Brit. J. Psychiat. (1972), **121**, 411–16

A Method for Measuring Depression: Validity Studies on a Depression Questionnaire

By I. PILOWSKY and D. SPALDING

Introduction

In previous papers (Pilowsky, Levine and Boulton, 1969; Pilowsky and Boulton, 1970; Pilowsky and McGrath, 1970) we have described the development of a depression questionnaire and its use in the classification and identification of depressed patients. In order to increase the flexibility of the questionnaire as a research tool we have derived a depression scale from the items in it and have investigated its validity as a measure of the depth of depression. The purpose of this paper is to describe the development of the depression scale and report the results of validation studies.

Subjects and Methodology

This study was performed in two stages. In the first stage the Depression Score was examined in relation to admission diagnoses, taxonomic classification and sensitivity to clinical change. Subjects for this part of the study were 80 female in-patients, all of whom were tested within a week of admission to the ward. The average age of the sample was 46·8 years (S.D. 15·6). Patients not included were the non-English-speaking, the excessively uncooperative and those with intellectual impairment. In addition, patients whose clinical diagnoses indicated organic cerebral involvement were later excluded. Testing was individual and in an office where privacy and comfort were reasonable.

In the second stage of the study, relationships between the Depression Score and two concurrent measures of depression were examined. The subjects involved in this stage were 37 in-patients (13 male, 24 female) and 23 out-patients (9 male, 14 female). The average age of this sample was 40·9 years (S.D. 14·2); males 38·0 years (S.D. 15·6); females 42·6 years (S.D. 13·1).

The Depression Questionnaire

The Depression Questionnaire (Pilowsky *et al.*, 1969) is a 57 item inventory requiring YES-NO answers.

A depression scale was based on items in this questionnaire, chosen on the basis of data gathered for the original classification study (Pilowsky *et al.*, 1969). Of the 200 subjects in that study, 106 were diagnosed by a clinician as depressed and 94 were not depressed. An item analysis showed that 25 of the questionnaire items discriminated between the depressive and non-depressive groups at the 0·05 level of significance. Of these 25 items, 19 were also found to correlate significantly ($p < 0·05$) and negatively with the taxonomic classification of non-depression (Class 'C' or '3'). These 19 items were selected for the depression scale and are shown in Table I with the responses contributing to the depression score *italicized*.

Relationship to diagnosis

Patients were grouped in categories according to the admission diagnosis (Table II), which was made known to the research team after all testing had been completed and before any results were disclosed to the ward psychiatrist.

Relationship to taxonomic classification

The relationship between the Depression Score and the LPD classification of depression (Pilowski *et al.*, 1969) was examined briefly. The subjects were classified into three groups using the decision rule described by Pilowsky and Boulton (1970).* Classes A, B and C of the previous study correspond with Classes 1, 2 and 3 respectively in this; Class 1 being 'non-

* Copies of the scoring sheets for the decision rule are available on request.

https://doi.org/10.1192/bjp.121.4.411 Published online by Cambridge University Press

TABLE I

Questionnaire items contributing to the depression score

1. Are you more irritable towards other people? ..	*Yes*	No	
2. Have you lost interest in watching television? ..	*Yes*	No	
4. Do you feel depressed all day long?	*Yes*	No	
5. Do you feel slowed up in your thinking?	*Yes*	No	Observer
10. Have you moved house in the past year?	*Yes*	No	
12. Is it more difficult to concentrate on your work? ..	*Yes*	No	Self
17. Are there times when you do not feel depressed? ..	Yes	*No*	
18. Do you have less interest in reading newspapers? ..	*Yes*	No	Self
22. Is your appetite normal? ..	Yes	*No*	
24. Do your feel you are a burden to others?	*Yes*	No	
30. Do you suffer from a dry mouth?	*Yes*	No	
35. Do you wish you were dead?	*Yes*	No	
40. Do you have less interest in things you usually enjoy? ..	*Yes*	No	
44. Is your depression the same all day long?	*Yes*	No	Sel.
46. Do you feel life is not worth living?	*Yes*	No	Self
51. Do you feel slowed up in doing things?	*Yes*	No	Observer
55. Do you feel time passing more slowly?	*Yes*	No	
56. Are you doing your work as well as you used to? ..	Yes	*No*	Self
57. Can you be easily cheered up?	Yes	*No*	Self

NOTE: Responses scored as depressed are *italicized.* Where the point biserial correlations of the Self and Observer ratings with the item response differed by 0·1 or greater, Self or Observer appears, indicating that the depressed response to that item is more likely to be associated with a 'high' rating of depression by the rater indicated.

TABLE II

Mean depression scores in diagnostic groups

Diagnosis	N	Mean	S.D.
Psychotic depressive reaction	26	14·2	3·9
Psychosis with depression ..	5	10·8	6·2
Non-depressed psychosis ..	10	10	5·2
Neurotic depressive reaction	26	12·1	4·3
Neurotic depressive reaction + anxiety	9	12·2	3·9
Neurosis with depression ..	4	10·5	5·8

endogenous' depression, Class 2 'endogenous' depression, and Class 3 non-depressive.

Sensitivity

The sensitivity of the questionnaire Depression Score to change in a patient's clinical status was examined by re-administering the questionnaire to patients shortly before discharge. The test-retest intervals ranged from 1 to 13 weeks, with a mean of 5 weeks. Changes in Depression Scores were also related to a clinical assessment of overall improvement for each subject, which was made by the senior psychiatrist in charge of the ward without knowledge of any testing results.

Relationship to concurrent measures of depression

The relationships between Depression Scores, Self-ratings of depression and Observer ratings were calculated in order to demonstrate the relevance of the Depression Score to objective and subjective assessments of depression. Immediately following completion of the questionnaire, subjects were asked to rate their depression on a 10 cm. line, the extremes of which were labelled 'I feel very depressed' and 'I do not feel depressed at all'. The use of this line was based on the method for measuring pain described by Pilowsky and Bond (1969). The observer rated the subject's mood on a similar line before administering the questionnaire. The subject did not see the observer's rating, and instructions for the Self-rating were phrased as neutrally as possible. The Observers included a psychiatrist who rated 15 subjects, a psychologist who rated 37, and two fifth year medical students who rated 8 subjects. Point biserial correlations (Ferguson, 1966) were calculated for each item with the Self and Observer depression ratings respectively.

RESULTS

(a) *Depression Scores and diagnosis*

The means and standard deviations of the Depression Scores of a number of diagnostic groups are shown on Table II. These indicate a tendency for patients diagnosed as predominantly depressed to have higher Depression Scores than those whose diagnoses did not

https://doi.org/10.1192/bjp.121.4.411 Published online by Cambridge University Press

specify depression. The group of patients diagnosed as 'psychotic depression' obtained the highest mean score.

(b) *Depression Score and taxonomic groupings*

The distributions of Depression Scores in the three taxonomic classes are shown on Table III. The two 'depression' classes have similar mean scores and are fairly clearly differentiated from the 'non-depressed' class on the basis of the scoring range.

TABLE III

Distribution of depression scores in the three taxonomic classes

Depression score	Class 1	Class 2	Class 3
1	—	—	2
2	—	—	1
3	—	—	3
4	—	—	7
5	—	—	5
6	—	—	5
7	—	—	8
8	—	—	5
9	1	—	6
10	—	1	3
11	—	10	2
12	2	4	—
13	3	4	1
14	2	2	—
15	5	7	—
16	2	7	—
17	3	8	—
18	1	3	—
19	—	2	—
20	—	1	—
N	19	51	48
Mean ..	14·53	14·47	6·48
S.D. ..	2·14	2·66	2·68

(c) *Sensitivity of Depression Score to clinical progress*

The initial distribution of Depression Scores for the female in-patients was compared with the distribution of scores at discharge (Table IV). The number of subjects retested was 54, and the difference between means of the two sets of scores was significant, $t = 6{\cdot}80$; $p < 0{\cdot}001$ (two tailed test). Only two subjects were rated by the clinician 'not improved' on discharge; both had lower Depression Scores, one less by 6 points and the other less by 10. Thirty-eight subjects were rated 'improved'. The change in mean scores was from 11·8 to 7·7; $t = 4{\cdot}48$; $p < 0{\cdot}001$ (see Table V).

TABLE IV

Change in mean depression scores for all subjects ($N = 54$)

Depression score	On admission	On discharge
Mean	12·0	7·0
S.D.	4·1	3·9

TABLE V

Change in depression scores for 'improved' group ($N = 38$)

Depression score	On admission	On discharge
Mean	11·8	7·7
S.D.	4·7	4·6

Fourteen subjects were rated 'much improved' (Table VI). The change in mean scores on the two occasions was from 12·0 to 4·7; $t = 5{\cdot}85$; $p < 0{\cdot}001$.

TABLE VI

Changes in mean depression scores for 'much improved' group ($N = 14$)

Depression score	On admission	On discharge
Mean	12·0	4·7
S.D.	4·0	2·1

The difference between the 'improved' and the 'much improved' groups at the time of retesting was not significant (Mann Whitney U Test $p = 0{\cdot}25$).

(d) *Depression Score and concurrent measures of depression*

The intercorrelations between the Depression Score, the Self rating of depression and the Observer rating of depression were as follows:

Self Rating × Depression Score
$r = 0{\cdot}59$, $n = 59$, $p < 0{\cdot}01$;
Self Rating × Observer Rating
$r = 0{\cdot}44$, $n = 59$, $p < 0{\cdot}01$;
Observer Rating × Depression Score
$r = 0{\cdot}62$, $n = 60$, $p < 0{\cdot}01$.

The Observer ratings were categorized into four groupings by dividing the 'rating line' into four parts labelled 'not depressed', 'slightly depressed', 'moderately depressed' and 'severely

https://doi.org/10.1192/bjp.121.4.411 Published online by Cambridge University Press

depressed'. The distribution of Depression Scores within the four Observer rating categories is shown in Table VII.

The correlation of 0·62 for Observer ratings and Depression Scores is for all three types of Observers' ratings pooled. It is of interest to note the correlations of Depression Scores with the individual raters. In the case of the psychiatrist's ratings, a correlation of $r = 0{\cdot}59$ ($N = 15$; $p < 0{\cdot}01$) was obtained with the questionnaire Depression Score. The psychologist's ratings correlated $r = 0{\cdot}57$ ($N = 37$; $p < 0{\cdot}01$) with the Depression Score, and the medical students' ratings correlated $r = 0{\cdot}97$ ($N = 8$; $p < 0{\cdot}01$). The distribution of Depression Scores for the three observers is shown on Table VIII. Possible explanations for these differences are discussed below.

TABLE VII
Distribution of depression scores in the observer rating categories

Dep. score	Observer rating: Not depressed	Slightly depressed	Moderately depressed	Severely depressed
1	—	—	—	—
2	2	—	—	—
3	3	1	—	—
4	2	1	1	—
5	1	1	1	2
6	1	3	2	—
7	1	1	1	—
8	1	1	1	—
9	—	3	2	—
10	1	—	2	1
11	1	2	—	—
12	—	3	—	1
13	—	2	5	2
14	—	—	1	—
15	—	—	1	1
16	—	—	1	2
17	—	—	—	—
18	—	—	1	1
19	—	—	—	—
20	—	—	—	—
N ..	13	18	19	10
Mean ..	5·2	8·7	10·6	12·3
S.D. ..	2·9	3·1	3·9	4·2

The point biserial correlations of the 57 questionnaire items with Self and Observer ratings are shown on Table IX. A number of the items were found to correlate more highly with one of the ratings than with the other. The possible influence of these items on the Depression Score is shown on Table I, where the items contributing to this score and their relationships with the ratings are shown.

Some indication of the validity of items included in the Depression Score may be inferred from the point biserial correlations with the ratings. Fifteen of the 19 items correlated significantly $p < 0{\cdot}05$ with the Self rating of depression, and 14 of the 19 items correlated significantly $p < 0{\cdot}05$ with the Observer rating of depression.

DISCUSSION

The results of the first stage of the study suggest that the Depression Score is related to clinical diagnosis of depression and that the score reflects alterations in the clinical status of the subjects. Depression Scores of the 'Improved' and the 'Much Improved' groups were not

TABLE VIII
Frequency distributions of depression scores of subjects seen by observers

Psychiatrist																				
Number of subjects	—	—	1	—	1	—	1	2	—	2	2	2	2	—	—	1	—	1	—	—
																				N=15
Depression score	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Psychologist																				
Number of subjects	—	—	3	3	4	6	2	1	3	2	1	2	6	1	1	2	—	—	—	—
																				N=37
Depression score	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Medical students																				
Number of subjects	—	2	—	1	—	—	—	—	2	—	—	—	1	—	1	—	—	1	—	—
																				N=8
Depression score	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

https://doi.org/10.1192/bjp.121.4.411 Published online by Cambridge University Press

TABLE IX

Point biserial correlations with (Yes) response
(Self ratings and observer ratings correlated with individual items)

Questionnaire items	Self rated dep. r_{pbi}	Observer rated dep. r_{pbi}
1. Are you more irritable towards other people?	0·30†	0·26*
2. Have you lost interest in watching television?	0·32†	0·39‡
3. Do you have difficulty in falling asleep without tablets?	−·02	0·23
4. Do you feel depressed all day long?	0·28*	0·25*
5. Do you feel slowed up in your thinking?	0·31†	0·44§
6. Have you any serious money worries?	0·01	−·08
7. Have you had any recent family worries?	0·01	0·13
8. Have you lost someone you love in the past year?	0·07	0·03
9. Do you feel you are a bad person?	0·12	−·03
10. Have you moved house in the past year?	−·25	−·24
11. Do you avoid company	0·56§	0·40‡
12. Is it more difficult to concentrate on your work?	0·45§	0·22
13. Have you any housing worries?	0·24	0·14
14. Do you wish you were able to cry?	0·14	0·13
15. Do you have a restless and and disturbed sleep without tablets?	0·22	0·29*
16. Do you feel most depressed in the evenings?	0·02	0·09
17. Are there times when you do not feel depressed?	−·21	−·22
18. Do you have less interest in reading newspapers?	0·33‡	0·18
19. Do you think you will get better?	−·24	−·01
20. Do you feel that people are sometimes talking about you?	0·14	0·02
21. Is it easy to fall asleep without tablets?	0·03	−·15
22. Is your appetite normal?	−·23	−·37‡
23. Have you less interest in sex?	0·11	0·24
24. Do you feel you are a burden to others?	0·26*	0·30†
25. Is life worth living?	−·41‡	−·17
26. Do you cry a lot?	0·32†	0·35‡
27. Are you unable to cry?	0·04	−·08
28. Have you become constipated?	0·03	0·04
29. Do you feel happier in the mornings?	−·22	−·02
30. Do you suffer from a dry mouth?	0·30†	0·30†
31. Have you less feeling for those close to you?	0·05	0·14
32. Do you feel you are letting other people down?	0·29*	0·25*
33. Have you lost your appetite?	−·09	0·29*
34. Have you had trouble at work in the past year?	0·25	0·12
35. Do you wish you were dead?	0·27*	0·34‡
36. Do you waken much earlier than your usual time without tablets?	0·04	0·22
37. Are you as good a person as most of your friends?	−·28*	0·08
38. Do you feel less depressed when you are with company?	−·03	0·12
39. Do you think your illness is a punishment that you deserve?	−·21	−·07
40. Do you have less interest in things you usually enjoy?	0·39‡	0·40‡
41. Can you sleep normally without tablets?	−·06	−·27*
42. Do you waken at your usual time without tablets?	−·11	−·00
43. Do you think there is something seriously wrong with your body?	−·04	0·12
44. Is your depression the same all day long?	0·30†	0·15
45. Do your find difficulty in relaxing?	0·28*	0·23
46. Do you feel life is not worth living?	0·46§	0·28*
47. Have you lost weight?	0·02	0·27*
48. Do you feel most depressed in the mornings?	0·19	0·19
49. Have you overheard people talking about you?	0·01	0·09
50. Do you feel this illness has been brought upon you by yourself?	0·12	0·12
51. Do you feel slowed up in doing things?	0·24	0·46§
52. Does the future look hopeful?	−·42§	−·23
53. Do you feel happier in the evenings?	0·01	−·17
54. Have you thought recently about ending your life?	0·14	0·26*
55. Do you feel time passing more slowly?	0·41‡	0·33‡
56. Are you doing your work as well as you used to?	−·45§	−·32†
57. Can you be easily cheered up?	−·52§	−·25*

Level of significance of correlation:
* = $p < 0·05$.
† = $p < 0·02$.
‡ = $p < 0·01$.
§ = $p < 0·001$.

https://doi.org/10.1192/bjp.121.4.411 Published online by Cambridge University Press

significantly different, but the criteria for improvement were not restricted to the dimension of depression, so that this lack of a clear difference in depression level is not unaccountable. The score changes of the two subjects rated as 'Not Improved' appear incongruous, but may perhaps be explained in a similar manner. The gross changes in Depression Scores from admission to discharge suggest that the questionnaire is reasonably sensitive to changes in the status of individual patients.

The results of the second stage of the investigation suggest that the Depression Score is related to both the patients' concepts of depression and the concepts of depression held by trained observers. The possible effects of the questionnaire on the ratings and vice versa should be considered in interpretation of the findings; however, the testing procedure was designed to minimize any direct distortion of this nature. Of special interest is the finding that both the observers' and subjects' ratings were more highly correlated with the Depression Score than they were with each other. From this it may be inferred that there is a difference in the two concepts of depression, and it is suggested that perhaps the difference is that of symptom versus syndrome. In support of this hypothesis is the finding that the correlation of Self and Observer ratings is far larger for the less specialized observers (the medical students $r = 0{\cdot}91$) than are those obtained by the psychiatrist and psychologist ($r = 0{\cdot}36$ and $r = 0{\cdot}40$ respectively). If it may be assumed that the students' concept of depression would be less influenced by theoretical formulations of syndromes, this may explain why their ratings should be closer to those of the subjects. Another possibility is that their subjects tended to be at the extremes of the depression—non-depression spectrum, thus making ratings somewhat easier to carry out (Table VIII).

A number of the questionnaire items in the Depression Score also correlate more highly with one rating than with the other; thus indicating the areas in which the differences between the two concepts of depression are most marked.

In view of the limitations inherent in rating scales where global concepts such as 'depression' are used, it would seem that the obtained correlations are sufficiently large to justify using the questionnaire as an index of degree of depression in future research projects.

Summary

A depression scale has been constructed using items in a depression questionnaire originally constructed for use in taxonomic studies. Evidence is provided for the sensitivity of the depression score to clinical change and for its validity as a measure of depression. The development of the scale thus increases the usefulness of the questionnaire, which may already be used as a means of classifying patients on the basis of the depressive component of their illness.

Acknowledgements

We wish to acknowledge the help and encouragement given to us during the course of this study by Dr. H. B. Prior, Medical Superintendent, and Dr. T. Pawloff, Psychiatrist, of the Broughton Hall Psychiatric Clinic.

References

Ferguson, G. A. (1966). *Statistical Analysis in Psychology and Education.* 2nd edition. McGraw-Hill.

Pilowsky, I., Levine, S., and Boulton, D. M. (1969). 'The classification of depression by numerical taxonomy.' *Brit. J. Psychiat.*, **115**, 937–45.

—— and Boulton, D. M. (1970). 'Development of a questionnaire-based decision rule for classifying depressed patients.' *Brit. J. Psychiat.*, **116**, 647–50.

—— and McGrath, M. D. (1970). 'Effect of ECT on responses to a depression questionnaire: implications for taxonomy.' *Brit. J. Psychiat.*, **117**, 685–8.

——and Bond, M. R. (1969). 'Pain and its management in malignant disease.' *Psychosom. Med.*, **31**, 400–4.

A synopsis of this paper was published in the July 1972 *Journal.*

I. Pilowsky, M.D., F.A.N.Z.C.P., M.R.C.Psych., *Associate Professor, Department of Psychiatry, University of Sydney, Australia. Present address: Professor of Psychiatry, University of Adelaide, Royal Adelaide Hospital, Adelaide, South Australia, 5001*

D. Spalding, B.A., *Research Assistant, Department of Psychiatry, University of Sydney*

(*Received 13 August 1971*)

https://doi.org/10.1192/bjp.121.4.411 Published online by Cambridge University Press