Brit. J. Psychiat. (1977), 130, 186–93

Group Exposure for Agoraphobics: A Replication Study

By J. D. TEASDALE, P. A. WALSH, M. LANCASHIRE and A. M. MATHEWS

Summary. A replication study was conducted of the treatment of agoraphobics by exposure in cohesive groups, as described by Hand, Lamontagne and Marks (1974). The continuing improvement during follow-up, with consequent large overall improvement, reported in the original study was not replicated. However, the present study confirmed the usefulness of this procedure as a highly cost-efficient treatment.

The assumed equivalence of the Gelder and Marks (1966) phobic rating scale and its modification by Watson and Marks (1971) was examined. Large discrepancies between the scales were obtained for initial assessments and change scores. It is suggested that there is a need for workers in this field to agree on methods of measurement.

In 1974 Hand, Lamontagne and Marks reported the effects of treatment of agoraphobics by exposure in cohesive groups. Their results demonstrated considerable clinical improvement at the cost of relatively little therapist time (approximately six therapist-hours per patient). To obtain some indication of the effectiveness of their procedure relative to other exposure treatments for agoraphobia the present authors compared their results with those of published studies which had used similar measuring instruments; all available studies which had used Watson and Marks' (1971) modification of Gelder and Marks' (1966) phobic rating scales, or the original Gelder and Marks' (1966) scales, were examined. In making this comparison we assumed, as have others (e.g. Watson and Marks, 1971), that the original Marks and Gelder 1–5 and the Watson and Marks 0–8 scale were roughly comparable, if the appropriate arithmetical transformation was made. The comparison showed that the cohesive group exposure treatment produced better overall improvement than that reported in any previous study. This superiority was largely the result of a continuing improvement during the follow-up period; in this respect cohesive group exposure treatment was unusual, for other studies found either effectively no change or actual deterioration during the follow-up period. It should, however, be pointed out that Hand *et al*'s follow-up data were incomplete, results being unavailable from one group.

Cohesive group exposure treatment appeared sufficiently promising to merit replication of Hand *et al*'s study. Of particular interest was the question whether the continuing improvement during follow-up could be replicated. An additional aim was to examine the assumed equivalence of the original Gelder and Marks (1966) rating scale and its modification by Watson and Marks (1971).

Method

Design

A simple design without control groups was considered sufficient for the purpose of the study. The design was the same as that used by Hand *et al*, with two exceptions. First, no unstructured treatment condition was employed; second, while Hand *et al* treated three separate structured (cohesive) groups, each containing four or five patients, and reported complete follow-up results on seven patients

https://doi.org/10.1192/bjp.130.2.186 Published online by Cambridge University Press

from two groups, the present study used three separate structured (cohesive) groups, two containing five and one containing eight patients, and follow-up data were available from all groups.

The design is described in detail in Hand *et al* (1974). Briefly, patients received three five-hour treatment sessions on separate days of a single week. Subsequently they met for discussions three days, four weeks, twelve weeks and twenty-four weeks after the last day of intensive treatment. Assessments were made three days before and three days, four weeks, twelve weeks and twenty-four weeks after the period of intensive treatment.

Patients

Characteristics of the patients treated were as follows, comparable data from the combined structured and unstructured groups of the original study being shown in brackets. Eighty-nine per cent of the eighteen patients were female (64 per cent of twenty-five patients). Mean age was 40 years, range 21–53 (mean age 35, range 17–56), and mean phobic duration 7·4 years, range 1–18 (mean duration 8·5 years, range 1·5–26). Eighty-three per cent (84 per cent) had received previous treatment for their phobias, including drugs and behavioural treatments. Thirty-nine per cent (36 per cent) had been treated previously for other psychiatric problems. Thirty-nine per cent (44 per cent) were working before treatment. Thus, with the possible exception of the sex ratio, where the difference was not significant ($P > \cdot 10$), patients in the original and replication studies were very similar on these criteria.

Therapists

A male psychiatrically-trained research nurse and a female clinical psychologist together treated the first two groups. Both had considerable previous experience in individual behavioural treatment of agoraphobia, and some experience of behavioural treatment of agoraphobics in groups. Two female clinical psychologists, one trained, one in training, treated the third group. Both had previous experience in individual behavioural treatment of agoraphobia.

Assessments

Each patient and an assessor rated phobic anxiety and avoidance on the 0–8 scales of Watson and Marks (1971). A separate assessor independently rated phobic severity on the 1–5 scale of Gelder and Marks (1966). These two assessors were different for each group, i.e. no assessor rated on a given scale for more than one group.

In order to replicate the original procedure, one assessor also administered a behavioural test, as described in Hand *et al* (1974). Hand *et al* had found this an unsatisfactory measure, so results from it will not be reported. At assessments following treatment, patients rated their overall improvement. Forms for all these measures were the same as used by Hand *et al*, and, like other written material employed, were very kindly provided by Dr Isaac Marks.

During and after treatment sessions, patients completed the measures of social cohesion used in the original study.

Patients kept detailed diaries of trips out of the house. No comparable measures were reported by Hand *et al*, so results from these will not be presented.

Treatments

Structured group exposure treatment was conducted as described in the report by Hand *et al*, to which the reader is referred for details. Two groups were conducted from a base in Oxford, one from a base in Aylesbury. Exposure situations included the shopping centres of Oxford, Aylesbury and Reading, and also train trips to London, where patients travelled on the tube. Therapist descriptions of treatment sessions were similar to those recorded by Hand *et al*, e.g. the surprising ease with which patients entered feared situations, patients proposing more difficult tasks than those suggested, etc.

Results

One patient failed to attend the twelve-week follow-up. Another patient was in hospital for the treatment of a physical condition at the four-week follow-up; she also received additional treatment for her phobia after the twelve-week assessment, so that her twenty-four-week assessment was disregarded. Data from

https://doi.org/10.1192/bjp.130.2.186 Published online by Cambridge University Press

the immediately preceding assessments were used as estimates of these missing data in the calculation of group means.

Hand *et al* reported their results in a form that does not allow detailed statistical comparison with the results of the present study. Comparison of means is all that is possible with the data available.

Social cohesion

Hand *et al* suggested that the higher social cohesion of their structured groups was an important factor leading to the superior performance of these groups compared with their unstructured group treatment.

Table I compares the means on measures of social cohesion for the combined groups of the present study with those reported for structured and unstructured groups by Hand *et al*. It can be seen that on all five measures means from the present study were intermediate between Hand *et al*'s structured and unstructured groups.

Change in agoraphobia

(a) *Watson and Marks (1971) rating scale*

Fig 1 shows changes in patients' averaged ratings of phobic anxiety and avoidance. Data from Hand *et al* (1974) are included for comparison. The immediate effects of treatment in the present study were similar to those in Hand *et al*'s structured and unstructured groups. However, results from the present study failed to replicate the continuing improvement during follow-up shown by Hand *et al*'s structured groups; in fact, the results are very similar to

TABLE I
Mean social cohesion ratings

	Help from whole group	Help from individual patients	Help to individual patients	Perceived liking from other patients	Liking of other patients
Present study	5·72	1·28	0·86	2·76	4·58
Hand *et al* (1974) unstructured ..	4·0	0·6	0	1·9	3·9
Hand *et al* (1974) structured ..	6·2	3·0	1·0	3·4	5·2

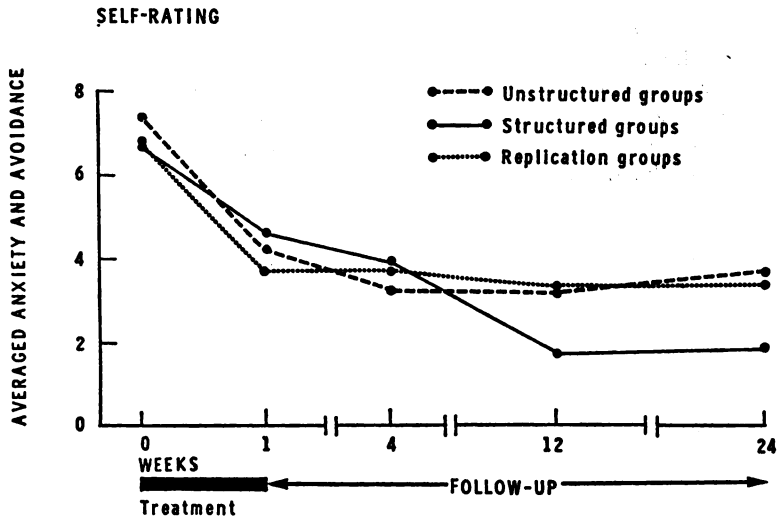


FIG 1.—Watson and Marks scales: averaged Self Ratings of Anxiety and Avoidance.

https://doi.org/10.1192/bjp.130.2.186 Published online by Cambridge University Press

those of Hand *et al*'s unstructured groups. As a result of the lack of further improvement during follow-up, the present study obtained smaller overall improvement than Hand *et al*'s structured (cohesive) groups (3·3 scale points compared with 5·1 scale points).

A similar picture emerges from assessors' averaged ratings of phobic anxiety and avoidance, shown in Fig 2.

(b) *Gelder and Marks (1966) rating scale*

Table III shows assessors' ratings for main phobia on the Gelder and Marks (1966) 1–5 scale. Most of the improvement occurs as an immediate response to treatment, with a further slight improvement from post-treatment to four-week follow-up. There is no further improvement during the follow-up period.

Comparison of Gelder and Marks (1966) and Watson and Marks (1971) rating scales

Table II shows the correlations between scores on Gelder and Marks' (1966) assessor ratings of main phobia and scores on Watson and Marks' (1971) averaged anxiety and avoidance ratings by patients and assessors. Different assessors independently made ratings on the Gelder and Marks and Watson and Marks scales. Correlations were computed for the 12-week and 24-week assessments, as variance on the ratings was greatest on these occasions.

Except for patient ratings of change pre-treatment to post-treatment, significant and substantial correlations were obtained between assessor ratings of main phobia on the Gelder and Marks (1966) scale, and patient and assessor averaged anxiety and avoidance ratings on the Watson and Marks (1971) scales. Generally, these were less than the correlations between patient and assessor ratings on the Watson and Marks (1971) scales obtained in the present study, but were comparable to previously reported reliabilities for each scale (Watson and Marks, 1971; Gelder and Marks, 1966). Thus, as far as correlations were concerned, there was reasonable agreement between the two scales.

Table III shows data comparing mean values on the two scales. The comparison was made by transforming Watson and Marks (1971) assessor ratings, averaged for anxiety and avoidance, from a 0–8 scale to a 1–5 scale (by dividing by two and adding one). These Watson and Marks ratings, expressed on a 1–5 scale, were then compared with the Gelder and Marks assessor ratings of main phobia actually obtained.

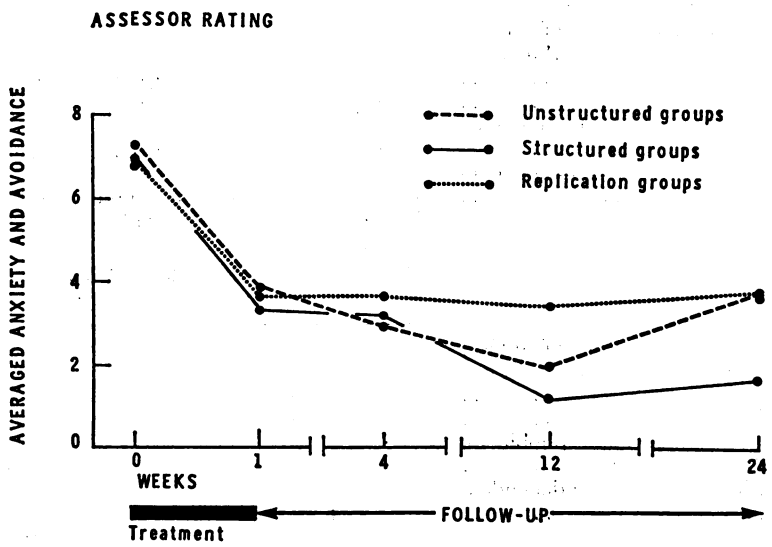


FIG 2.—Watson and Marks scales: averaged Assessor Ratings of Anxiety and Avoidance.

https://doi.org/10.1192/bjp.130.2.186 Published online by Cambridge University Press

TABLE II

Correlations between Watson and Marks (1971) 0–8 and Gelder and Marks (1966) 1–5 rating scales for phobias (n = 18 for all correlations)

	12-week assessment			24-week assessment		
	0–8 (assessor)	0–8 (patient)	1–5 (main phobia)	0–8 (assessor)	0–8 (patient)	1–5 (main phobia)
0–8 (assessor) ..	—	·95**	·76**	—	·96**	·76**
0–8 (patient) ..		—	·80**		—	·74**
1–5 (main phobia)			—			—
	Change pre-treatment to post-treatment			Change pre-treatment to 24-week follow-up		
	0–8 (assessor)	0–8 (patient)	1–5 (main phobia)	0–8 (assessor)	0–8 (patient)	1–5 (main phobia)
0–8 (assessor) ..	—	·65*	·76**	—	·92**	·85**
0–8 (patient) ..		—	·42		—	·80**
1–5 (main phobia)			—			—

** $P < ·001$; * $P < ·01$.

TABLE III

Comparison of means from Watson and Marks (1971) 0–8 and Gelder and Marks (1966) 1–5 rating scales for phobias (n = 18 for all means)

		Pre-treatment	Post-treatment	4-week follow-up	12-week follow-up	24-week follow-up
I	Mean assessor averaged rating of anxiety and avoidance on Watson and Marks 0–8 scale	6·94	3·77	3·64	3·43	3·64
II	I transformed to 1–5 scale	4·47	2·89	2·82	2·72	2·82
III	Mean assessor rating of main phobia on Gelder and Marks 1–5 scale	3·64	2·83	2·60	2·60	2·58
IV	II compared with III by t-test for dependent data	$t = 8·42$ $P < ·001$	$t < 1$ NS	$t = 1·26$ NS	$t < 1$ NS	$t = 1·31$ NS
		Change pre-treatment to post-treatment		Change pre-treatment to 24-week follow-up		
I	Mean assessor averaged rating of anxiety and avoidance on Watson and Marks 0–8 scale	3·17		3·30		
II	I transformed to 1–5 scale	1·59		1·65		
III	Mean assessor rating of main phobia on Gelder and Marks 1–5 scale	0·81		1·06		
IV	II compared with III by t-test for dependent data	$t = 5·34$ $P < ·001$		$t = 4·01$ $P < ·001$		

https://doi.org/10.1192/bjp.130.2.186 Published online by Cambridge University Press

With the exception of the pre-treatment measure, transformed Watson and Marks ratings were quite similar to Gelder and Marks ratings, and did not differ significantly from them. However, there was a substantial and statistically significant difference on the pre-treatment measure. This probably also accounted for the large and statistically significant differences on the two change scores; the immediate change from pre-treatment to post-treatment on the Gelder and Marks scale was approximately half that on the transformed Watson and Marks scale.

Discussion

Replication of the results of Hand, Lamontagne and Marks (1974)

The superiority of Hand *et al*'s cohesive group treatment to previously reported methods depended on continuing improvement during the follow-up period. This makes it possible to draw useful conclusions from a replication study without comparison groups, such as the present, because *within*-group changes are an important source of evidence. A definitive demonstration of the superiority of this method over others would require a large-scale trial in which it was compared with leading alternatives. As this would be extremely costly, a much cheaper, simple study, such as the present, is a useful first stage; if the pattern of change observed *within* group cannot be replicated, it may not be worthwhile to conduct the larger *between*-group study.

The continuing improvement during follow-up observed by Hand *et al* was not replicated in the present study. Consequently, overall improvement was also less than that reported in the original study, and not apparently superior to that produced by other treatments. While this is disappointing, it should be stressed that the present results do confirm group exposure as an extremely cost-efficient psychological treatment for agoraphobia, and to that extent are encouraging.

Why were Hand *et al*'s results not replicated? One possibility is that in the original study, by chance, the structured group which was omitted from the calculation of follow-up means contained more severe agoraphobics than the other two structured groups. This would produce an apparent improvement in mean score for the structured groups during follow-up, for artifactual reasons. Data are not available to evaluate this possibility.

Another possibility might be differences in patient populations between the original and replication studies. However, in terms of the simple data available the two groups of patients were actually quite similar.

Hand *et al* suggested that the higher social cohesion of their structured groups was an important factor in producing the continuing improvement during follow-up, compared with their unstructured groups. The present study obtained social cohesion ratings intermediate between Hand *et al*'s structured and unstructured groups. It is possible that there is a certain minimum level of social cohesion necessary to produce further improvement during follow-up, and that groups in the present study, while higher in social cohesion than Hand's unstructured groups, were still below this minimum. If the level of social cohesion attained in Hand *et al*'s structured groups was not replicated, it would suggest that the treatment procedure they described, which was followed exactly, does not reliably produce the required level of social cohesion.

It is not immediately apparent why high social cohesion should lead to continuing improvement during follow-up. Hand *et al* suggested that it might operate by increasing the motivation for continued exposure to phobic situations, this further exposure then leading to further improvement. Hand (personal communication) has reported that during follow-up meetings cohesive groups worked out loose contracts (home work assignments) among individual members, and discussed intensively their achievements and failures since the last meeting. He suggests that such contracts, in the context of high social cohesion, motivated sufficient further practice to produce continuing improvement. Follow-up meetings were conducted in a similar way in the present study. It may be that, unless more tightly structured specifically to encourage continuing practice, such meetings may not reliably pro-

https://doi.org/10.1192/bjp.130.2.186 Published online by Cambridge University Press

duce this result. Hafner and Marks (1976) employed groups described as 'moderately cohesive' in group exposure treatment of agoraphobia. The data they present graphically suggest that they too failed to obtain continuing improvement during follow-up, and consequently achieved improvements smaller than those reported by Hand *et al.* However, it is possible that this was the result of floor effects on the measures used.

One of the striking features to emerge from our initial examination of published studies of psychological treatment of agoraphobia (p 186) was the virtual absence of continuing improvement once intensive treatment had stopped. The development of more effective treatments will probably depend on the design of procedures which will encourage the patient to continue therapeutic exposure by himself during the follow-up period. A recent study (Mathews, Teasdale, Johnston and Munby, unpublished) showed that it was possible to obtain continuing improvement during follow-up by the use of a home-based individual treatment programme specifically designed to achieve this effect. A modification of Hand *et al*'s cohesive group exposure treatment to produce continuing exposure to phobic situations more reliably would probably represent an extremely effective and efficient treatment.

Comparison of the rating scales of Gelder and Marks (1966) and Watson and Marks (1971)

Results from these two scales were broadly similar, in terms of correlations and for means on assessments made after treatment (using the appropriate transformations). However, there were serious discrepancies in the means of pre-treatment assessments, and in mean changes as a result of treatment. The extent of these discrepancies is illustrated by the fact that the immediate change as a result of treatment on the Watson and Marks scale (after the appropriate transformation) is almost twice that on the Gelder and Marks scale. As a different combination of assessors used the scales for each of the three groups, the discrepancies appear to be a function of the nature of the scales rather than of the eccentricities of a particular assessor in using one of them.

Examination of the scales suggests a cause for the discrepancies. In the present study, as in previous studies conducted in Oxford (Gelder, Bancroft, Gath, Johnston, Mathews and Shaw, 1973; Mathews, Johnston, Lancashire, Munby, Shaw and Gelder, in press) the Gelder and Marks scale was used to give a fairly global estimate of the extent of agoraphobia. This actually represents a slight modification of the original 'main phobia' rating on these scales; whereas in the original Gelder and Marks scale separate ratings were made of 'main phobia' and 'other phobias', in the form used by the Oxford group, a single rating of 'phobic severity' was made on this scale. Using the scale in this way, a maximum score of 5 could only be obtained by someone who was almost totally housebound by agoraphobia.

In contrast, in the present study, as in Hand *et al* (1974), the Watson and Marks scales were used to give estimates of anxiety and avoidance for five fairly specific phobic situations, e.g. travelling on buses with automatic doors. A maximum score of 8 on averaged anxiety and avoidance ratings would be obtained by someone who invariably avoided and was terrified by each of the five specific phobic situations chosen. It is therefore possible, by the assessor selecting suitable situations, for someone with only moderately severe agoraphobia to obtain maximum score on this scale. This appears to have actually happened in Hand *et al*'s study and in the present study, and would account for the discrepancies in pre-treatment assessments and change scores observed between the Gelder and Marks and Watson and Marks scales.

The assumption of broad equivalence between these two scales, which was made in our initial examination of published studies (p 186), and has been made by other workers, is no longer tenable. Currently, research on agoraphobia conducted at the Maudsley Hospital by Marks and his colleagues, and in Holland by Emmelkamp and his colleagues employs the Watson and Marks scales, whereas research conducted by the Oxford group uses a slightly modified form of the Gelder and Marks scales. While there are difficulties in comparing the results of studies using the same measuring instruments,

https://doi.org/10.1192/bjp.130.2.186 Published online by Cambridge University Press

these are far less than the difficulties of comparing studies using different measuring instruments. Accordingly, some uniformity in the measures of agoraphobia used by different workers would be useful.

Conclusions

1. The present study failed to replicate the unusually large overall improvement, and the continuing improvement during follow-up, reported by Hand *et al* (1974), for treatment by exposure in cohesive groups.

2. The present study confirmed that Hand *et al*'s cohesive group treatment is—nonetheless—one of the most cost-efficient psychological treatments of agoraphobia.

3. Further effort should be devoted to programming exposure during follow-up as an integral part of treatment.

4. The phobic rating scales of Gelder and Marks (1966) and their modification by Watson and Marks (1971) cannot be assumed to be equivalent.

5. Agreement on a common method of measurement by researchers in agoraphobia would be valuable.

Acknowledgements

This research was supported by the Medical Research Council of the United Kingdom.

Many thanks are due to Dr M. Johnston and Miss V. Rezin, who were therapists, to Dr P. M. Shaw who screened and assessed patients, to Mrs M. Munby, who was an assessor, and to staff at the Manor House Hospital, Aylesbury, who provided the base for the Aylesbury group.

References

Gelder, M. G., Bancroft, J. H. J., Gath, D. H., Johnston, D. W., Mathews, A. M. & Shaw, P. M. (1973) Specific and non-specific factors in behaviour therapy. *British Journal of Psychiatry*, **123**, 445–62.

—— & Marks, I. M. (1966) Severe agoraphobia: a controlled prospective trial of behaviour therapy. *British Journal of Psychiatry*, **112**, 309–19.

Hafner, J. & Marks, I. M. (1976) Exposure *in vivo* of agoraphobics: contributions of diazepam, group exposure and anxiety evocation. *Psychological Medicine*, **6**, 71–88.

Hand, I., Lamontagne, Y. & Marks, I. M. (1974) Group exposure (flooding) *in vivo* for agoraphobics. *British Journal of Psychiatry*, **124**, 588–602.

Mathews, A. M., Johnston, D. W., Lancashire, M., Munby, M., Shaw, P. M. & Gelder, M. G. (1976) Imaginal flooding and exposure to real phobic situations: treatment outcome with agoraphobic patients. *British Journal of Psychiatry*, **129**, 372–7.

Watson, J. P. & Marks, I. M. (1971) Relevant and irrelevant fear in flooding—a crossover study of phobic patients. *Behaviour Therapy*, **2**, 275–93.

J. D. Teasdale, M.A., Ph.D., Dip.Psych., *Senior Research Psychologist, Department of Psychiatry, University of Oxford, The Warneford Hospital, Oxford OX3 7JX*

P. A. Walsh, R.M.N., *Senior Nursing Officer (Special Projects), Berkshire AHA—East District*

M. Lancashire, B.Sc., M.Sc., *Senior Psychologist, The London Hospital*

A. M. Mathews, B.Sc., Ph.D., Dip.Psych., *Senior Research Psychologist, Department of Psychiatry, University of Oxford, The Warneford Hospital, Oxford OX3 7JX*

(*Received 23 April; revised 15 June 1976*)

https://doi.org/10.1192/bjp.130.2.186 Published online by Cambridge University Press