present at the interview itself. Similarly, recording the interviewer's judgments in a comprehensive inventory facilitates identification of the specific behaviors that form the basis of the interviewer's over-all conclusions.

Do patients resent the structured interview? Generally speaking, they do not, and if the interviewer is familiar with the instrument, he can administer it in a fairly natural manner so that the patient does not even recognize that he is reading standard questions. Those few patients who seem to resent the structured interview are usually so paranoid and negativistic that they would resent any interview procedure.

In 1953, Zubin suggested the creation of a centrally located control group consisting of a large group of mental patients who would receive no treatment and whose course and outcome could be used as a standard against which the various therapies could be compared. (5). His suggestion was never put into effect. On behalf of Zubin and myself, we now suggest that instead of putting the patients on ice, it would be preferable to select a group of standard evaluators proficient in administering standardized interview techniques, such as the Mental Status Schedule, and the Psychiatric Status Schedule. Patients receiving psychotherapy could be sent to this objective Bureau of Standards for an independent assessment at the initiation and termination of their treatment and at follow-up. If the psychotherapist explained to his private patient the research value of such an independent assessment, I believe patients would generally not object to the inconvenience of the research evaluation.

At the first national scientific meeting of this Association, Stanley Lesse introduced a panel discussion on "Criteria for the Evaluation of the Results of Psychotherapy" with this statement:

Before results can be evaluated in meaningful fashion, significant criteria that can be duplicated and applied by many observers must first be delineated. In the absence of general agreement as to what constitutes the necessary basic elements of these criteria, there can be no adequate communication among the host of research and clinical psychotherapists who have been applying various techniques to a wide spectrum of patients. This specification of criteria is of equal importance whether one is interested in primarily the study of the process of psychotherapy or in the evaluation of results.

It is extremely unlikely that any one instrument can ever serve as a suit-
able criterion measure for all of the kinds of psychopathologic disturbances that are treated with the various forms of psychotherapy. I believe, however, that the MSS can make a significant contribution toward the delineation of criteria for evaluating response to treatment that could be acceptable to many observers of different theoretical persuasions.

REFERENCES


