THE PROGNOSTIC VALUE OF THE METENYM TEST IN A FOLLOW-UP STUDY OF PSYCHOSURGERY PATIENTS AND THEIR CONTROLS

JOSEPH ZUBIN AND CHARLES WINDLE
Psychiatric Institute, Columbia University

INTRODUCTION

The purpose of this study was to determine the value of the Metenym (multi-meaning word) Test in predicting outcome of psychosurgery. This test, previously used by Cappa under the name of Homograph Test, consists of a list of common stimulus words: (note, dock, cell, hatch, pole, round, quarter, file, club, band, diamond, gin) each of which has at least five different meanings. The subject is required to give as many meanings as he can for each word. The score is the total number of different meanings (excluding the first) and has been regarded as a measure of verbal flexibility. The data presented here have been previously reported in the Columbia-Greystone study. At that time follow-up information was not available, but by the end of the second year after operation it was possible to select sixteen patients with clearly defined outcome. Eight patients who stayed out of the hospital continuously since the time of discharge (out-group) were contrasted with eight who had remained continuously in the hospital in an unimproved condition (in-group). Four patients in each of these two selected groups had been subjected to psychosurgery, the other four being controls.

DEVELOPMENT OF SCORING SYSTEM

A scoring system was developed similar to the one utilized in scoring the vocabulary of the Stanford Binet Test. Acceptable responses included definitions listed
in the Shorter Oxford English Dictionary\(^{(1)}\) and specific usages of these meanings. In the preoperative performance on the Metenym Test the average number of shifts in meaning (meanings in addition to the first) given by the in-group was 18.9, while the mean of the out-group was 14.0, a difference significant at the .10 level. A tabulation of the number of different meanings given for each stimulus word revealed that in ten of the words, the in-group exceeded the out-group, in one word the two groups were equal, while in the last word the out-group excelled. The probability of one group exceeding the other by chance alone in 10.5 out of 12 comparisons is less than .001, if the comparisons are assumed to be uncorrelated. It may be concluded that the in-group showed a higher capacity for shifting on this test than did the out-group. An item analysis of each word revealed that the critical point of differentiation between the two groups was the zero shift category (one response only), the incidence of non-shifting being more than twice as great in the out-group than in the in-group.

To achieve a minimum overlap between outcome groups, each word was weighted in accordance with the degree to which the frequency of zero shifts differentiated between the two criterion groups.\(^{*}\) The weights given for the first shift were as follows: note—10; dock—8; cell, hatch and pole—4; and the remaining words—2. For each additional shift a weight of one half of that given for the first shift was credited. The total weighted shift score obtained was subjected to an analysis of variance in order to determine the relationship of the preoperative score to outcome and operative status. The prognostic value of the test was significant at the .05 level, but the difference between the operated and control groups was negligible.

**Validity of the Scoring System**

Before accepting the difference between the in- and out-groups as dependent upon basic ability on the Metenym Test, it is necessary to investigate incidental factors which may produce differences in performance on this test. A comparison of the frequency of similar or repeated response definitions indicated no differences in this respect between outcome groups. Differences in word knowledge or in tendencies to give responses of differing degrees of difficulty were investigated by means of the Semantic Word Count of Lorge and Thorndike.\(^{(1)}\) No differences between outcome groups were found in tendency to give meanings of different levels of difficulty. The Metenym Test was found to correlate .67 with the Wechsler-Bellevue IQ. In order to determine whether the test derived any part of its prognostic strength from the IQ, the IQ level was partialled out by means of the analysis of covariance technique. The prognostic value of the Metenym Test remained on the .05 level.

**Prognostication for Total Group**

Our scoring system was developed empirically with criterion patients of clear-cut outcome. The next step was to apply the scoring system to all the patients in the Columbia-Greystone Project. This meant the addition of nineteen patients whose follow-up status was less well defined. They could, however, be classified into three groups: the in-group (unimproved), out-group (improved) and out-in group (fluctuating status). The mean score of the total in-group was 49.8, the mean of the out-group was 32.6, while the out-in group mean was 35.7. In an analysis of variance it remained clear that the preoperative test score predicted outcome significantly regardless of operative status.

**Postoperative Effects**

An analysis of variance of the immediate postoperative test scores obtained within the first month after operation indicated no significant differences between outcome groups or between operated and control groups. However, when the preoperative differences between groups were eliminated by the analysis of covariance, a significant difference did emerge. The postoperative test score of the operated group was significantly higher than the score of the control group.

\(^{*}\)This goal could have been obtained more directly by means of the discriminant function \(^{(1)}\), but in view of the smallness of the sample, inspection was used to determine the proper weights.
a significant difference between the operated and control groups was revealed. Apparently the operation either reduces the ability in the Metenym test immediately after operation or interferes with the usual improvement expected from repetition of the test. The preoperative IQ level was found to be significantly related to this loss, the diller patients tending to show losses more often than the brighter patients (P < .05), especially within the operated group. It may be concluded that the tendency to lose in the immediate postoperative testing is most marked in the diller operated patients regardless of eventual outcome. The second and third postoperative tests were administered approximately three months and one year after operation respectively. For neither testing period did analyses of variance or covariance reveal significant differences between outcome or treatment groups.

**Discussion**

Although no cross validation of these findings is now available, the fact that the prognostic value of the test is not altered when the sample size is increased to include patients whose outcome was less clearly determined lends further credence to the reliability of the findings. There are, however, two aspects of the results which are somewhat unexpected and therefore demand careful scrutiny. First, why should the patients who do poorly on this test have better prospects of recovery than those who do well? Second, why should the prognosis hold equally well for the operated and the unoperated? One is tempted to provide theoretical speculation to account for these two findings, but until there is verification from cross validation studies, it is fruitless to theorize.

**Summary**

The Metenym Test was found in retrospect to be prognostic of outcome in a two year follow-up study of 19 operated and 16 unoperated cases. The prognostic score obtained empirically on selected criterion cases held up when applied to the entire group of patients, but showed no difference between treatment groups. Contrary to previous studies, good performance on this test was not indicative of good, but of poor prognosis. The immediate postoperative testing indicated a loss in the operated group, but the expected practice gain in the unoperated group. These postoperative changes bore no relationship to outcome. Later postoperative testings were unrelated to either outcome or treatment groups.

**Bibliography**

1. Baruch, R. M., and Rust, R. Other tests. Ch. 22 in *.*
5. Landis, C. Test results one year after operation. Ch. 22 in *.*