A UNIDIMENSIONAL DEPRESSION SCALE FOR THE MMPI

PAUL DEMPSEY

University of California, Davis

To develop a depression scale that differentiates consistently not only between normal and abnormal samples but also within such samples, items from the MMPI D scale were selected on the basis of consistent relationship to the major dimension or factor underlying the 60 items as determined by a contextual analysis of responses in each of 4 normative samples of 40 Ss (normal and abnormal divided also by sex). The final 30-item scale, called the \( D_{30} \) scale, showed greatly improved within-group distinctions, particularly among normals, dimensionality coefficients (\( r_d \)) of .97, .95, and .98 being obtained in cross validation in contrast to values of .67, .45, and .37 for the original. Split-half reliability showed improvement despite shorter length, and test-retest estimates in 2 normal samples were .88 and .92. Since part-whole correlations indicated that \( D_{30} \) scale scores account for the systematic differences in scores on the 60-item scale, the \( D_{30} \) scale was recommended for general use in lieu of the original. T score norms based on 424 Ss were presented and scale content discussed.

The wealth of research stimulated by the MMPI is an aid to further exploration of the instrument, but at the same time it does exert certain constraints. Chief among these is the desirability of fitting subsequent work into the framework of knowledge and practice that has already built up around the test. Particularly is this important for research directed toward improving the MMPI as a measuring instrument, for any improvement that calls for so gross a change in either use or interpretation that past work is of questionable applicability is not an improvement at all. Rather it constitutes the introduction of a new instrument under circumstances calculated to produce maximum confusion.

This paper proposes changes in the MMPI Depression scale (Scale 2) that are basic and in some ways drastic, but nevertheless can reasonably be described as improvements. However gross the changes may appear to be, they require no alteration in established modes of using and interpreting the MMPI in general or Scale 2 in particular (Dahlstrom & Welsh, 1960). Instead, the net effect will be to make standard interpretations of the scale more accurate.

Specifically, this paper presents a Depression scale composed of 30 of the original 60 items and recommends that it be used in lieu of the original. The revised instrument, called the \( D_{30} \) scale, expresses the dimensional differences of the original, but has eliminated extraneous or error variation. Moreover, it consists of items that differentiate systematically not only between hospitalized and normal samples, but also within hospitalized samples and within normal samples. The present report will consider briefly the development of the \( D_{30} \) scale, but a comparison of the old and new scales as to dimensionality and other salient performance characteristics will be its main concern. The report will conclude with a description of scale content and the presentation of norms.

DEVELOPMENT OF THE \( D_{30} \) SCALE

The need for revision of the original Depression scale (here called also the \( D_{60} \) scale) reflects the changing use of the MMPI. The original scale was developed primarily to identify severely depressed persons rather than to make distinctions among them or among the members of any other group (Hathaway & McKinley, 1942). The items selected for the \( D_{60} \) scale were those that differentiated between groups: 49 between groups of normal subjects and an especially chosen criterion group of severely depressed patients, and 11 between the latter and a group of hospital patients not diagnosed as depressed. This approach to scale construction defines dimensionality in terms of consistent group
A Unidimensional Depression Scale for the MMPI

365

differences rather than consistent individual differences.

While there can be no objection to this approach for making group comparisons or for attempting to assign individuals to groups, it leaves much to be desired as a basis for individual comparisons either within or between groups. If, for example, two depressed patients obtain different scores, it is gratuitous to assume that the higher scorer is more depressed or more anything else than the other. Also, if a patient classified as non-depressed obtains a higher score than one classified as depressed, the method of scale construction provides no basis for interpretation. The inclusion of the 11 correction items makes the interpretation of individual differences even more equivocal. Particularly among normal subjects the presence of these items seems likely to increase nonsystematic or error variation. In short, while correlation studies, profile interpretation, and other current uses of the MMPI call for a depression scale whose dimensionality is based on consistent individual differences both within and between groups, the existence of such differences has rested primarily on assumption.

Until recently few studies of the MMPI have managed to avoid the a priori assumption of the dimensionality of the individual scales. Two notable exceptions were factorial studies of the intercorrelations of the items of the D60 scale, one by O'Connor, Stefan, and Gresock (1957) and the other by Conrey (1957). But while these studies did not assume the scale's dimensionality, neither did they test it. Both reported a complex factorial structure for the scale and recommended that it be interpreted in terms of one or more component parts, but neither presented evidence that the basic instrument was not dimensional. Two direct investigations of the dimensionality of the MMPI clinical scales have recently been carried out by the writer, however, and these indicate that the original D scale makes fairly consistent differentiations among abnormal subjects of both sexes but untrustworthy ones among normals. In a sample of male hospital patients of varying diagnoses the coefficient of dimensionality, $r_D$, was .65, and in a corresponding sample of females the coefficient was .87. Among ab-

normal subjects, in other words, $D_{60}$ scale scores accounted for 75% of the variance attributable to individual differences on the major dimension or factor underlying responses to the 60 items. In a parallel study of normal subjects (Dempsey, 1963) the coefficients of dimensionality were only .67 among males and .45 among females; $D_{60}$ scale scores accounted for only 45% of the dimensional variance among males and 20% among females. The lower dimensionality shown in the normal samples did not appear to be wholly a function of the more restricted range of talent among normals, however, for there was considerably more systematic variation than the scale scores successfully tapped. The presence of untapped systematic variation within both normal and abnormal samples made an attempt to improve the scale appear feasible.

To improve the $D_{60}$ scale so that it makes consistent differentiations within every subpopulation in which the MMPI is used calls for a manifold classification of subjects. In the present study a three-way classification was used. Two of the breakdowns were based on characteristics of the subjects, namely, normal-abnormal and male-female dichotomies. The third breakdown met a purely methodological requirement, the subjects being divided into normative and cross-validating groups. This three-way classification required eight separate samples. Four of these samples were available from the studies of dimensionality cited above (Dempsey, 1963). These constituted the normative group and were designated as follows: Sample 1, Normal males; Sample 2, Normal females; Sample 3, Abnormal males; and Sample 4, Abnormal females. Corresponding cross-validating samples have been designated as Samples 5–8. Every sample was composed of 40 subjects except Sample 8, which included 24, making the total N in the eight basic samples 304.¹

For purposes of test-retest reliability esti-

¹ The writer is grateful to Donald W. Mackinnon of the Institute of Personality Assessment and Research, University of California, Berkeley, and Robert E. Harris and the late G. Hamilton Crook, both of the Langley Porter Clinic, University of California Medical Center, San Francisco, for permission to use data gathered under their direction.
mates of the final $D_{30}$ scale, two additional cross-validating samples were used, one (Sample 9) composed of 34 normal males and the other (Sample 10) of 103 normal females. $T$ score norms for the $D_{30}$ scale are based on the performance of 424 subjects, 280 normals, and 144 hospital patients.

Further breakdowns of the subject samples by age and educational level would have been desirable, but the additional labor entailed was prohibitive. Instead a partial compensation for such differences was possible because age and educational level were relatively homogeneous within samples, but varied across samples. The normal females in the basic samples were lower-division college students, and the normal males upper-division and graduate students, while the hospital samples were somewhat older and showed a mixed educational background. The abnormal subjects represented a wide variety of diagnoses, with schizophrenia the most common classification.

If a scale has evidenced some degree of dimensionality, a straight-forward way to improve it is to eliminate the least consistent items. To do this for the $D_{30}$ scale, the procedure followed was to make use of the major dimension or factor underlying performance on the 60 items of the scale, as separately determined in each of the normative samples. The method used was contextual analysis (Dempsey & Baumhoff, 1963), a technique that was especially advantageous here because it yielded an explicit dimensional position or factor score for each subject and thus made possible a detailed analysis of individual items. Items that related in the same direction to the major underlying dimension in all samples met the minimum standard of acceptability. This may appear to be a far from exacting standard, but it must be remembered that the items of the $D_{30}$ scale have already survived extensive scrutiny. Except for the 11 correction items every one of them has been found to discriminate between normal and abnormal groups. What this additional standard asks in essence is whether the items are also consistent with individual differences in overall performance within normal and abnormal samples for both sexes. The dimensionality of the resulting $D_{30}$ scale indicates that the standard was sufficiently rigorous.

A total of 30 items met the minimum standard; hence the $D_{30}$ scale. The scale would have been several items longer if the criterion had specified consistent differentiation only among normals; it would have been considerably shorter if items showing a zero frequency among normals were not included. It is interesting to note that not 1 of the 11 correction items met the minimum standard. One anomaly occurred. MMPI Item 39, "At times I feel like smashing things," is scored false on the $D_{30}$ scale, but in every one of the normative samples, answering false related negatively to the underlying dimension. Further, the false response occurred with greater frequency among normals than normals. When these relationships also appeared in each of the cross-validating samples, Item 39 was kept in the final scale, but was keyed true.

To evaluate the $D_{30}$ scale several comparisons were made. The most important of these was testing the dimensionality of the new scale in cross validation. The responses to the 30 items made by the subjects in Samples 5, 6, and 7 were subjected to contextual analysis, each sample being analyzed separately. This provided for the subjects of each sample contextual or factor scores indicating their positions on the major dimension underlying responses to the 30 items. All subjects were also scored on the $D_{30}$ scale by using the appropriate portion of the original $D_{30}$ scoring key (with the modification for Item 39). The final step was to calculate the major coefficient of dimensionality, $r_a$, in each of the samples by correlating contextual scores with $D_{30}$ scale scores. The closer this coefficient approaches +1.00, the more closely scale scores account for systematic differences in performance, and hence the more nearly dimensional the scale.

The coefficients of dimensionality shown by the $D_{30}$ scale in cross validation are given in Table 1. Among normal males (Sample 5), $r_a = .97$; in the sample of normal females (Sample 6), $r_a = .95$; in the sample of male hospital patients (Sample 7), $r_a = .98$. These figures may be contrasted with the dimen-
sionality coefficients obtained for the original $D_{70}$ scale in the corresponding normative samples. These coefficients were .67, .45, and .86, respectively, as reported above. Thus the $D_{80}$ scale represents a considerable improvement in dimensionality over the original Depression scale, especially among normal subjects.

**OTHER SALIENT PERFORMANCE CHARACTERISTICS OF THE $D_{70}$ SCALE**

Besides dimensionality, two other performance characteristics of the $D_{70}$ scale are of concern here: its reliability and its interpretability vis-à-vis the original $D_{60}$ scale. The latter characteristic is the more crucial, but since the former can be summarily dealt with, it will be considered first.

For scales composed of items having different levels of difficulty, alternate form or test-retest estimates of reliability are generally more appropriate than are split-half estimates (Guilford, 1956). In the eight normative and cross-validating samples of the present study, however, split-half estimates were the only ones readily obtainable for both the $D_{60}$ and $D_{80}$ scales. These estimates are presented in Table 2. In all eight samples the reliability coefficient of the $D_{70}$ scale is higher than the corresponding coefficient of the original scale, even though the latter is twice as long. Both scales show adequate reliability in abnormal samples, with coefficients ranging from .85 to .95, but only the $D_{80}$ scale shows coefficients at a reasonable level in the normal samples, the range being from .77 to .86. These last contrast with coefficients of .53 to .68 found for the longer $D_{60}$ scale among the same subjects.

A test-retest reliability estimate was obtained in an undergraduate class in introductory psychology, the students retaking the $D_{60}$ scale after an interval of from 3 days to 3 weeks. The reliability coefficient amongst the 103 girls was .88, and amongst the 34 boys was .92. These results are shown in Table 2 under the headings of cross-validating Samples 9 and 10.

The original MMPI $D$ scale can be split into two halves that are of special interest. One half is the $D_{70}$ scale; the other is composed of the 30 items excluded from the original scale in order to improve dimensionality. These latter will be called here the $D_{ex}$ items for purposes of identification. As can be seen in Table 3, the correlation of these two halves of the original scale is negligible in every normative and cross-validating sample tested, the coefficients ranging from high values of .31 and .27 to minimum values of -.07 and -.09. There are several ways of accounting for the fact that there is no significant relationship between these two components of the original scale, the most plausible being that several small independent factors are included among the 30 $D_{ex}$ items. Such an interpretation is consistent with the large number of small factors reported by Comrey (1957) in his factor analysis of the $D_{60}$ items.

**TABLE 2**

**RELIABILITY OF THE $D_{70}$ SCALE AND THE ORIGINAL MMPI DEPRESSION SCALE ($D_{60}$)**

<table>
<thead>
<tr>
<th>Sample</th>
<th>$D_{70}$</th>
<th>$D_{60}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Normal males</td>
<td>.84</td>
<td>.61</td>
</tr>
<tr>
<td>2 Normal females</td>
<td>.89</td>
<td>.53</td>
</tr>
<tr>
<td>3 Male hospital patients</td>
<td>.89</td>
<td>.85</td>
</tr>
<tr>
<td>4 Female hospital patients</td>
<td>.95</td>
<td>.85</td>
</tr>
<tr>
<td>Cross-validating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Normal males</td>
<td>.86</td>
<td>.68</td>
</tr>
<tr>
<td>6 Normal females</td>
<td>.77</td>
<td>.64</td>
</tr>
<tr>
<td>7 Male hospital patients</td>
<td>.91</td>
<td>.83</td>
</tr>
<tr>
<td>8 Female hospital patients</td>
<td>.95</td>
<td>.90</td>
</tr>
<tr>
<td>9 Normal males</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td>10 Normal females</td>
<td>.88</td>
<td></td>
</tr>
</tbody>
</table>

Note.—$N=40$ in all samples except 8, 9, and 10 where $N=24, 34$, and 103, respectively. *All entries corrected split-half estimates except for test-retest estimates in Samples 9 and 10.

**TABLE 1**

**DIMENSIONALITY OF THE $D_{70}$ Scale in Cross Validation**

<table>
<thead>
<tr>
<th>Sample</th>
<th>$r_{xy}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Normal males</td>
<td>.97</td>
</tr>
<tr>
<td>6 Normal females</td>
<td>.95</td>
</tr>
<tr>
<td>7 Male hospital patients</td>
<td>.98</td>
</tr>
</tbody>
</table>

Note.—$N=40$ in each sample.
TABLE 3
PART-PART AND PART-WHOLE CORRELATIONS OF THE ORIGI NAL MMPI D SCALE (D90), THE IMPROVED D SCALE (D80), AND THE THIRTY DISCARDED ITEMS (D10)

<table>
<thead>
<tr>
<th>Sample</th>
<th>D90 versus</th>
<th>D80 versus</th>
<th>D10 versus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dα</td>
<td>Dα</td>
<td>Dα</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>r</td>
<td>r</td>
</tr>
<tr>
<td>Normative</td>
<td>Dα</td>
<td>Dα</td>
<td>Dα</td>
</tr>
<tr>
<td>Normal male</td>
<td>-.18</td>
<td>.33</td>
<td>.85</td>
</tr>
<tr>
<td>Normal females</td>
<td>-.13</td>
<td>.46</td>
<td>.79</td>
</tr>
<tr>
<td>Male hospital patients</td>
<td>.19</td>
<td>.58</td>
<td>.90</td>
</tr>
<tr>
<td>Female hospital patients</td>
<td>-.07</td>
<td>.40</td>
<td>.88</td>
</tr>
<tr>
<td>Cross-validifying</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal males</td>
<td>-.09</td>
<td>.42</td>
<td>.82</td>
</tr>
<tr>
<td>Normal females</td>
<td>-.27</td>
<td>.38</td>
<td>.78</td>
</tr>
<tr>
<td>Hospital patients</td>
<td>.10</td>
<td>.53</td>
<td>.89</td>
</tr>
<tr>
<td>Female hospital patients</td>
<td>.31</td>
<td>.62</td>
<td>.93</td>
</tr>
</tbody>
</table>

Note.—N = 40 in all samples except 8, when N = 24.

The fact that negligible correlations exist between these two halves of the original scale makes it possible to view subjects's scores on the original as the sum of one highly dimensional component, the D90 scale scores, and of a second component that has only a chance relationship to the first. This is like saying that the D90 scale scores are made up of the D80 scores plus chance variation. But if this is the case, there should be gross differences in the part-whole correlations between D80 and D90 on the one hand and between Dα and D80 on the other. These part-whole relationships in the eight basic normative and cross-validating samples are shown in Table 3. For the Dα component, the coefficients range from .33 to .62, with a median of .44. For the D90 component the coefficients are considerably higher, ranging from .78 to .93, with a median of .86. These differences indicate that the overwhelmingly important component of the original scale is represented by the D90 scale. It seems likely that whatever consistencies of interpretation the original scale has made possible in the past are attributable to its D90 component. Indeed, if the unrelated contribution of the Dα items were eliminated, there is good reason to believe that the standard interpretations of the scale could be made more consistently and more accurately. In short, it appears reasonable to use the D90 scale in lieu of the original Depression scale.

CONTENT AND NORMS

The items of the D90 scale are listed in Figure 1, together with their direction of scoring and their order number in the booklet form of the MMPI (Hathaway & McKinley, 1951). In content the items are all fairly straightforward, each calling for the admission of what the subject undoubtedly considers a limitation or weakness. Few of them refer directly to feelings of unhappiness or depression; rather they refer to conditions, perceived or actual, under which depression occurs. The specific somatic items and the fundamentalist religious items of the original scale have disappeared, along with the other items included for purposes of correction.

THE D90 SCALE

1. My daily life is full of things that keep me interested. (8, F)
2. I am about as able to work as I ever was. (9, F)
3. I find it hard to keep my mind on a task or job. (32, T)
4. At times I feel like smashing things. (39, T)
5. I have had periods of days, weeks, or months when I couldn’t take care of things because I couldn’t “get going.” (41, T)
6. My sleep is fitful and disturbed. (43, T)
7. I prefer to pass by school friends, or people I know but have not seen for a long time, unless they speak to me first. (32, T)
8. I am a good mixer. (57, F)
9. I wish I could be as happy as others seem to be. (57, T)
10. I am certainly lacking in self-confidence. (86, T)
11. I usually feel that life is worth while. (88, F)
12. I don’t seem to care what happens to me. (104, T)
13. I am happy most of the time. (107, F)
14. I seem to be about as capable and smart as most others around me. (122, F)
15. I do not worry about catching diseases. (131, F)
16. Criticism or scolding hurts me terribly. (138, T)
17. I certainly feel useless at times. (142, T)
18. Most nights I go to sleep without thoughts or ideas bothering me. (152, F)
19. During the past few years I have been well most of the time. (153, F)
20. I cry easily. (158, T)
21. I cannot understand what I read as well as I used to. (159, T)
22. I have never felt better in my life than I do now. (160, F)
23. My memory seems to be all right. (178, F)
26. I enjoy many different kinds of play and recreation. (207, F)
27. I brood a great deal. (236, T)
28. I believe I am no more nervous than most others. (242, F)
29. I have difficulty in starting to do things. (259, T)
30. I work under a great deal of tension. (290, T)

In a very real sense the items represent different levels of difficulty, for each tends to differentiate at a specific part of the scale. The admission of chronic ill health (Item 19), for example, has a zero frequency in the lower reaches of the scale, occurring only among the highest scorers. While there are undoubtedly chronic invalids whose spirits are generally good, still prolonged illness, projected or real, is a potent depressive factor. Similarly, the environmental or perceptual poverty reported in Item 1 distinguishes not at all among high scorers, all of whom admit to boredom, nor among low scorers, all of whom deny it; Item 1 differentiates in the middle ranges of the scale. Two easy items, i.e., items that differentiate only at the lower reaches of the scale, are: 9. I wish I could be as happy as others seem to be, and 10. I am certainly lacking in self-confidence. Moderate and high scorers almost invariably get points on these items.

There are no sex differences in total scores on the $D_{90}$ scale, but there are some interesting differences on individual items. The most striking of these perhaps is Item 20—I cry easily. This item differentiates consistently in both sexes, but at greatly different levels of the scale. Among females, it is a relatively easy item, differentiating at the low to moderate levels. Among males, on the other hand, it appears almost pathognomonic of severe depression, occurring among only the very highest scorers in the hospital samples.

The items of the $D_{90}$ scale show extensive overlap with those of a scale suggested by Comrey (1958) as a result of his factor analysis of the original Depression scale in a sample which included both normal and abnormal subjects. Of the 33 items that make up the two scales, 23 are common to both, and two more are listed by Comrey as optional. Such a high level of agreement from different methods and different samples undoubtedly contributes to the trustworthiness of both scales, but it is the differences between them that are of interest here. An analysis of the trace lines of these items in the various samples of this study indicates that the differences arise because of the test behavior of normal subjects. The items unique to Comrey's scale show very poor differentiation among normals, while the items appearing solely on the $D_{90}$ scale differentiate consistently among them. On the basis of these differences, it seems likely that the $D_{90}$ scale will be considerably more effective than Comrey's for work with normal subjects, but that both will do well with abnormal subjects.

$T$ score norms for the $D_{90}$ scale are presented in Table 4. These norms are based on the performances of 424 subjects, 280 normals, and 144 hospitalized patients. As may be expected, the distribution is skewed among normal subjects, although it appears to be fairly symmetrical among hospital patients. Only about 3% of the subjects in the normal group scored as high as the mean of the mental patients, which in $T$ score terms occurs very close to 70. The lowest $T$ score, 37, occurs about 3% of the time; the highest possible $T$ score, 107, did not appear in the normative sample. It is of course too early to say what effect the use of $D_{90}$ will have on the relative frequency with which Scale 2 appears as a profile peak either among normals or among abnormal subjects (Hathaway & Meehl, 1951).
REFERENCES


(Received September 30, 1953)