Meehl, P. E., & Jeffery, M. (1946). The Hunt-Minnesota Test for Organic Brain Damage in cases of functional depression. *Journal of Applied Psychology, 30*, 276-287.

The Hunt-Minnesota Test for Organic Brain Damage in Cases of Functional Depression *

Paul E. Meehl and Mary Jeffery
The University of Minnesota

Among the several tests which have been devised for the detection of intellectual deterioration, one of the most efficient is the Hunt-Minnesota Test for Organic Brain Damage (Hunt, 1943a,b, 1944, 1945)

The author developed this instrument specifically for the diagnosis of organic damage. While the detection and measurement of a decrement in intellectual function, however caused, is an important part of the clinical psychologist's work, methods must eventually be developed for distinguishing between two kinds of deterioration. On the one hand are those which are secondary to "emotional-motivational" factors (e.g. in schizophrenia) and, on the other, those which represent the direct effect of organic central nervous system pathology.[1] Since many varieties of behavior disorders are characterized by a certain amount of psychological deficit, the psychologist will obviously be playing a more significant clinical role if he can make a definite contribution to differential diagnosis (e.g. as between "functional" and "organic" deficit) instead of merely reporting a deviation from the " normal" or "optimal" level.

Such an added report would carry no implication as to the ultimate etiology of the disorder which he has thus labeled as showing either "functional" or "organic" deterioration. Because even if "organic" (endocrine, metabolic, or autonomic) factors should finally be established as primary causes of the development of a schizophrenia, it would still be possible for an observed intellectual deficit to occur as a function of motivation, itself dependent upon the organic factors. In cases with "functional" deterioration, techniques to effectively motivate the patient may cause him to return temporarily to his "true" level, a phenomenon which has often been observed by clinicians. Whereas, in the strictly "organic" case, no such motivational improve-

* The authors are indebted to Dr. B. C. Schiele and Dr. A. B. Baker, Department of Neuropsychiatry, for their cooperation in this study.

[1] As Hunt has pointed out, test results must be considered as only a part of the evidence required for diagnosis, and must always be interpreted in the light of data from all sources: "Deterioration test scores are thus not a final index … but rather a diagnostic and prognostic aid. The extent to which they aid diagnosis and prognosis depends, to a substantial degree, upon the skill and clinical acuity of the interpreting clinician" (Hunt, 1945).

ment can make up for a deficit in function directly related to nerve cell destruction, such as occurs in senile dementia or paresis.

In the construction of psychometric devices for detecting organic brain damage, therefore, a difficulty arises because deficit can reflect multiple causation. Consequently, in such psychometric devices, we must either employ tasks which shall yield no decrement for subjects who are merely suffering from test-anxiety, boredom, preoccupation with fantasy, or depressive retardation; or, if this is impossible, we must have means for identifying such special decrements.

The approach suggested by the first alternative is very difficult because possibly it might only be attained by eliminating some degree of test sensitivity to organic deficit. Such a loss of sensitivity might reduce the effectiveness of a test to a point where it would detect only an amount of intellectual loss so gross as to be detectable on other grounds. The ultimate aim of such tests must be to detect *minimal* amounts of damage so that the psychologist can contribute independent evidence of the presence of pathology in the same way as the serologist or roentgenologist can do in cases which are relatively asymptomatic. Psychological tests which merely detect intellectual loss in a person, known on other grounds to be brain damaged, do not contribute maximally to clinical work.

There is reason to believe that this "maximal contribution" is possible because complex intellectual processes are very sensitive to even slight cortical disturbances; and there is already sufficient evidence that the Hunt-Minnesota Test has achieved the increased delicacy desired. Yet just here is the "difficulty" referred to. For, no sooner has the delicacy of the test been stepped up to a point where it can pick up small losses such as those in an early senile change or an undiagnosed encephalitis, than it is also markedly affected by the motivational and emotional factors which are present in other types of cases. In short the increased delicacy is rarely specific for the *kind* of decrement we wish to detect. This would seem to be the crucial problem confronting the clinical psychologist in the field of mental deterioration.

Experience with the Hunt-Minnesota Test at the University of Minnesota Hospitals has demonstrated its high validity as an indicator of organic brain damage. Some of the data have already been published (Hunt, 1944), and other validation studies are in progress. Experience, however, in testing cases of depression aroused a doubt that the specificity of the device for organic damage was as great as had originally been hoped. This feeling was first expressed by Hunt, himself, in this journal (Hunt, 1945) when he wrote: "In the development of the Hunt test, an attempt was made to provide a special means for identifying those pathologic scores attributable to emotional-moti-

vational disturbances so that the test would then be a specific test for the deterioration associated with brain damage. This attempt has been only partially successful."

Hunt had attempted to identify the scores that he referred to, by including a set of "validity" tests (called *interpolated* tests in his manual) such as digit span, attention, and saying the months forward and backward. His theory was that persons who are disturbed or uncooperative so much as to invalidate their test results would fail the interpolated tests, and thus the examiner would have an index for avoiding an interpretation of deterioration due to organic brain damage. As will be seen from study of the manual, the standard of scoring is extremely lenient; the criterion of invalidity being "failure" on three or more of the nine interpolated tests.

As it stands, the Hunt test showed results that were gratifying with the majority of cases at the University Psychopathic Unit. However, with some persons showing anxiety and depression, high deterioration scores were obtained without other evidence of organic brain damage. Most of these patients seemed quite capable of cooperating as judged by the interpolated tests, usually passing them by a wide margin. To corroborate this clinical impression the present study was undertaken.

It is important that investigations of this sort should avoid unintentionally including subjects already suffering from minimal organic damage. The mere absence of a diagnosis of pathology cannot be taken as proof of normality, without a systematic check in the form of careful history taking and neurological examination. Even using neurology as the criterion, it is unfortunately true that, among the so-called "false positives," an unknown number of persons are actually correctly "positive." However, all one can do is to include only cases which have been neurologically studied and are negative, and then to make the assumption that only a small minority of the group (in the absence of other evidence) have any minimal damage over and above that due to age, for which the Hunt test presumably supplies an adequate correction.

Originally it was intended to obtain retests upon all cases, initially tested during a state of depression, following recovery. This plan was abandoned for several reasons. First of all, research by Arkola (1945) indicated the existence of a practice effect of some magnitude. This was apparent even after the lapse of considerably more time than would have passed before "recovery" in our cases. Secondly, the great majority of depressed patients were treated with electroshock therapy which, in itself, may result in unknown amounts of minimal brain damage. The result, then, would have been a combined effect of three variables; two of them (recovery and practice) would tend to lower the T-score by an indeterminate amount, and the other (shock therapy)

would tend possibly to raise it. Accordingly, this scheme of investigation had to be abandoned.

In choice of subjects, several restrictions were necessary. The age limits of 20 to 55 years, for which Hunt claims maximum effectiveness for his test, were imposed. It was required that there be no hint of organic findings or history of shock therapy of any kind during previous episodes. Over a period of eleven months, despite a large number of patients "considered" as subjects, only seventeen subjects fulfilled our requirements. Of these, two were later eliminated because such suggestive signs as slight retinal arteriosclerosis or markedly elevated blood pressure appeared during subsequent neurological study. That the final group of fifteen cases of clearly "functional" depression is small, reflects the extreme care with which cases were selected. The findings, however, are so clear-cut and the Hunt test is being used so extensively that the writers feel further delay in reporting results is ill-advised. Dr. Hunt concurs in this opinion.

The Group

The group studied consists of all in-patient cases with prominent symptoms of depression admitted to the Psychopathic Unit of the University Hospitals from October 1944 through November 1945. Of those who met the required conditions, there were 13 females and 2 males, all between the ages of 34 and 55. The median age was 50 years, with a mean age of 48.7 and SD of 6.4 years. Education varied from 7th grade through two years of college, the mean education attained being 10th grade with a SD of 2.4 years. Vocabulary level, on the Stanford Binet list, varied from 15 words (M.A. about 13 years) to 29 words (not quite Superior Adult III), with a mean of 22,7 words (Superior Adult I). The *t* of this mean, from a hypothetical supply mean of 20 words, is 2.086 which lies between the 5% and 10% levels of probability.

All of the patients tested had previously received thorough physical and neurological examinations as well as routine laboratory studies. In each case, these were all negative, and no case had a history of head trauma, addiction, or encephalitis.

One patient had a blood pressure of 170/100 but was included because her chart gave three much lower readings for examinations of about 18 months previous. She showed no evidence of cerebral arteriosclerotic changes, neurologically or opthalmoscopically.

Nine cases were entirely unsedated when tested, and the remaining six were under sedation with either phenobarbital (1½ grains), sodium amytal (3 grains), nembutal (1½ grains), or seconal (3 grains). An unpublished study by Arkola (1945) has shown that this amount of sedation with the barbiturates does not produce measurable effects

upon Hunt scores, even when administered by injection, and tested at the peak of the sedative effect. Most of the present cases were tested several hours after the oral-administration of the sedatives. Furthermore, the mean T-score of the six sedated patients is 66.8 whereas that of the nine unsedated ones is 72.4 (medians 67.5 and 75 respectively). Consequently, it seems safe to assume that these slight degrees of sedation cannot by any means account for the elevations to be reported below.

The staff diagnoses of the fifteen cases were as follows: Involutional melancholia, 5; psychoneurosis, reactive depressive, 4; manic depressive psychosis, depressed, 2; and one each of involutional psychosis, depressed and paranoid; manic depressive psychosis, mixed (agitated); psychoneurosis, mixed (reactive depressive and psychaesthenia); and psychoneurosis, anxiety state.

The mean Multiphasic Personality Inventory profile for these 15 cases was as follows: ? 50.8, L 56.5, F 58.8, Hs 66.6, D 86.3, Hy 72.8, Pd 69.7, Pa 72.1, Pt 70.7, Sc 67.4, Ma 52.6, Mf 55.7. In 10 of the cases the depression score (D) was the peak of the profile, and in 11 cases it was above 70. Among the four cases in which D was less than 70, two showed T-scores of 63 on the "lie" scale (L). However, one of these cases was not tested with the Multiphasic until some 55 days after administration of the Hunt, at a time when her psychiatric condition had improved considerably. The median time elapsing between the administrations of the Hunt and of the Multiphasic was three days, although in two cases an interval of over eight days had elapsed between the administration of the two tests.

It should be pointed out that although all of these patients were depressed in varying amounts, many of them were at some stage of improvement when tested. No patient was tested whose momentary psychiatric condition was such as to preclude his at least claiming ability to cooperate, and apparently doing so. This will be more evident when we later consider the results obtained on the nine interpolated tests.

One case, called "anxiety state" and lacking the word "depression" in her diagnosis, was included because depression, crying, weakness, and insomnia were prominent in her complaints, and because her most marked elevation on the MMPI was on the Depression scale (T-score = 98).

The testing procedure was that described by Hunt in his manual; however, the special urging and explanation required to secure adequate cooperation was possibly more than would be employed routinely. But no actual "coaching" or allowance of leeway in time limits occurred. As was suggested by the author, the "long form" of the Hunt test was administered. A brief, semi-standardized interview was used

following the Hunt test in an attempt to form some impression of the more qualitative aspects of the patient's response to the test situation. The implications of these responses will be discussed below.

The testing was done more or less alternately by the authors, but, due to special circumstances, nine cases were tested by one author and six by the other. Since the mean T-score of these two sets of cases do not differ significantly ($P > .20$), all of the data have been combined for interpretation.

Results

The long-form T-scores of these 15 functionally depressed patients were as follows, in order of magnitude: 88, 87, 87, 87, 83, 75, 74, 73, 69, 68, 65, 62, 55, 44, and 36. The mean of these scores is 70.2 and the median, 73. The sample SD is 15.41 and the best estimate of the supply variability is 15.95 T-score units. Even with a sample this small, it is quite evident that the central tendency of T-scores for depressed patients is considerably above that of the supply mean (of 50) used in interpretation of scores.

Testing the hypothesis that such a sample could have arisen from a population with parameter mean of 50, the Student t is 4.906 which, with 14 d.f., is highly significant ($P < .0002$). We may conclude with confidence, therefore, that the scores of depressed persons cannot be evaluated on the basis of a non-brain-damaged supply mean of 50 T-score.

The obtained estimate of the SD is 15.954, about half again as large as the norm sigma of 10 points. Making use of the fact that the ratio of a sample variance to the supply variance is distributed as χ^2/n, we find a χ^2 of 35.604 which, with 14 d.f., is again highly significant ($P < .008$). It is clear, then, that neither the mean nor the variability of the depressed population can be assumed to be the same as those of the norms.

The confidence belt for the mean (using t) extends down to a T-score of 61.37, using the 5% level of confidence. On the basis of our obtained sample, we may therefore say that the "true" mean of depressed patients is almost certainly not less than about 61, i.e., a full standard deviation above the mean of the general population norms. A similar application of the χ^2 distribution indicates that, at the 5% level of confidence, the "true" SD cannot reasonably be assumed to be less than 12.26 T-score points.

With only 15 cases it was not practicable nor legitimate to make a normal curve fit and test for normality. However, the ω test of Geary (Peters & Van Voorhis, 1940), employing the ratio of the MD to the SD, was done since it is quite exact even for this small a sample. The

MD of these cases is 12.32, which bears a ratio of .800 to the sample SD. This is almost precisely the mean of the sampling distribution of ω, and there is no reason for assuming that the distribution of scores in the supply is abnormal.

When this approximating assumption has been made, the question arises: How many depressed patients may be expected to show T-scores above the "critical line" of 70? If the sample mean is taken as the best estimate, it is apparent that about half of all cases may be expected to show such spuriously "organic" scores.

Or, more generously, the extreme (most favorable) limits of the confidence belt for the mean and sigma of the supply may be taken. That is, if it is assumed that the true mean is as low as 61.368, as indicated above (a very improbable sampling error), and that the true standard deviation is as small as 12.260, the critical score of 70 is about .704 standard deviations above the mean in such a population distribution. On the assumption of normality, this implies that about 24%, or nearly one in four, depressed persons can be expected to have "pathological" T-scores.

If, as suggested by Hunt in his second article (1943b), the critical score of 66 were used, the line would be set at .378 sigma above the hypothesized supply mean and therefore 35%, or about one in three, depressed cases would show a "pathological" result. Inspection of the distribution and the mean-median relationship would suggest that, to the extent the assumption of supply normality does not hold, it is because of negative skewness, possibly due to the presence of the rare depressed person whose emotional state leaves his motor and cognitive functions relatively intact. Such a skewness would of course make the proportion of spuriously deteriorated scores even higher.

In summary of these analyses, it is clear that the present sample makes it practically certain that the elevations of T-score in depressed persons cannot be evaluated in terms of the published norms *if* the desired interpretation, that of deterioration due to organic brain damage, is to be made. At the very best, we see that about one in four functionally depressed patients will show scores above the critical line of 70, or about one in three using the score Hunt advises. A much more plausible estimate in terms of the sample statistics is, of course, that about half of the patients will show such elevated scores.

How well do the interpolated tests function in their purpose of detecting such spuriously "pathological" cases? Of the entire group of 15 depressed cases, only one case failed as many as three interpolated tests, Hunt's criterion that the test is invalid. Indeed, only four of the present group failed *any* of the nine interpolated tests; and inspection of the protocols shows, additionally, that the great majority of the cases were even far removed from the "danger line" on any interpo-

lated test. For those four cases who failed one or more interpolated tests, the T-scores were 87, 87, 83 and 73. The one patient whose test would have been identified as invalid on the basis of interpolated test scores (with a failure of six out of nine) had a T-score of 73.

Arbitrarily, rough weights were assigned to the scores on each interpolated test, and the weights for all nine interpolated tests were summed for each patient. There was no significant relation in our sample between this quantity and the size of T-score ($r = .15$, $P > .50$).

Whether or not the scoring on the interpolated tests could be made more rigorous as a method of solving the present problem cannot be determined from our data. But the good performance of most of the cases and the lack of correlation with T-score, suggests that such a strategem might not work. In order for the majority of functionally depressed patients to fail them, the scoring of the interpolated tests would have to be so rigorous that they probably would begin identifying cases of actual deterioration as "invalidly tested." This seems very likely since these tests have already been used with some success as indicators of deterioration by Babcock [1930] and others. However, such a possibility would need to be explored further.

It should be noted that the examiners, on the basis of their previous clinical experience with the Hunt test as used with depressed cases, were able to supplement the interpolated tests in assessing the validity of each test. The test, then, did not "miss" diagnostically as often as the statistics would indicate.

Any reasonably competent clinician would, of course, use his judgment in cases where the psychiatric condition of the patient made invalidity a serious possibility. The examiners, however, would not have been able to distinguish the spuriously high scores adequately here, even though probably influenced by test performance. Before actually scoring the test, each examiner made a rating as to the apparent validity of the testing, trying to exclude estimates of the quantitative results; and to judge, both in terms of the performance as it appeared qualitatively, and in terms of results from the short, post-testing interview. These ratings fell into three categories, namely: probably valid (6 cases), doubtful (4 cases), and probably invalid (5 cases). Dividing the 15 T-scores into three categories from high to low in the same proportions, a chi-square test on the resulting nine-fold table was not significant ($\chi^2 = 5.379$, 4 d.f., $P > .20$).

The results of a short semi-standardized interview, following the administration of the Hunt test, might be discussed briefly. Answers to the question: "How did you like taking this test?" were rated jointly by the authors, independently of knowledge of test-scores; three categories (favorable, neutral, unfavorable to the test) were used. A chi-square between these ratings and the size of the T-score (9-fold

table) was 16.35, which with 4 d.f. is significant at the 1% level. The contingency coefficient based upon this χ^2 is .724, indicating some relationship between how badly the patient performed and his own emotional reaction of disfavor toward the test situation.

An arbitrary weighting of a check list for emotional responses (crying, trembling, etc.) shown by the patient together with subjective judgments by the examiner as to the patient's degree of retardation, motivation, etc., correlated .45 with the T-score which the patient obtained on the Hunt test. That is, high scores were associated with a higher degree of emotional disturbance. With only 15 cases, this correlation lies between the 5% and 10% levels using Fisher's *t*.

Between magnitude of T-score and score on the Minnesota Multiphasic Depression scale, there was an insignificant association ($r = -.13$, $P > .60$).

The four highest T-scores are those of psychotics, but so are the two lowest. On the whole, the diagnoses seem to be scattered randomly among the test scores. The mean score for the nine cases of psychosis was 72.8, and that for the six psychoneurotic cases was 66.3, a difference which is quite insignificant statistically ($P > .40$). Breaking the set of T-scores into "High" and "Low" and then obtaining a chi-square on the resulting fourfold table, again shows an insignificant association between severity of T-score on the Hunt test and diagnosis ($\chi^2 = .028$, 1 d.f., $P > .80$). It should be recalled, however, that the numbers here become so very small that quite possibly the study of larger groups, of psychotics compared with neurotics, would yield a difference.

From these various findings, we may tentatively, with suitable caution because of the small sample, conclude that examiner judgments of validity, amount of upset shown by the patient, diagnosis of psychosis or neurosis, or a measure of depression such as that of the Multiphasic Depression scale, would not enable one to separate valid from invalid testings. It would seem, then, that the best approach is to either avoid giving the test to depressed patients at all, or look upon its results in such cases as indicators of loss in intellectual efficiency without implication of underlying organic pathology.

With a sample this small, such correlations mean little, but it may be worth while to indicate such trends as the relative absence of relation between the T-score and certain other variables. Since the T-score is based upon a deviation from the multiple regression plane (learning score regressed upon age and vocabulary), one would not expect any relation to exist here. Correlation of T-score with age is insignificant ($r = -.22$, $P > .40$) as is that with vocabulary ($r = -.14$, $P > .50$). The correlation of T-score with maximum grade reached in school is also insignificant , ($r = -.30$, $P > .20$).

Qualitative Observations

When questioned, the majority of the patients stated that they felt they could have performed better had they been tested before they became ill. And, it was observed that a number of them showed overt signs of upset such as crying, tremor, and peculiarities of voice and speaking rate. A few expressed a lack of interest in the proceedings, as would be expected in depressed persons. However, all were sufficiently cooperative to be willing apparently to attend to the test material; only two were inclined to admit that they were not really trying very hard.

The explanations patients gave of poor performance varied—that their thoughts tended to be on other things, that they felt too sad to care about the test, and in some cases that they were really trying to make a good showing but simply could not remember adequately. It was not possible, from either the quantitative or the qualitative data at hand, to form any clear hypotheses as to the manner in which depression interferes with the intellectual output.

However, it is likely that the simple fact of retardation could lead to a considerable elevation of T-scores, considering the rather split-second timing which the Hunt test employs. Preoccupation with "other things" is, of course, a possibility; but few would admit to this and, indeed, the examiners' impression is that this was not a very real factor, considering the more-than-adequate performance of the great majority of the cases when taking the interpolated tests.

Considering the foregoing, the writers are convinced that, on the basis of the subject's behavior in the testing situation, the examiner cannot adequately judge whether psychiatric upset is seriously impairing validity.

Summary

The Hunt-Minnesota Test for Organic Brain Damage was administered to a group of 15 persons with functional depressions, of whom nine were psychotic and six neurotic. All of these cases were between the ages of 34 and 55 years, and were neurologically and serologically negative for organic brain damage. None of them had a history of alcohol or drug addiction, head trauma, or encephalitis. All were cooperative to the extent of being willing to take the test and to apparently pay attention to the stimulus materials. Only one of the 15 was disturbed so greatly as to fail as many as three of the interpolated tests, and 11 subjects did not fail any of them. The findings were:

1. The mean T-score of the entire group was 70.2, with a SD of 15.41 points. Both the mean and the standard deviation differ significantly from a hypothetical supply with a mean of 50 and a SD of 10.

2. By the setting up of confidence belts for the estimation of population mean and variance, it is shown that at the very least, one can expect about one in four functionally depressed patients to have "pathological" scores (T > 70); or, setting the critical score at 66, about one in three patients.

3. The best estimate is that about half of functionally depressed patients may be expected to show scores over 70 on the Hunt test.

4. It is not possible, from the external manifestations of the patient's emotional disturbance, for the examiner to separate "valid" from "invalid" testings.

5. It is concluded that the Hunt-Minnesota Test for Organic Brain Damage, as it now stands, is not entirely specific for organic brain damage. Significant scores on this test obtained upon cases with depression as an important component of their illness cannot be interpreted except as a decrement in intellectual function of undetermined etiology.

It would be a mistake to extend this interpretation to the test scores of all patients with functional disorders, however, for over half of Hunt's original standardization group was composed of such cases. The mere presence of psychiatric involvement, as in a severe psychoneurosis, is by no means sufficient to invalidate the results, as will be shown by data soon to be published. However, examiners should interpret with caution a significant Hunt score which is obtained on a patient depressed to a considerable degree.

Received February 9, 1946.

References

Arkola, A. (1945). *The effect of sodium amytal upon performance on the Hunt-Minnesota test for organic brain damage.* Unpublished M.A. Thesis, University of Minnesota.

Babcock, H. [1930] *An experiment in the measurement of mental deterioration. Arch. Psychol*., No. 117.

Fisher, R. A. (1937). *Design of experiments*. London: Oliver and Boyd.

Fisher, R. A. (1941). *Statistical methods for research workers* (Eighth edition). London: Oliver and Boyd.

Hunt, Howard F. (1943a). A practical, clinical test for organic brain damage. *J. appl. Psychol.*, *27*, 375-886.

Hunt, Howard F. (1943b). *The Hunt-Minnesota test for organic brain damage.* Minneapolis: The University of Minnesota Press.

Hunt, Howard F. (1944). A note on the clinical use of the Hunt-Minnesota test for organic brain damage. *J. appl. Psychol.*, *28*, 175-178.

Hunt, Howard F. (1945). A note on the problem of brain damage in rehabilitation and personnel work. *J. appl. Psychol.*, *29*, 282-288.

Hunt, J. McV. (1944). *Personality and the behavior disorders.* New York: Ronald Press. [Not cited in text]

Peters, C. C., & Van Voorhis, W. R. (1940). *Statistical procedures and their mathematical bases.* New York: McGraw-Hill.

pdf by LJY, February 2016.