

ITEM CHARACTERISTICS OF THE MMPI
EMOTIONAL DISORDER (Ed.) SCALE IN
A NON-AMERICAN SAMPLE¹

E.O. OLADELE

The emotional disorder scale of the MMPI (Dahlstrom, Welsh & Dahlstrom 1975) is very likely to attract the attention of clinicians or researchers who are looking for simple timesaving instruments suitable for use in the general population. The need to employ quick and efficient procedures for survey and screening purposes in Community mental health is obviously greater in the developing countries where there are serious shortages of resources, including highly trained mental health professionals.

However, since the emotional disorder scale was constructed and validated in the United States of America, using American subjects within an American culture, it seems proper to question the appropriateness of its use in a non-American culture or setting.

The rationale for such a question lies in the fact that performances on such scales are often influenced by factors other than the amount of the construct present in the subject. Butler & Jones (1979) have noted that most items in mental health scales represent either mild forms of psychological disorder or types of distress that are sensitive to environmental situational factors. In

¹ The author wishes to express his gratitude to Professor E.O. Adebo, head of the Department of Nursing, University of Ibadan, Nigeria for her assistance: during the process of data collection for this study.

other words, emotional disorder is by no means a simple construct to define or measure.

/p. 312/ A review of the psychological literature showed that there has been very little published information concerning the use and the psychometric characteristics of this scale outside the United States. This is rather surprising, given the scale's apparent simplicity and the amount of innovative efforts that went into its construction and validation. One would expect it to be, at least, as popular as its older counterparts (i.e. MacMillan 1957; Gurin 1960; Langer 1962).

Generally speaking, the component items of the emotional disorder scale do not possess the degree of homogeneity required for internal consistency reliability. The advantages and disadvantages of this fact are not addressed in this paper. However, it is important to bear in mind that this is not a unifocal scale and that the commonly used indices of test reliability designed to measure internal consistencies of uni-dimensional scales are less efficient when applied to multidimensional scales with heterogeneous items. The complexity of the concept of psychometric reliability as applied to inventory scales is highlighted by Dahlstrom, Welsh & Dahlstrom (1975). However, this complexity does not in any way invalidate attempts to evaluate such scales or their component parts; it only calls for a modification of one's procedure and interpretation.

The present study was an attempt to investigate the characteristics of the component parts of the MMPI emotional disorder scale, with particular references to the discriminatory power and the construct validity of each item.

The writer is aware that the term "item discrimination" is sometimes used to include both internal consistency item discrimination and item validity (Davis 1951). However, for the purpose of this study, the term "item discrimination" is intended to mean the extent to which an item differentiates or segregates persons scoring higher on the emotional disorder scale from those scoring lower on the same scale. On the other hand,

the term “item validity” is used in a sense similar to Vernon’s (1948) concept of “item consistency”, i.e.; the extent to which an item measures what the total score on the scale is measuring, a sort of construct validity. Here also, “item popularity is used to refer to the more commonly encountered /p. 313/ concept” of item difficulty (proportion marking the item in the keyed direction) since this is a nonaptitude scale.

Method

Subjects and Setting

The 35 subjects of this study were first year undergraduate students enrolled in a nursing degree programme at one of Nigeria’s main Universities. This number constituted 78% return of a set of questionnaires distributed for voluntary completion and return by interested participants. Respondents were all Nigerians from various ethnic backgrounds (12 of Nigeria’s 19 states were represented).

Subjects’ ages ranged from 23 to 36 years with a means of 26.74 years and a standard deviation of 3.00 years. The mean age for the males was 27.36 years with a standard deviation of 3.23 years, while the females had a mean age of 26.46 years and a standard deviation of 2.84 years. There were 24 females and 11 males in this group of professional nurses.

Procedure

A set of questionnaires including the 20 items of the MMPI emotional disorder scale was administered to a group of volunteer undergraduates in a classroom setting. It was made clear to participants that this was not a timed “test” and that there were no right or wrong responses to any of the items.

The questionnaires included measures of self-consciousness (11

items), conscience (12 items), and a measure of the subject's self-concept (20 items). Only 18 out of the 20 items emotional disorder scale were utilized in this study because 2 items 1 (F); and 18 (I)² were answered in the same (healthy) direction by all subjects.

A number of techniques which have been advocated as suitable for item analyses when samples are small were employed in the present study. They included the following: /p. 314/

1. The upper-lower number-right difference (d) as an index of item discriminatory power. The table provided by Olukpe (1982) was considered as applicable to samples of widely varying sizes (from N = 8 to N = 1000). When the sample size is as large as 12 or more, the researcher may use any of three alternative proportions (i.e. P = .5; P = .33; and P = .25) of the sample from each tail of the distribution as long as the distribution could be assumed to be normal.

2. The Upper-Lower Index proposed by Johnson (1951) advocates the comparison of the responses of 27 % of the sample from each tail of the distribution.

The proportions are regarded as being independent of the size of the sample. Formulas for determining the standard error as well as the difficulty level of each item are provided.

3. Cureton's Chi test has been described by DAVIS (1951) as more appropriate than many other tests of item discrimination when the samples are small. The Chi test requires no assumptions regarding the shape of the distribution of scores on the variable measured by the items.

4. Point-biserial correlation (item-total) was computed between the total scores and scores on each of the 18 items utilized. Henrysson's (1963) formula for correcting point-biserial correlation for spurious overlap was applied to each of the coefficients obtained. The main purpose here was to examine how well each of the items measured what the scale as a whole was measuring. Similarly, intercorrelations among the compo-

² MMPI item numbers 305 and 2 (booklet form).

nent parts of the scale were computed to see how closely they related to one another and how closely each part related to the total scores.

5. The contribution of each item to the total variance of scores on the scale was estimated, using a formula suggested by Ferguson (1971). Such contribution to the total scale variance, which utilizes information regarding item popularity and discriminating power is deemed to reflect the quality of each item. In other words, the more an item contributes to the total variance, the better that item is. The interested reader is also referred to Ebel's (1967) formula for the relationship between the total variance on a test and the discriminatory power of the items on that test. A direct relationship between $\sqrt{p \cdot q}$ mean discrimination index and test reliability was also suggested from this formula.

6. Inter-correlations and internal consistencies of the different parts of the scale were also estimated.

Results

The Upper-lower number-right difference (d) was very similar to the upper-lower index as indicated by the number of items selected as fair or good ones by each technique. With $P = .50$ and $N = 17$, the (d) technique selected 12 items as fair or good while the upper-lower index (27% from each tail of the sample distribution) selected 13 items with similar discriminative qualities. Table 1 shows the items which were selected by each technique and the 11 items which were selected by both methods as statistically discriminative at the 10% significance level (one-tailed). It should be noted that only 4 items (Nos 6, 7, 8 and 20) were selected by the (d) method when $P = .33$ and $N = 12$. The remaining 14 items were categorized as poor.

According to Cureton's Chi test, only 3 items (all of them belonging to the psychologic symptom index part of the scale) were found to be statistically discriminative at the 5% level of

significance. These were items 6, 7 and 8, which were also selected as good item by both methods described above.

Table 2 shows the corrected item-total point-biserial correlation for each item, the popularity of individual items, as well as the percentage of total variance contributed by each item. The mean item-total *r* (corrected) was .217 while the item popularity ranged from 5 to 50 per cent. The mean discrimination index (D) of the component items of the scale (18 items) estimated by Ebel's (1967) formula was .30 (.29 by direct calculation). This would imply that quite a good number of the items were of marginal discriminative quality.

/p. 316/

TABLE 1

The discriminatory powers of emotional disorder scale items as reflected by their selection or rejection by the (d) technique. The upper-lower index and both techniques Selected or rejected by

Item N° ³	(d)	U-L Index	Both methods
2	+	+	++
3	-	-	
4	+	+	++
5	+	+	++
6	+	+	++
7	+	+	++
8	+	+	++
9	-	-	
10	-	-	
12	+	+	++
13	+	-	
14	-	+	
15	-	-	
16	+	+	++
17	-	+	
19	+	+	++
20	+	+	++
22	+	+	++

- rejected
+ Selected

³ Item 1, 11 and 18 are excluded ; 20 and 21 are equivalents of MMPI item number 230.

++ Selected by both

TABLE 2
Corrected item-total r , item popularity and % of total variance contributed
By each item

(Ed) Item N ^{o4}	r_{pbis} (corrected)	Popularity	% Variance Contributed
2	.060	15	4.5
3	.254	5	1.9
4	.198	10	3.3
5	.051	15	3.6
6	.483*	50	12.2
7	.476*	25	9.7
8	.482*	50	12.6
9	.140	5	1.3
10	-.043	10	1.6
12	.413**	15	7.0
13	.137	20	5.7
14	.201	25	5.4
15	.119	30	5.8
16	.031	50	5.5
17	.122	10	2.0
19	.145	25	4.1
20	.486*	40	10.4
22	.0145	10	3.4

* $P < .01$

** $P < .05$

TABLE 3
The inter-correlations of the parts of the emotional disorder (ed)
scale with one another and with total scores

	1	2	3	4	Total
1. Psychological		.255	.200	.206	.835
2. Psychophysiological			-.258	.192	.569
3. Physiological				0.13	.286
4. Ambiguous					.540

⁴ Item 1, 11 and 18 are excluded ; 20 and 21 are equivalents of MMPI item number 230.

Other Correlations

When taken together, the 5 items with statistically significant point-biserial correlations ($P < .05$ or better) correlated with the total scores .83 and had an internal consistency reliability of .63 (K-R 20). These items also contributed 51.9% of the total variance of scores on the scale in this sample.

The 10 items with the highest individual contributions to the total scale variance as indicated in table 2 correlated .91 with the total scores. They contributed 78.8% of the total variance and had an internal consistency of .54 (K-R 20). Obviously, the lack of homogeneity of the items is reflected in such a low internal consistency.

The correlations of scores on these 10 items with scores on the psychological symptom index and the psychophysiological symptom index were .73 and .63 respectively.

The internal consistency reliabilities of the scale and its parts were also estimated by the use of Kuder-Richardson formula 20 (K-R 20).

Results were as following:

Total scale (18 item; .50), psychological symptom index (9 items; .37), psychophysiological symptom index (4 items; .35), physiological symptoms (2 items; .35), and ambiguous items (3 items; .37).

Discussion and conclusions

The techniques used in the present study are among those commonly considered as suitable when the samples are relatively /p. 318/ small. However, it is still necessary to be cautious in interpreting the findings, especially when attempting to generalize the results to individuals or groups outside this sample. The homogeneity (all subjects were professional nurses) and the non-randomness of the sample also need to be borne in mind.

In this study with $N = 35$, the (d) index of item discrimina-

tory power did not prove to be more efficient than the upper-lower index earlier suggested by Johson (1951). In fact, it appeared to be less effective when $P = .33$, being able to select only the best items and excluding items of fair or moderate qualities. However, this is consistent with the caution offered by Oluikpe (1982) that, when dealing with small samples, the higher the proportion used, the more confidence the result deserves.

Cureton's Chi test, as applied to this sample, seemed to be quite stringent. Only 3 items (6, 7 and 8) were identified as statistically discriminative ($P < .05$). These 3 items were among the most valid ones in the scale and were always selected as good items no matter what method was used.

It is the opinion of the present writer that the MMPI emotional disorder scale, as it is presently constituted, needs to be modified and validated locally before it can be used with a reasonable degree of confidence among groups of subjects similar to those of the present study.

In addition to the need for large heterogeneous samples, due consideration should be given to the social desirability factor and the semantic aspect of item construction. For example, the statement "I have a good appetite" (item 18), which all the subjects of this study answered in the healthy way, could hardly be expected to be answered otherwise in this culture. This is mainly because the quality of one's appetite is often considered as an important indicator of that person's health. In other words, a statement such as "I have a poor appetite" is frequently intended to convey a message similar to saying "I am not well". Obviously, relating the quality of appetite to the present or current health status is only one of the ways in which appetite can be viewed.

Those 10 items that contributed the highest percentages of the total test variance would probably be most useful /p. 319/ as the nucleus for a reconstructed and improved scale to measure emotional disorders in a non-patient population in settings similar to that of the present study.

It is hoped that this paper would be of an adequate heuristic

value in stimulating the interest of researchers and Clinicians concerning the study and use of the MMPI emotional disorder scale. An improved version of the scale could be quite useful outside the United States, especially in situations where simple and efficient estimates of emotional disorders are needed for research and screening purposes.

Emmanuel O. OLADELE,
M.S. (Psychiatric-Mental Health Nursing),
MA. (Clinical Psychology), Boston University, USA.
Lecturer, Department of Nursing
University of Ibadan, Ibadan (Nigeria).

REFERENCES

- BUTLER M.C. & JONES A.P. (1979) "The health Opinion Survey reconsidered: dimensionality, reliability, and validity." *J. Clin. Psychol.* 35: 554-559.
- DAHLSTROM W.G., WELSCH G.S. & DAHLSTROM L.E. (1975) *An MMPI Handbook. Vol. II: Research applications*. St. Paul, North Central Publishing Co.
- DAVIS F.B. (1951) "Item Selection techniques". In B.F. LINQUIST (ed.) *Educational Measurement*. Washington D.C., American Council on Education.
- EBEL R.L. (1967) "The relation of item discrimination to test reliability". *J. Educ. Meas.*, 4: 125-128.
- ELY J .H. (1951) "Studies in item analysis 2: Effects of various methods upon test reliability." *J. App. Psychol.* 35: 194-203.
- FERGUSON GA. (1971) *Statistical analysis in Psychology and Education* (3rd Bd.) New York, McGraw-Hill Book Company: 359-360.
- GUILFORD J.P. (1965) *Fundamental Statistics in Psychology and Education*. New York, McGraw-Hill Book Company.
- GURIN G., VEROFF J. & FELD S. (1960) *Americans view their mental health: a nation-wide survey*. New York, Basic Books.
- /p. 320/ HENRYSON S. (1963) "Correction of item-total Correlations in item analysis". *Psychometrika* 28: 211-218.
- JOHNSON A.P. (1951) "Notes on a Suggested index of item validity: the U-L Index." *J. Educ. Psychol.* 42: 499-504.
- KILLY T.L. (1939) "The Selection of upper and lower groups of the Validation of test items." *J. Educ. Psychol.* 30: 17-24.
- LANGNER T.S. (1962) "A twenty-two item screening scale of psychiatric Symptoms indicating impairment." *J. Health and Human Beh.* 3: 269-276.

- MACMILLAN A.M. (1957) "The Health Opinion Survey: techniques for estimating prevalence of psychoneurotic and related types of disorders in communities." *Psychol. Rep.* 3: 315-339.
- MANIS, J.G. BRAWER, M.J., HEND, C.I. & KERCHER I.C. (1964) "Validating a mental health scale." *Am. Soc. Rev.* 29: 84-89.
- OLUIPKE C.G. (1982) "A table of upper-lower number-right difference (d) as an index of item discriminatory power." *Br. J. Educ. Psychol.* 52: 121-124.
- ORTAR C.R. (1963) The transfer of psychological diagnostic measure from one culture to another." *Acta Psychologica* 21: 218-230.
- PYRCZAK F. (1973) "Validity of the discrimination index as a measure of item quality." *J. Educ. Meas.* 10: 221-231.
- RICHANDSON M.W. (1936) "Notes on the rationale of item analysis." *Psychometrika* 1, 1: 69-76.
- SWINEFORD F. (1936) "Validity of test items." *J. Educ. Psychol.* 27: 68-78.
- VERNON P.E. (1948) "Indices of item consistency and validity." *Br. J. Psychol. Stat.* Section, 1: 152-166.
- WOLF R. (1967) "Evaluation of Several formulae for correction of item-total correlations in item analysis." *J. Educ. Meas.* 4, 2: 21-25.
/p. 321/

SUMMARY:

Items on the MMPI emotional disorder scale were evaluated, using techniques commonly regarded as suitable for small samples. The subjects were 35 undergraduate Nigerian students. Analysis of 18 out of the 20 items comprising the scale indicated that it needed considerable modification and local validation in order to make it suitable for use with subjects having similar backgrounds with those of this study. The roles of semantics as well as the social desirability factor were also stressed.

Key words: • MMPI (Minnesota Multiphasic Personality Inventory)
• Semantics • Screening • Consistency • Reliability • Multidimensional • Validity • Discriminative • Popularity • Homogeneity
• Nigeria • Undergraduate • Students.

RÉSUMÉ :

CARACTÉRISTIQUES DES ITEMS DE L'ÉCHELLE
"DÉSORDRE ÉMOTIONNEL" (Ed) DU MMPI
DANS UN ÉCHANTILLON DE SUJETS NON AMÉRICAINS
(NIGÉRIA)

Les items de l'échelle "désordre émotionnel" du MMPI ont été évalués suivant les techniques habituellement utilisées dans l'étude de petits échantillons. La population étudiée était composée de 35 étudiants Nigériens. L'analyse de 18 des 20 items de l'échelle montre que celle-ci nécessite des modifications considérables et une validation locale pour la rendre utilisable dans une population partageant des caractéristiques de base communes à l'échantillon d'étude. Le rôle de la sémantique aussi bien que celui du facteur de désirabilité sociale sont soulignés.

Mots clés : • MMPI (*Minnesota Multiphasic Personality Inventory*) • Échelle de personnalité • Sémantique • Criblage • Consistance • Sûreté • Multidimensionnel • Validité • Discrimination • Popularité • Homogénéité • Nigéria • Étudiants.