The Dynamics of "Structured" Personality Tests

Paul E. Meehl
University of Minnesota Hospitals

Originally published in JCLP, 1, 296–303 (1945). © 2000 John Wiley & Sons, Inc. J Clin Psychol 56: 367–374, 2000.

In a recent article in this Journal,[9] Lt. Max L. Hutt of the Adjutant General's School has given an interesting discussion of the use of projective methods in the army medical installations. This article was part of a series describing the work of clinical psychologists in the military services, with which the present writer is familiar only indirectly: The utility of any instrument in the military situation can, of course, be most competently assessed by those in contact with clinical material in that situation, and the present paper is in no sense to be construed as an "answer" to or an attempted refutation of Hutt's remarks. Nevertheless, there are some incidental observations contained in his article which warrant further critical consideration, particularly those having to do with the theory and dynamics of "structured" personality tests. It is with these latter observations rather than the main burden of Hutt's article that this paper is concerned.

Hutt defines "structured personality tests" as those in which the test material consists of conventional, culturally crystallized questions to which the subject must respond in one of a very few fixed ways. With this definition we have no quarrel, and it has the advantage of not applying the unfortunate phrase "self-rating questionnaire" to the whole class of question-answer devices. But immediately following this definition, Hutt goes on to say that "it is assumed that each of the test questions will have the same meaning to all subjects who take the examination. The subject has no opportunity of organizing in his own unique manner his response to the questions."

These statements will bear further examination. The statement that personality tests assume that each question has the same meaning to all subjects is continuously appearing in most sources of late, and such an impression is conveyed by many discussions even when they do not explicitly make this assertion. It should be emphasized very strongly, therefore, that while this perhaps has been the case with the majority of question-answer personality tests, it is not by any means part of their essential nature. The traditional approach to verbal question-answer personality tests has been, to be sure, to view them as

© 2000 John Wiley & Sons, Inc.

self-ratings; and it is in a sense always a self-rating that you obtain when you ask a subject about himself, whether you inquire about his feelings, his health, his attitudes, or his relations to others.

However, once a "self-rating" has been obtained, it can be looked upon in two rather different ways. The first, and by far the commonest approach, is to accept a self-rating as a second best source of information when the direct observation of a segment of behavior is inaccessible for practical or other reasons. This view in effect forces a self-rating or self-description to act as surrogate for a behavior-sample. Thus we want to know whether a man is shy, and one criterion is his readiness to blush. We cannot conveniently drop him into a social situation to observe whether he blushes, so we do the next best (and often much worse) thing and simply ask him, "Do you blush easily?" We assume that if he does in fact blush easily, he will realize that fact about himself, which is often a gratuitous assumption; and secondly, we hope that having recognized it, he will be willing to tell us so.

Associated with this approach to structured personality tests is the construction of items and their assembling into scales upon. an a *priori* basis, requiring the assumption that the psychologist building the test has sufficient insight into the dynamics of verbal behavior and its relation to the inner core of personality that he is able to predict beforehand what certain sorts of people will say about themselves when asked certain sorts of questions. The fallacious character of this procedure has been sufficiently shown by the empirical results of the Minnesota Multiphasic Personality Inventory alone, and will be discussed at greater length below. It is suggested tentatively that the relative uselessness of most structured personality tests is due more to *a priori* item construction than to the fact of their being structured.

The second approach to verbal self-ratings is rarer among test-makers. It consists simply in the explicit denial that we accept a self-rating as a feeble surrogate for a behavior sample, and substitutes the assertion that a "self-rating" constitutes an intrinsically interesting and significant bit of verbal behavior the non-test correlates of which must be discovered by empirical means. Not only is this approach free from the restriction that the subject must be able to describe his own behavior accurately, but a careful study of structured personality tests built on this basis shows that such a restriction would falsify the actual relationships that hold between what a man says and what he *is*.

Since this view of question-answer items is the rarer one at the present time, it is desirable at this point to elucidate by a number of examples. For this purpose one might consider the Strong Vocational Interest Blank, the Humm-Wadsworth Temperament Scales, the Minnesota Multiphasic Personality Inventory, or any structured personality measuring device in which the selection of items was done on a thoroughly empirical basis using carefully selected criterion groups. In the extensive and confident use of the Strong Vocational Interest Blank, this more sophisticated view of the significance of responses to structured personality test items has been taken as a matter of course for years. The possibility of conscious as well as unconscious "fudging" has been considered and experimentally investigated by Strong and others, but the differences in possible interpretation or *meaning* of items have been more or less ignored—as well they should be. One is asked to indicate, for example, whether he likes, dislikes, or is indifferent to "conservative people." The possibilities for differential interpretation of a word like *conservative* are of course tremendous, but nobody has worried about that problem in the case of the Strong. Almost certainly the strength of verbs like "like" and "dislike" is variably interpreted throughout the whole blank. For the present purpose the Multiphasic (referred to hereinafter as MMPI) will be employed because the present writer is most familiar with it.

One of the items on the MMPI scale for detecting psychopathic personality (Pd) is "My parents and family find more fault with me than they should." If we look upon this as a rating in which the *fact* indicated by an affirmative response is crucial, we immediately begin to wonder whether the testee can objectively evaluate how much other people's parents find fault with them, whether his own parents are warranted in finding as much fault with him as they do, whether this particular subject will interpret the phrase "finding fault" in the way we intend or in the way most normal persons interpret it, and so on. The present view is that this is simply an unprofitable way to examine a question-answer personality test item. To begin with, the empirical finding is that individuals whose past history and momentary clinical picture is that of a typical psychopathic personality tend to say "Yes" to this much more often than people in general do. Now in point of fact, they probably should say "No" because the parents of psychopaths are sorely tried and probably do not find fault with their incorrigible offspring any more than the latter deserve. An allied item is "I have been quite independent and free from family rule" which psychopaths tend to answer *false*—almost certainly opposite to what is actually the case for the great majority of them. Again, "Much of the time I feel I have done something wrong or evil." Anyone who deals clinically with psychopaths comes to doubt seriously whether they could possibly interpret this item in the way the rest of us do (*cf.* Cleckley's[2] "semantic dementia"), but they *say* that about themselves nonetheless. Numerous other examples such as "Someone has it in for me" and "I am sure I get a raw deal from life" appear on the same scale and are significant because psychopaths tend to *say* certain things about themselves, rather than because we take these statements at face value.

Consider the MMPI scale for detecting tendencies to hypochondriasis. A hypochondriac says that he has headaches often, that he is not in as good health as his friends are, and that he cannot understand what he reads as well as he used to. Suppose that he has a headache on an average of once every month, as does a certain "normal" person. The hypochondriac says he often has headaches, the other person says he does not. They both have headaches once a month, and hence they must either interpret the word "often" differently in that question, or else have unequal recall of their headaches. According to the traditional view, this ambiguity in the word "often" and the inaccuracy of human memory constitute sources of error; for the authors of MMPI they may actually constitute sources of discrimination.

We might mention as beautiful illustrations of this kind of relation, the non-somatic items in the hysteria scale of MMPI[8]. These items have a statistical homogeneity and the common property by face inspection that they indicate the person to be possessed of unusually good social and psychiatric adjustment. They are among the most potent items for the detection of hysterics and hysteroid temperaments, but they reflect the systematic distortion of the hysteric's conception of himself, and would have to be considered invalid if taken as surrogates for the direct observation of behavior.

As a last example one might mention some findings of the writer, to be published shortly, in which "normal" persons having rather abnormal MMPI profiles are differentiated from clearly "abnormal" persons with equally deviant profiles by a tendency to give statistically rare as well as psychiatrically "maladjusted" responses to certain other items. Thus a person who says that he is afraid of fire, that wind-storms terrify him, that people often disappoint him, stands a better chance of being normal in his non-test behavior than a person who does not admit to these things. The discrimination of this set of items for various criterion groups, the intercorrelations with other scales, and the content of the items indicate strongly that they detect some verbal-semantic distortion in the interpretation and response to the other MMPI items which enters into the spurious ele-

vation of scores achieved by certain "normals." Recent unpublished research on more subtle "lie" scales of MMPI indicates that unconscious self-deception is inversely related to the kind of verbal distortion just indicated.

In summary, a serious and detailed study of the MMPI items and their interrelations both with one another and non-test behavior cannot fail to convince one of the necessity for this second kind of approach to question-answer personality tests. That the majority of the questions seem by inspection to require self-ratings has been a source of theoretical misunderstanding, since the stimulus situation seems to request a self-rating, whereas *the scoring does not assume a valid self-rating to have been given.* It is difficult to give any psychologically meaningful interpretation of some of the empirical findings on MMPI unless the more sophisticated view is maintained.

It is for this reason that the possible differences in interpretation do not cause us any *a priori* concern in the use of this instrument. Whether any structured personality test turns out to be valid and useful must be decided on pragmatic grounds, but the possibility of diverse interpretations of a single item is not a good *theoretical* reason for predicting failure of the scales. There is a "projective" element involved in interpreting and responding to these verbal stimuli which must be recognized, in spite of the fact that the test situation is very rigidly structured as regards the ultimate response possibilities permitted. The objection that all persons do not interpret structured test items in the same way is not fatal, just as it would not be fatal to point out that "ink blots do not look the same to everyone.

It has not been sufficiently recognized by critics of structured personality tests that what a man says about himself may be a highly significant fact about him even though we do not entertain with any confidence the hypothesis that what he says would agree with what complete knowledge of him would lead others to say of him. It is rather strange that this point is so often completely passed by, when clinical psychologists quickly learn to take just that attitude in a diagnostic or therapeutic interview. The complex defense mechanisms of projection, rationalization, reaction-formation, etc., appear dynamically to the interviewer as soon as he begins to take what the client *says* as itself motivated by other needs than those of giving an accurate verbal report. There is no good *a priori* reason for denying the possibility of similar processes in the highly structured "interview" which is the question-answer personality test. The summarized experience of the clinician results (one hopes, at least) in his being able to discriminate verbal responses admissible as accurate self-descriptions from those which reflect other psychodynamisms but are not on that account any the less significant. The test analogue to this experience consists of the summarized statistics on response frequencies, at least among those personality tests which have been constructed empirically (MMPI, Strong, Rorschach, etc.).

Once this has been taken for granted we are prepared to admit powerful items to personality scales regardless of whether the rationale of their appearance can be made clear at present. We do not have the confidence of the traditional personality test maker that the relation between the behavior dynamics of a subject and the tendency to respond verbally in a certain way must be psychologically obvious. Thus it puzzles us but does not disconcert us when this relation cannot be elucidated, the science of behavior being in the stage that it is. That "I sometimes tease animals" (answered *false*) should occur in a scale measuring symptomatic depression is theoretically mysterious, just as the tendency of certain schizophrenic patients to accept "position" as a determinant in responding to the Rorschach may be theoretically mysterious. Whether such a relation obtains can be very readily discovered empirically, and the wherefore of it may be left aside for the moment as a theoretical question. Verbal responses which do not apparently have any self-reference at all, but in their form seem to request an objective judgment about social

phenomena or ethical values, may be equally diagnostic. So, again, one is not disturbed to find items such as "I think most people would lie to get ahead" (answered *false*) and "It takes a lot of argument to convince most people of the truth" (answered *false*) appearing on the hysteria scale of MMPI.

The frequently alleged "superficiality" of structured personality tests becomes less evident on such a basis also. Some of these items can be rationalized in terms of fairly deep-seated trends of the personality, although it is admittedly difficult to establish that any given depth interpretation is the correct one. To take one example, the items on the MMPI scale for hysteria which were referred to above as indicating extraordinarily good social and emotional adjustment can hardly be seen as valid self-descriptions. However, if the core trend of such items is summarily characterized as "I am psychiatrically and socially well adjusted," it is not hard to fit such a trend into what we know of the basic personality structure of the hysteric. The well known *belle indifference* of these patients, the great lack of insight, the facility of repression and dissociation, the "impunitiveness" of their reactions to frustration, the tendency of such patients to show an elevated "lie" score on MMPI, may all be seen as facets of this underlying structure. It would be interesting to see experimentally whether to the three elements of Rosenzweig's "triadic hypothesis" (impunitiveness, repression, hypnotizability) one might add a fourth correlate—the chief non-somatic component of the MMPI hysteria scale.

Whether "depth" is plumbed by a structured personality test to a lesser extent than by one which is unstructured is difficult to determine, once the present view of the nature of structured tests is understood. That the "deepest" layers of personality are not verbal might be admitted without any implication that they cannot therefore make themselves known to us via verbal behavior. Psychoanalysis, usually considered the "deepest" kind of psychotherapy, makes use of the dependency of verbal behavior upon underlying variables which are not themselves verbalized.

The most important area of behavior considered in the making of psychiatric diagnosis is still the form and content of the *speech* of the individual. I do not mean to advance these considerations as validations of any structured personality tests, but merely as reasons for not accepting the theoretical objection sometimes offered in criticizing them. Of course, structured personality tests may be employed in a purely diagnostic, categorizing fashion, without the use of any dynamic interpretations of the relationship among scales or the patterning of a profile. For certain practical purposes this is quite permissible, just as one may devote himself to the statistical validation of various "signs" on the Rorschach test, with no attempt to make qualitative or really dynamic personological inferences from the findings. The tradition in the case of structured personality tests is probably weighted on the side of non-dynamic thinking; and in the case of some structured tests, there is a considerable amount of experience and clinical subtlety required to extract the maximum of information. The present writer has heard discussions in case conferences at the University of Minnesota Hospital which make as "dynamic" use of MMPI patterns as one could reasonably make of any kind of test data without an excessive amount of illegitimate reification. The clinical use of the Strong Vocational Interest Blank is another example.

In discussing the "depth" of interpretation possible with tests of various kinds, it should at least be pointed out that the problem of validating personality tests, whether structured or tin-structured, becomes more difficult in proportion as the interpretations increase in "depth." For example, the validation of the "sign" differentials on the Rorschach is relatively easier to carry out than that of the deeper interpretations concerning the basic personality structure. This does not imply that there is necessarily less validity in the latter class of inferences, but simply stresses the difficulty of designing experi-

ments to test validity. A very major part of this difficulty hinges upon the lack of satisfactory external criteria, a situation which exists also in the case of more dynamic interpretations of structured personality tests. One is willing to accept a staff diagnosis of psychasthenia in selecting cases against which to validate the Pt scale of MMPI or the F% as a compulsive-obsessive sign on the Rorschach. But when the test results indicate repressed homosexuality or latent anxiety or lack of deep insight into the self, we may have strong suspicions that the instrument is fully as competent as the psychiatric staff. Unfortunately this latter assumption is very difficult to justify without appearing to be inordinately biased in favor of our test. Until this problem is better solved than at present, many of the "depth" interpretations of both structured and unstructured tests will be little more than an expression of personal opinion.

There is one advantage of unstructured personality tests which cannot easily be claimed for the structured variety, namely, the fact that falsehood is difficult. While it is true for many of the MMPI items, for example, that even a psychologist cannot predict on which scales they will appear nor in what direction certain sorts of abnormals will tend to answer them, still the relative accessibility of defensive answering would seem to be greater than is possible in responding to a set of ink-blots. Research is still in progress on more subtle "lie" scales of MMPI and we have every reason to feel encouraged on the present findings. Nevertheless the very existence of a definite problem in this case and not in the case of the Rorschach gives the latter an advantage in this respect. When we pass to a more structured method, such as the T. A. T., the problem reappears. The writer has found, for example, a number of patients who simply were not fooled by the "intelligence-test" set given in the directions for the T. A. T., as was indicated quite clearly by self-references and defensive remarks, especially on the second day. Of course such a patient is still under pressure to produce material and therefore his unwillingness to reveal himself is limited in its power over the projections finally given.

In conclusion, the writer is in hearty agreement with Lieutenant Hutt that unstructured personality tests are of great value, and that the final test of the adequacy of any technique is its utility in clinical work. Published evidence of the validity of both structured and unstructured personality tests as they had to be modified for convenient military use does not enable one to draw any very definite conclusions or comparisons at the present time. There is assuredly no reason for us to place structured and unstructured types of instruments in battle order against one another, although it is admitted that when time is limited they come inevitably into a very real clinical "competition" for use. The present article has been aimed simply at the clarification of certain rather prevalent misconceptions as to the nature and the theory of at least one important structured personality test, in order that erroneous theoretical considerations may not be thrown into the balance in deciding the outcome of such clinical competition.

References

1. Benton, A. C. The interpretation of questionnaire items in a personality schedule Archiv. Psychol. 1935, No. 190.
2. Cleckley, H. The Mask of Sanity. St. Louis: Mosby, 1941.
3. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule: I. Construction of the schedule. J. Psychol., 1940, 10, 249–254.
4. McKinley, J. C., and Hathaway, S. R. A multiphasic personality schedule: II. A differential study of hypochondriasis. J. Psychol., 1940, 10, 255–268.
5. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule: II. The measurement of symptomatic depression. J. Psychol., 1942, 14, 73–84.

6. McKinley, J. C., and Hathaway, S. R. A multiphasic personality schedule: IV. Psychasthenia. J. Appl. Psychol., 1942, 26, 614–624.
7. Hathaway, S. R., and McKinley, J. C. Manual for the Minnesota Multiphasic Personality Inventory. Minneapolis: University of Minnesota Press, 1943.
8. McKinley, J. C., and Hathaway, S. R. The Minnesota Multiphasic Personality Inventory: V. Hysteria, hypomania, and psychopathic deviate. J. Appl. Psychol., 1944, 28, 153–174.
9. Hutt, Max L. The use of projective methods of personality measurement in army medical installations. J. Clin. Psychol., 1945, 1, 134–140.
10. Landis, C., and Katz, S. E. The validity of certain questions which purport to measure neurotic tendencies. Jour. Appl. Psychol., 1934, 18, 343–356.
11. Landis, C., Zubin, J., and Katz, S. E. Empirical evaluation of three personality adjustment inventories. J. Educ. Psychol., 1935, 26, 321–330.
12. Leverenz, C. W. Minnesota Multiphasic Personality Inventory: An evaluation of its usefulness in the psychiatric service of a station hospital. War Med., 1943, 4, 618–629.
13. Maller, J. B. Personality tests. In J. McV. Hunt (ed.), Personality and the Behavior Disorders. New York: Ronald Press, 1944, pp. 170–213.
14. Meehl, P. E. A general normality or control factor in personality testing. Unpublished Ph.D. Thesis, University of Minnesota Library, Minneapolis, 1945.
15. Mosier, C. I. On the validity of neurotic questionnaires. J. Soc. Psychol., 1938, 9, 3–16.
16. Rosenzweig, S. An outline of frustration theory. In J. McV. Hunt (ed.), Personality and the Behavior Disorders. New York: Ronald Press, 1944.
17. Strong, E. K. Vocational Interests of Men and Women. Stanford: Stanford University Press, 1943.