

Experimental Evaluation of Sensitivity Training¹

CAROLE E. BARE

RIE R. MITCHELL

This study was undertaken to: (1) determine whether significant behavioral changes occur as a result of sensitivity training, and (2) compare the effectiveness of three sensitivity training group time patterns: spaced (an average of 30 hours spaced over a period of approximately 10 weeks), massed (an average of 24 hours in one continuous training session), and a combination (a weekend, with 10 hours spent in training, followed by five two-hour sessions once a week and terminating with another weekend).

Behavioral changes were measured by individually constructed rating scales. The criterion scores included self, trainer, and group member ratings, as well as an independent assessment by outsiders who knew the individual well.

Experimental subjects showed more rated behavioral change ($p < .01$) than did control group subjects at the conclusion of training and after a three-month follow-up period. The data also indicate at the short post that the combination pattern of time was more effective than the spaced pattern, but no inter-pattern differences were significant at the three-month follow-up.

Some major questions which should be answered by a comprehensive evaluation of sensitivity training are: Are training goals reached; i.e., do experimental groups differ significantly from

1. Manuscript received June 1, 1970.

Carole E. Bare, formerly associate professor of clinical psychology at Long Island University, is presently a clinical psychologist in private practice in Los Angeles. Rie R. Mitchell is assistant professor of psychological foundations, San Fernando Valley State College, Northridge, California; and is employed part time as a psychologist by the Glendale Guidance Clinic and Behavior Science Foundation.

comparable control groups on an outcome such as behavioral change? Are gains brought about by the training maintained, as shown by follow-up evaluations? Which training time pattern is most effective in achieving training goals?

This investigation had two purposes: (1) to determine whether significant behavior changes occurred as a result of sensitivity training when experimental subjects are compared with control subjects who did not receive sensitivity training; and (2) to compare the effectiveness of three sensitivity training group time patterns—spaced, massed, and combination.

Three major patterns of sensitivity training that have evolved in recent years were studied in the present investigation. A group that meets for a total of 20 to 50 hours, spaced over a period of about eight to 15 weeks, is designated a *spaced group*. A second time pattern is the marathon group, designated as a *massed group*, which generally meets for an entire weekend, with approximately 24 to 30 hours spent in one continuous sensitivity training session. A combination of the former two patterns constitutes the third. The *combination groups* included in this study met, first, for a weekend, during which approximately 10 hours were spent in sensitivity training, followed by five two-hour sessions once a week, and a weekend similar to the first one.

All three sensitivity training time patterns are purported to effect positive behavioral change. They offer, however, quite distinct differences in rationale and approach. Massed sensitivity training groups are rather intensive in approach. "Constructive aggression" and deliberately instigated group pressure focused on behavioral change are supposed to enhance "transparency" and a genuine encounter among participants. One- or two-hour sessions are considered insufficient time for an individual to remove substantially his social façade or for the creation of the group pressure necessary for significant behavioral change. Bach (1966) stated that: ". . . becoming transparent, leveling, exposing to influence-pressure, attempting changes and practicing new behavior, we believe, is a 'natural Gestalt,' i.e., a unit of learning experience which should not be broken up into bits and pieces but should occur as a whole,

mediating a significant turn-point, a big step toward becoming what one can be!" (p. 996).

Even the exhaustion and fatigue produced by the massed session is said to be of value: because tired people do not have the energy to play games it can lead to a more truthful expression of feelings. On the other hand, some critics claim that the participants are also too tired to practice new behavior and that, because of their weakened state during the long group meeting, massed group training may have little effect. Advocates of massed group training point out, however, that new learnings can occur more readily and past learnings and practices can be counteracted by taking the individual out of his old environment for an extended time. Nevertheless, even if behavioral changes come about in the massed group environment, it does not necessarily follow that such changes will be translated into the participant's "real" environment. The group member may not sustain the one-trial behavioral changes if there are no follow-up sensitivity training meetings to reinforce the new behavior in the old environment. In fact, some critics charge that a massed group may open up new problems for some participants. When such a participant returns to the "real" environment, he may not be able to deal with these added problems without the support of the training group; he may therefore feel worse and function less well than before training.

The most significant assumption in the rationale for spaced sensitivity training is that behavioral changes require a substantial period of time. Gradual insights in spaced sessions are said to allow an individual time to practice and integrate the suggested new behaviors, both during the sensitivity training sessions and in the "incubation" periods in the "real" environment between sessions. Family, friends, and business associates may accommodate an individual's gradually changing behavior more readily than a change that is relatively sudden. It may therefore be easier to transfer to the actual environment behaviors learned and practiced in spaced groups rather than behaviors practiced in a massed group.

The incubation period can, however, also work against positive behavioral changes if the intervening time is used to

rationalize the events of the sessions and to solidify old façades and behaviors. Critics of the spaced group pattern believe that the intensive and dramatic character of the massed group experience is more difficult to forget and to rationalize than the spaced group experience and that members of the latter are probably less involved because they are not taken out of their old environment for as long as the members of the former group.

Combination sensitivity training groups incorporate many of the strengths, but also the weaknesses, of the massed and spaced schedules. In general, a combination training group is characterized by a less intensive and aggressive approach than the massed training group. Nevertheless, because of the weekend sessions at the beginning and the end of the spaced sessions, it is still more intensive than the spaced groups. Furthermore, since the weekend sessions are generally held away from the home environment, it is felt that previous habits are less likely to interfere.

METHOD- OLOGICAL ISSUES

There are many methodological problems encountered in the evaluation of sensitivity training groups, but one of the most basic issues centers around the development of the criterion instrument used to measure precisely specified outcomes of training. In this study, outcome criteria are defined in terms of observable behaviors, rated not only by the subject, but also by the group trainer, fellow group members, and individuals closely associated with the subject outside the group.

A second problem involves individual differences in the need for particular behavioral changes. To assess whether an individual reached an appropriate goal, one which helped him function more fully, individualized behavioral change prescriptions were constructed from a known baseline of functioning for each participant.

It is also difficult to obtain control groups of similar motivation for behavioral change, to minimize data attrition, and to obtain a sufficiently large sample size. In this study, control group members were equivalent to the experimental group members in age, sex, marital status, occupation, and personality test scores. To ensure comparable motivation, only individuals who (like the experimental subjects) were not involved in sensitivity

training or any type of therapy at the time of the study but wished to participate in sensitivity training, were accepted into the control group. There was little data attrition (97 per cent of the total possible data were obtained at the termination of the groups and 95 per cent at the three-month follow-up study). A large sample size ($N = 165$) was used. It included heterogeneous subjects (in the sense that the subjects were not employees of the same organization), so that presumed outcomes of sensitivity training could be separated from organizational context factors.

The trainer sample is frequently heterogeneous with respect to the goals of training, which may reduce the value of the findings. In this study, trainers who subscribed to the overall goal of promoting long-term positive behavioral change were chosen.

HYPOTHESES

The following three hypotheses were investigated:

1. Significantly more positive behavioral change is exhibited by participants in one of the three experimental conditions—spaced, massed, and combination—than by control group subjects with no sensitivity training, when measured on the basis of individualized and independently rated behavioral change criteria, both immediately after the termination of the sensitivity training groups and after a period of three months.
2. Participation in the combination sensitivity training groups produces significantly more positive behavioral change than participation in spaced or massed groups, but participation in spaced groups produces significantly more positive behavioral change than massed group participation, when measured by the above criteria.
3. When the behavioral change criterion scores at the termination of the training sessions are compared with the behavioral change criterion scores three months after training, subjects in the massed condition show a significant decrease in behavioral changes gained during training, but there is no significant increase or decrease in the behavioral change criterion scores for subjects in the spaced, combination, and control conditions.

PROCEDURES

The sample consisted of 165 adults in the greater Los Angeles

area (mean age 32.2; 79 males and 86 females). Individuals who had undergone individual or group therapy or sensitivity training within the previous two years were excluded because delayed effects of such therapy could contaminate the study results.² The 165 subjects were distributed among four groups: the control group ($N = 42$) and three types of experimental groups—spaced, massed and combination. The 45 subjects who received spaced training and the 42 subjects in the combination training pattern were in each case obtained from four sensitivity training groups. The 36 subjects who received massed training were obtained from seven different training groups. The sizable number of excluded subjects in the massed group probably both promoted and hindered the progress of the subjects in this study; it is therefore probable that a random influence occurred.

The comparability of the four types of groups with respect to age, sex, marital status, occupation, and personality test scores was analyzed statistically. The results were satisfactory.

The mean number of training hours was comparable for the three types of groups: 30 hours over a period of 13-15 weeks for the spaced groups; 24 hours for the massed groups; and 28 hours for the combination groups.

The ten trainers selected for the study all had previous experience in leading groups. Eight had doctorates and were either university professors or in private practice. The remaining two trainers were doctoral candidates. Although trainer style differed somewhat, each trainer subscribed to the overall goal of promoting positive behavioral change, primarily through conventional verbal, rather than nonverbal or body movement encounter.

Experimental and control subjects were given a pretraining test using four instruments: the Gordon Personal Profile, the Gordon Personal Inventory, the Edwards Personal Preference Schedule, and a Behavior Check List. The two Gordon Inventories contain the following eight factors: Cautiousness (impulse control), Original Thinking (flexibility and enjoyment of challenging tasks), Personal Relations (trust in people), Vigor

2. Because of involvement in individual therapy or group experience within the previous two years, three participants in the combined group, two in the spaced group, and 23 in the massed group had to be excluded from the experimental population.

(psychological energy investment), Ascendancy (social leadership), Responsibility (determination, perseverance, and dependability), Emotional Stability (evenness of moods), and Sociability (gregariousness). The Gordon Inventories were selected for the study because the eight personality factors which they assess represent behaviors found in normal populations and because they are forced-choice instruments that minimize the social desirability variable.

The Edwards Personal Preference Schedule contains 15 manifest need scales: Achievement, Deference, Order, Exhibition, Autonomy, Affiliation, Intracception, Succorance, Dominance, Abasement, Nurturance, Change, Endurance, Heterosexuality, and Aggression. This test was chosen because its factors augment the interpretation of the Gordon factors. For example, a low score (below the 25th percentile) on the Gordon personality factor of Personal Relations (trust) together with a low score on the EPPS Nurturance, Succorance, and Affiliation needs suggests that the interpersonal relationship difficulties are related to an inability to accept help from others and to fulfill others' needs. The same Gordon score on Personal Relations may, however, be interpreted differently: a high score (above the 75th percentile) on the EPPS need of Aggression together with a low score on need of Abasement suggests that it is anger and a tendency to blame others for one's misfortunes that underlie the difficulty in trusting people.

A Behavior Check List containing 34 potential problem areas was designed to determine an individual's self-awareness. The schedule, developed from interviews with individuals about their test scores on the three standardized instruments, translates the factors found on these tests into behavioral terms so that an individual can evaluate himself. Each subject in the study was asked to check the extent to which these areas were a problem for him.

From the scores on these instruments, four to eight problem dimensions representing behavioral areas in which each individual subject needs to change if he is to move toward greater psychological health were determined for each subject. Some dimensions were taken directly from the individual test scores; others were constructed from a pattern of test scores (inter-rater

reliability of the individualized dimensions was checked for a random sample of subjects; an r of 0.81 was found). For example, a Behavior Change Rating Sheet of a subject scoring low on the Cautiousness scale of the Gordon Personal Inventory (below the 25th percentile) may contain the following dimensions: "Increased self-discipline—better able to think through situations before acting and less inclined to make snap judgments." Conversely, if a subject's score on the Cautiousness scale were high (above the 75th percentile), the dimension may read: "Becoming freer and more flexible—less structured and more willing to try new ideas and alternatives." If an individual scores low (below the 25th percentile) on the Edwards Personal Preference Schedule on Nurturance, Succorance, and Affiliation, this pattern of test scores may be used as the following dimension for change: "Closer, more meaningful relationships—more comfort in accepting help from others and better able to give others what they need and want."

The individualized change dimensions were entered on the Behavior Change Rating Sheet, (BCRS) which was used as the criterion instrument to evaluate behavioral change. Each rater was asked to indicate one of five degrees of change for each dimension: 0 = None or Negative Change, 1 = Little Change, 2 = Some Change, 3 = Fair Amount of Change, and 4 = Much Change, checking the category which best described the extent to which a change had taken place in each dimension since a particular time. The experimental subjects completed the Behavior Change Rating Sheets at the end of sensitivity training and approximately three months later. Control subjects completed the Behavior Change Rating Sheets about two and one-half months after the pretraining test battery (a period comparable with the longest time spent by any experimental subject in sensitivity training) and again some three months later.

Five other individuals—the group trainer, two randomly selected group members, and two outside raters who were closely associated with the subject, e.g., spouse, friend—were also asked to complete a BCRS for each experimental subject. In the control condition, each subject and two closely associated individuals were asked to complete a Behavior Change Rating Sheet. Approximately 97 per cent of the rating sheets were returned

by the raters. In the three-month follow-up study, each experimental and control subject and their two outside raters were asked to complete once more a Behavior Change Rating Sheet identical to that given before. At this time, 92 per cent of the rating sheets were returned.

To determine whether there were significant differences between the total experimental group and the control group in the extent to which behavior change took place, the criterion scores derived from the Behavior Change Rating Sheets were compared, using a 4×1 analysis of variance and Scheffé tests. Two different sets of comparisons were made: (a) the criterion scores of all raters (self, outside raters, group trainer, and group members) were combined within each condition and the four conditions were compared with each of the others; and (b) the single criterion scores of each of the four raters were compared across the four conditions. In addition, analyses of variance were conducted comparing follow-up Behavior Change Rating Sheet data with the data derived from the posttraining Behavior Change Rating Sheets in each of the four conditions.

RESULTS

Hypothesis 1, which states that experimental groups show significantly more positive behavior changes than exhibited by control groups can be accepted. The data were analyzed for the combined and for the individual criterion scores. The analyses of variance and the Scheffé tests on the combined criterion scores³ for the experimental group (self, outside, group members, and trainer change ratings) and for the control group (self and outside ratings) show the following results:

1. Immediately after training, subjects in the three experimental conditions—spaced, massed, and combination—scored significantly greater behavioral change ($p < .01$) than did control group subjects (Table 1).
2. Follow-up data gathered two and a half months after the termination of training indicate that the groups in the spaced and combination conditions produced greater behavioral changes than did the control group ($p < .01$),

3. Further information concerning criterion scores for each type of rater (self, outside, group members, and trainer) can be obtained from Mitchell (1969).

TABLE 1. *Analysis of Variance—Behavior Change Ratings^a*

Case	Number of Subjects	Sample Size	Mean	Standard Deviation
Combined Ratings, Posttraining				
Spaced Condition	45	179	1.89	0.78
Massed Condition	36	128	1.94	0.93
Combined Condition	42	164	2.18	0.82
Control Condition	42	83	1.27	0.88
(F-Ratio = 21.70*)				
Combined Ratings, Follow-Up				
Spaced Condition	45	83	2.05	0.75
Massed Condition	36	62	1.85	1.04
Combined Condition	42	78	2.16	0.87
Control Condition	42	75	1.44	0.88
(F-Ratio = 9.89*)				
Self Ratings, Posttraining				
Spaced Condition	45	45	2.06	0.83
Massed Condition	36	35	1.82	0.82
Combined Condition	42	42	2.14	0.79
Control Condition	42	41	1.29	0.95
(F-Ratio = 8.54*)				
Self Ratings, Follow-Up				
Spaced Condition	45	43	2.16	0.74
Massed Condition	36	32	1.89	0.97
Combined Condition	42	40	2.17	0.76
Control Condition	42	38	1.45	0.97
(F-Ratio = 6.25*)				
Outside Ratings, Posttraining				
Spaced Condition	45	44	1.87	0.66
Massed Condition	36	28	1.95	0.97
Combined Condition	42	38	2.12	0.86
Control Condition	42	42	1.25	0.82
(F-Ratio = 8.65*)				
Outside Ratings, Follow-Up				
Spaced Condition	45	40	1.93	0.76
Massed Condition	36	30	1.81	1.14
Combined Condition	42	38	2.15	0.99
Control Condition	42	37	1.44	0.79
(F-Ratio = 3.95*)				
Trainer Ratings, Posttraining				
Spaced Condition	45	45	1.81	0.94
Massed Condition	36	29	2.27	0.97
Combined Condition	42	42	2.32	0.81
(F-Ratio = 3.99†)				
Group Member Ratings, Posttraining				
Spaced Condition	45	45	1.80	0.65
Massed Condition	36	36	1.76	0.95
Combined Condition	42	42	2.15	0.81
(F-Ratio = 2.82)				

^a Behavior Change Ratings are expressed on a four-point numerical scale

0 = none or negative change

1 = little change

2 = some change

3 = fair amount of change

4 = much change

* Significant at .01 level of confidence

† Significant at .05 level of confidence

while the subjects in the massed groups produced behavioral changes greater than those of the control subjects at the .06 level of significance.

Comparison of the experimental and control groups using individual criterion scores (self and outside ratings) at the close of training and again three months later yielded the following results: For the *self-rating* criterion, the combination and spaced groups showed significantly greater behavioral changes than did the control group, both at the close of training and three months later ($p < .01$). At the end of training, the massed group showed greater behavioral change than did the control group ($p < .06$). The difference between the massed group and the control group was not, however, significant when change was measured three months later.

Using the outside rating criterion, all three experimental groups exhibited significantly greater behavioral change than did the control group at the conclusion of training ($p < .01$). The follow-up data indicate, however, that the difference between the control group and the spaced and massed experimental groups is not significant. The combination group does show a greater behavioral change than did the control group ($p < .05$).

Hypothesis 2, which states that participation in combination sensitivity training groups produces significantly more positive behavioral change than does participation in spaced or massed groups and that participation in spaced groups produces significantly more positive behavioral change than massed group participation, can be only partially accepted (data tables available on request).

The results of the analyses of variance and Scheffé tests for multiple criterion scores (self, outside, group trainer, group member) at the end of training indicate that the groups in the combination condition showed significantly more behavior change than the group in the spaced condition ($p < .05$). All other data were nonsignificant. The trends of the data indicate, however, the following order of conditions in terms of increasing amount of change: (1) spaced, massed, combination (at the termination of sensitivity training), and (2) massed, spaced, combination (in the follow-up study).

Hypothesis 3 states that in a comparison of posttraining behavior change criteria with follow-up behavior change criteria, massed group subjects show a significant decrease in behavioral change during the three-month interval after sensitivity training, but spaced, combination, and control subjects show no significant increase or decrease. This hypothesis must be rejected, as only slight trends in the predicted direction are discernible (data table available on request).

DISCUSSION The results of the study indicate that sensitivity training does produce positive behavioral changes. These findings are in line with previous studies by Miles (1960, 1965), Bunker (1965), Valiquet (1964, 1968), Beer and Kleisath (1967), Zand, Steele, and Zalkind (1967), and Friedlander (1967), which support the notion that sensitivity training can be used to effect positive behavioral change. These studies should dispel the many doubts that have been raised about the utility of therapeutic interventions. Furthermore, significant behavioral changes, tailored to individual needs, not only occurred, but were maintained for three months. Although the findings were not significant, trends in the data suggest that long-term behavioral changes were less for groups who received massed training than for the other two conditions. This may suggest that spaced or distributed learning of social skills and personal habits is more effective than massed learning. It can be argued that a massed type training was included in the combined training condition, but this experience really consisted of spaced learning periods during the weekend sessions, quite unlike the learning experience in the massed condition. It is probable, however, that the addition of this type of massed training was instrumental in making the combined training condition more effective than the spaced condition.

Although data from this investigation indicate that positive behavioral changes result from participation in sensitivity training groups, the question whether sensitivity training groups also tend to promote negative changes is left unanswered. Two studies (Underwood, 1965; Yalom and Lieberman, 1971) have found that, besides producing positive changes, sensitivity training groups also produced a high percentage of unfavorable change. It is also possible that negative behavioral change is sometimes evidence of a painful loosening of old habits and can

be regarded as temporary until new positive behavior changes evolve.

There are many areas concerning sensitivity training, such as the role of the trainer, that have not been considered in this study. The results of this investigation suggest additional research on the following questions:

1. How would the results of this study be affected over a long period of time; i.e., would the data be substantially the same if behavioral change were measured a year or more after sensitivity training?
2. Is the use of individualized criteria an effective method of measuring behavioral change? If the problem dimensions were given at random to individual subjects, would they judge their change to be as great as do subjects who are given criteria that are individualized on the basis of their own test scores?
3. Do sensitivity training groups produce negative as well as positive behavior change? Are these negative changes transitory—do they lead to positive change at a later date?
4. Do sensitivity training groups that involve extended periods (two weeks or more) away from the home environment promote more behavioral change than do the types of sensitivity training groups investigated in this study?
5. Do subjects in training groups which rely principally on cognitive inputs manifest more behavioral change than do subjects in training groups which emphasize the expression of feelings, body awareness, and other forms of nonverbal communication?
6. How can a subject learn to bridge the gap between sensitivity training and the "back-home" environment more effectively?

- REFERENCES
- Bach, G. R. The marathon group: Intensive practice of intimate interaction. *Psychol. Rep.*, 1966, 18, 955-1002.
- Beer, M., & Kleisath, S. W. The effects of the Managerial Grid lab on organizational and leadership dimensions. In S. S. Zalkind (Chm.), Research on the impact of using different laboratory methods for interpersonal and organizational change. Symposium presented at the meeting of the American Psychological Association, Washington, D.C., September 1967.

- Bunker, D. R. Individual applications of laboratory training. *J. appl. Behav. Sci.*, 1965, 1 (2), 131-148.
- Friedlander, F. The impact of organizational training laboratories upon the effectiveness and interaction of ongoing work groups. *Personnel Psychol.*, 1967, 20, 289-308.
- Miles, M. B. Human relations training: Processes and outcomes. *J. counsel. Psychol.*, 1960, 7, 301-306.
- Miles, M. B. Changes during and following laboratory training: A clinical-experimental study. *J. appl. Behav. Sci.*, 1965, 1 (3) 215-242.
- Mitchell, Rie R. An evaluation of the relative effectiveness of spaced, massed, and combined sensitivity training groups in promoting positive behavior change. Unpublished doctoral dissertation. Univer. of California at Los Angeles, 1969.
- Underwood, W. J. Evaluation of laboratory method training. *Training Directors J.*, 1965, 19, 34-40.
- Valiquet, M. I. Contribution to the evaluation of management development program. Unpublished master's thesis, Massachusetts Institute of Technology, 1964.
- Valiquet, M. I. Individual change in a management development program. *J. appl. Behav. Sci.*, 1968, 4 (3), 313-325.
- Yalom, I. D., & Lieberman, M. A. A study of encounter group casualties. *Archives of General Psychiatry*, 1971, 25, 15-30.
- Zand, D. E., Steele, F. I., & Zalkind, S. S. The impact of an organizational development program on perceptions of interpersonal group and organizational functioning. In S. S. Zalkind (Chm.), Research on the impact of using different laboratory methods for interpersonal and organizational change. Symposium presented at the meeting of the American Psychological Association, Washington, D.C., September 1967.

There is nothing that can be changed more completely than human nature, when the job is taken on early enough.

—G. B. Shaw, Preface
to *On the Rocks*