Donald W. Taylor
Paul C. Berry
Clifford H. Block

Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking?

In an experiment designed to answer the title question, twelve groups of four men each and forty-eight individuals followed the four basic rules of brainstorming in attacking the same three problems in the same order. Upon completion of the experiment, a table of random numbers was used to divide the forty-eight individual subjects into twelve nominal groups of four men each. The performance of each nominal group was then scored as though its members had actually worked together. The achievement of these nominal groups thus provided a measure of the performance to be expected if group participation neither facilitates nor inhibits creative thinking. When compared with that of the twelve nominal groups, the performance of the twelve real groups was found to be markedly inferior with respect to: (a) mean total number of ideas produced; (b) mean number of unique ideas produced; (c) three different measures which weighted the ideas produced differentially with respect to quality. To the extent that the results of the present experiment can be generalized, it must be concluded that group participation when using brainstorming inhibits creative thinking.[1]

[1]The experiment reported here is one in a series of psychological studies of problem solving being carried out under Project NR 150–166 and supported by Contract Nonr 609(20) between Yale University and the Office of Naval Research. The present article presents essentially the same information, omitting only certain details of

Donald W. Taylor is professor of personnel administration in the Department of Industrial Administration and also professor in the Department of Psychology, Yale University. Paul C. Berry and Clifford H. Block are graduate students and assistants in research in the Department of Psychology.

BRAINSTORMING was originated and first used by Alex F. Osborn in 1939 in the advertising agency Batten, Barton, Durstine & Osborn, which he then headed.[2] Within recent years its use has grown rapidly. A large number of major companies, units of the Army, Navy, and Air Force, and various federal, state, and local civilian agencies have employed the technique, and instruction has been given in a number of colleges and universities in its use.[3] Although an occasional critical voice has been raised,[4] brainstorming may be said to have achieved wide acceptance as a means of facilitating creative thinking.

The purpose of brainstorming is to free individuals from inhibition, self-criticism, and criticism by others in order that in response to a specific problem they may produce as many different ideas as possible. The assumption is that the larger the number of ideas produced, the greater the probability of achieving an effective solution. Brainstorming is characterized by four basic rules:

(1) *Criticism is ruled out.* Adverse judgment of ideas must be withheld until later.

(2) *"Free-wheeling" is welcomed.* The wilder the idea, the better; it is easier to tame down than to think up.

(3) *Quantity is wanted.* The greater the number of ideas, the more the likelihood of winners.

(4) *Combination and improvement are sought.* In addition to contrib-

experimental procedure and statistical analysis, as the following technical report previously issued under the contract: Donald W. Taylor, Paul C. Berry, and Clifford H. Block, *Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking?* (New Haven: Departments of Industrial Administration and Psychology, Yale University, 1957). Copies of this technical report are available in major university libraries. Permission is granted for reproduction, translation, publication, use, and disposal of the present article in whole or in part by or for the United States Government.

[2]Alex F. Osborn, *Applied Imagination,* rev. ed. (New York, 1957).

[3]*Ibid.* See also Brainstorming—New Ways to Find New Ideas, *Time,* 69 (Feb. 18, 1957), 90.

[4]Bernard S. Benson, Let's Toss This Idea Up...*Fortune,* 56 (Oct. 1957), 145–146.

uting ideas of their own, participants should suggest how ideas of others can be turned into *better* ideas; or how two or more ideas can be joined into still another idea.[5]

Brainstorming ordinarily involves not only following the four basic rules but also group collaboration in attacking the problem. Osborn emphasizes the value of group interaction in facilitating the flow of ideas.[6] It was this characteristic of brainstorming which was of primary interest in the present study.

Certain rather informal experiments very briefly described by Osborn appear to support the view that group collaboration in brainstorming does increase production of ideas.[7] Moreover, many practical experiences described by Osborn also seem to provide support persuasive of this view. An adequate answer to the question, however, demands adequately controlled experiments appropriately designed and with results subjected to thorough statistical analysis. No such experiment was known to the authors of the present report—certainly none has been published in sufficient detail to permit evaluation of the design, the analysis, and the conclusions reached.

The present experiment employed a design previously developed by Taylor for use in studies of group problem solving where the problems involved have logically correct solutions.[8] Earlier studies of such group problem solving were concerned with a comparison of the achievement of groups of various sizes with that of individuals. However, the performance of a group should be superior to that of an individual, simply because in the group more individuals are working on the problem. On the assumption of the appropriate null hypothesis, namely, that working in a group has no effect either positive or negative upon individual performance, Taylor[9] and Lorge and Solomon[10] independently have presented a simple

[5]Osborn, *Applied Imagination,* p. 84.

[6]*Ibid.,* pp. 82–83.

[7]*Ibid.,* pp. 82, 228.

[8]Donald W. Taylor, "Problem Solving by Groups," in *Proceedings of the Fourteenth International Congress of Psychology, Montreal, June 1954* (Amsterdam, 1955), pp. 218–219. See also Donald W. Taylor and Olga W. McNemar, "Problem Solving and Thinking," in Calvin P. Stone, ed., *Annual Review of Psychology* (Stanford, Calif., 1955), VI, 455–482.

[9]Taylor, *op. cit.*

[10]Irving Lorge and Herbert Solomon, Two Models of Group Behavior in the Solution of Eureka-Type Problems, *Psychometrika,* 20 (1955), 139–148.

mathematical model for predicting the performance of a group of a given size from a knowledge of individual performance. By comparing actual group achievement with that predicted from the model, one can determine whether group participation facilitates or inhibits problem solving.

Taylor has also developed an experimental design which provides an alternative method of testing the same null hypothesis as that represented by the model.[11] Individuals are randomly assigned to work either alone or in groups of a given size on a series of problems. The number of individuals working alone should be about equal to that working in groups. After the experiment is completed, those who actually worked alone are divided at random into nominal groups of the same size as the real groups. The performance of the nominal groups is then scored as though the members of the group had worked together. The achievement of the nominal groups thus provides a measure of the performance to be expected under the null hypothesis. If the performance of the real groups is superior to that of the nominal groups, group participation facilitates performance; if it is inferior, group participation inhibits it.

This design, with appropriate modification in the scoring of responses for nominal groups, was employed in the present experiment to provide an answer to the question: Does group participation when using brainstorming facilitate or inhibit creative thinking?

THE EXPERIMENT

Subjects

The ninety-six Yale juniors and seniors who served as subjects in this experiment were all at the time enrolled in a course in Psychology of Personnel Administration taught by the first author. Each week, in addition to two lectures to the entire class, the course included an analysis of a case carried out in small discussion groups;[12] each group had its own student leader, this task being rotated among the members of the group. As a result of such case

[11]Taylor, *op. cit.*

[12]Each period was devoted to discussion of a case selected from John D. Glover and Ralph M. Hower, *The Administrator* (Homewood, Ill., 1957). By the beginning of the present experiment, each group had completed a total of six cases.

discussion and of the way in which subjects were assigned, each real group in the present experiment was not, as often must be the case in studies of group problem solving, an *ad hoc* group of individuals meeting for the first time; instead, each real group included men who not only knew each other but who also had worked together effectively in small-group discussion over a considerable period of time. At the same time, the procedure used in assigning subjects was such that those assigned to work in groups and those assigned to work alone could legitimately be regarded as random samples from the same population.

For the case discussions, the class was divided into five sections of about equal size which met at different times during the week. These five sections were in turn divided into two discussion groups of about nine to eleven men each which met simultaneously in separate rooms; the instructor divided his time between the two groups meeting simultaneously. The following procedure was used to assign half of the ninety-six men to work in real groups of four and the other half to work alone. From each of the ten discussion groups in the class, four men were picked at random to form an experimental group, thus providing ten experimental groups; and from two of the ten discussion groups, an additional four men were picked at random to provide two more groups, for a total of twelve experimental groups. The remaining men in the ten discussion groups, forty-eight in all, served as individual subjects. Two points perhaps should be added here. The original registration of members of the class in the five sections was dependent primarily upon class schedules and had no known relation to ability. The division of these sections into two discussion groups was also unrelated to ability.

Problems

Considerable pretesting was carried out in an effort to ensure that the problems selected for use in the present experiment would be as suitable as possible for attack by the brainstorming procedure and also for use with Yale students. Initially, several problems were selected from those suggested by Osborn as appropriate for use with brainstorming;[13] others were originated by the present writers,

[13]Alex F. Osborn, *A Manual of Instruction for Teachers Using the Textbook "Applied Imagination"* (New York, 1953).

and some were obtained from other sources. With Yale students both individually and in small groups used as subjects, a total of about twenty problems was tried out; none of those, of course, who participated in these pretests served as subjects in the main experiment. On the basis of this pretesting, three problems were selected which seemed to be of interest to Yale students, productive of many and varied responses, and appropriate for use with brainstorming. The three problems were as follows:

1. Each year a great many American tourists go to visit Europe. But now suppose that our country wished to get many more European tourists to come to visit America during their vacations. What steps can you suggest that would get more European tourists to come to this country?

2. We don't think this is very likely to happen, but imagine for a moment what would happen if everyone born after 1960 had an extra thumb on each hand. This extra thumb will be built just as the present one is, but located on the other side of the hand. It faces inward, so that it can press against the fingers, just as the regular thumb does now. Here is a picture to help you see how it will be. (A line drawing of a hand with two thumbs was shown by the experimenter at this point in the reading of the problem and then left in full view on the table during the entire period of work on the problem.[14]) Now the question is: What practical benefits or difficulties will arise when people start having this extra thumb?

3. Because of the rapidly increasing birth rate beginning in the 1940s, it is now clear that by 1970 public school enrollment will be very much greater than it is today. In fact, it has been estimated that if the student-teacher ratio were to be maintained at what it is today, 50 per cent of all individuals graduating from college would have to be induced to enter teaching. What different steps might be taken to insure that schools will continue to provide instruction at least equal in effectiveness to that now provided?

For brevity's sake, the three problems were referred to as the "Tourists Problem," the "Thumbs Problem," and the "Teachers Problem."[15]

[14]For a copy of the line drawing, see Taylor, Berry, and Block, *op. cit.*, p. 6.

[15]Although the third problem would be more accurately described as the education problem, it was called the "Teachers Problem" for the sake of symmetry.

Procedure

In a single experimental session lasting about one hour, the three problems were presented in the order in which they are listed above to each of the twelve groups and to each of the forty-eight individuals. The second and third authors of the present report conducted the sessions, each one conducting six group and twenty-four individual sessions. The assignment of groups and individuals to experimenters was largely a matter of chance. Group and individual sessions were alternated in such a way that on any given date about the same proportion of group and individual sessions had been completed.

Both experimenters were advanced graduate students in psychology. Both were familiar with Osborn's writing concerning brainstorming. Both had participated in the pretesting described above and thereby gained experience with the procedures of brainstorming. Both personally believed that group brainstorming was an effective procedure for facilitating the production of ideas.

Very shortly before the present experiment began, one lecture to the class as a whole was devoted to creative thinking, with particular attention to brainstorming. The origin, nature, and widespread use of brainstorming was described, the purpose being to create interest in the procedure and as favorable as possible an attitude toward it. The shortage of controlled experimental studies of this and similar procedures was described. Finally, the students were asked to participate as subjects in the experiment and were promised a report of the results when it was available.

During the pretesting, both with individuals and with small groups, attention was devoted to the question of what length of time should be allowed for work on each of the problems selected for use. What was wanted was a span of time long enough so that members of groups of four would have adequate opportunity to express all the ideas which occurred to them within the working period and at the same time short enough so that individuals would not become bored by being forced to continue work on a problem long after they had essentially exhausted their ideas. Pretesting showed that the rate at which ideas were produced on the problems decreased with time. A time limit of twelve minutes was finally selected as one which would permit group members to express

all ideas occurring to them within the work period (though not to exhaust all possible ideas) and yet which would not result in excessive periods of silence for individual subjects. Actually, for both individuals and groups, appreciable periods of silence appeared between responses near the end of the twelve minutes.

Instructions

The experimenter began each group or individual session by reading aloud instructions,[16] the first part of which was designed to ensure a high degree of subject motivation. First, it was explained that the experiment was part of a program of research on problem solving and creative thinking being carried on under a contract with the Office of Naval Research. Secondly, each subject was asked to read the first three paragraphs (about 60 per cent) of a one-page feature article on brainstorming which, by coincidence, had appeared a few weeks earlier in *Time*.[17] These three paragraphs were quite favorable to brainstorming, describing its use in major U.S. companies. Finally, subjects were specifically asked to do as well as they could, and it was impressed upon them that the success of the experiment was contingent upon such effort.

The second part of the instructions was designed to make certain that subjects fully understood the nature of brainstorming. The four basic rules, as already described, were explained.

The instructions included a brief explanation that, because writing was slow and inaccurate, the discussion of each problem would be recorded (using an Edison *Voicewriter* and conference microphone). Subjects were also requested to avoid discussing any of the problems with anyone at Yale until the experiment was completed; evidence available indicates that this request was generally honored.

Opportunity was provided for questions and, finally, the four basic rules were reiterated. The experimenter presented each problem by reading it aloud to the individual or group. If any question was raised, it was answered by rereading part or all of the original statement. The actual copy of the problem was never given to the individual or group. On those few occasions when a critical com-

[16]For a copy of the instructions, see Taylor, Berry, and Block, *op. cit.*, pp. 37–39.

[17]Brainstorming—New Ways to Find New Ideas, *Time*, 69 (Feb. 18, 1957), 90.

ment was made during work on a problem, the experimenter called attention again to the basic rule against criticism. Comments volunteered by many of the students during the experiment support the belief that with a few possible exceptions they were motivated to do as well as they could. Those who participated as members of groups seemed particularly enthusiastic about their experience.

RESULTS

The first step following the completion of the experimental sessions was the division of the forty-eight individual subjects into twelve nominal groups of four each. This was done in order to permit comparison of real group performance not only with that of individuals but also with that to be expected on the basis of the hypothesis that working in the group has no effect either positive or negative upon the performance of its members. A table of random numbers was employed to divide the twenty-four individual subjects who had worked with the first experimenter into six nominal groups of four; the same procedure was used to divide the twenty-four who had worked with the second experimenter into an additional six nominal groups. This particular procedure was necessary if a test were to be made of any possible difference between the two sets of six nominal groups resulting from differences between the experimenters. Inspection of the data later obtained, however, revealed no possible significant difference between experimenters in the results.

Next, an essentially complete typewritten transcript was made of the recording of the responses of each real group and of each individual subject to each of the three problems. An analysis of all the transcripts for the Tourists Problem resulted in the preparation of a master list of all the different steps suggested to "get more European tourists to come to this country." Similar analyses yielded for the Thumbs Problem a master list of all the different suggestions of "what practical benefits or difficulties will arise when people start having this extra thumb," and for the Teachers Problem a master list of all the different steps that "might be taken to insure that schools will continue to provide instruction at least equal in effectiveness to that now provided." The three master lists were, of course, organized in terms of appropriate categories and

subcategories. After preparation, each list was carefully reexamined to make certain that no essentially similar item appeared twice. What constituted a discriminable difference between two suggestions depended, of course, upon the judgment of the individual preparing the list, but in general it may be said that whenever the second suggestion appeared clearly to add something to the first, it was regarded as a different suggestion. The master list for the Tourists Problem included a total of 483 different steps, for the Thumbs Problem a total of 791 different consequences, and for the Teachers Problem a total of 513 different steps. The magnitude of such lists may be visualized from the fact that, when typed on sheets measuring 8½ by 11 inches, the Tourists list covered 14 pages, the Thumbs list 32 pages, and the Teachers list 17 pages.

For each master list, sets of data sheets were prepared matching the list in such a way that the analysis of the protocol for any individual or real group could be recorded simply by placing a check in the row and column corresponding respectively to the particular idea and the particular group or individual. Four adjoining columns were used to record the analyses of the protocols for the individuals included in each nominal group of four. In a single adjacent column, a check mark was placed whenever a check had been made in that row for any one or more of the four individuals. Thus this column provided a record of nominal-group performance, consistent with the requirement of the experimental design that the nominal group be scored as though the members had worked together.[18] If any member of the nominal group presented an idea, the group was regarded as having presented it. If two or more members of the nominal group presented the same idea, it was still recorded as only one idea. The total number of

[18]The procedure described had not only the advantage of clerical and computational convenience but also the more important advantage of avoiding any possible effect of bias upon the scoring of nominal-group performance. Judgments of whether two similar ideas are the same or different are necessarily somewhat subjective. Had nominal-group scores been obtained by working directly with the original protocols, such judgments would have been made with knowledge of whether the two similar ideas had been produced by two members of the same nominal group. Such knowledge might conceivably have biased the judgments. In the procedure employed here, such judgments were made in the course of the construction of the master lists and without such knowledge.

different ideas presented by each real group, individual, or nominal group was obtained simply by taking the total number of check marks in the appropriate column.

Table 1 presents the mean number of responses by individuals and real groups to each of the three problems. On each of the three problems, the mean number of ideas presented by real groups is much larger than that presented by individuals. The appropriate analysis of variance[19] briefly summarized in Table 1 shows that this difference between real groups and individuals with an F of 71.2 is significant at well beyond the .0001 level.[20] The analysis also shows that the differences among the three problems in mean number of

Table 1. Mean total number of responses to each problem by individuals and real groups.

	Tourists	Thumbs	Teachers	Mean of means
Individuals	20.7	19.9	18.2	19.6
Real groups	38.4	41.3	32.6	37.5
Mean of means	29.6	30.6	25.4	

Analysis of variance	d. f.	F	p
Individuals *vs.* real groups	1, 58	71.2	.0001
Among problems	2, 116	8.5	.001
Interaction	2, 116	4.96	.01

[19]The analyses of variance reported in Tables 1, 2, 3, 5, 6, and 7 all involve what is described as a "Type I Mixed Design" by E. F. Lindquist, *Design and Analysis of Experiments in Psychology and Education* (Boston, 1953), pp. 266–273. To save space, only the most essential information resulting from the several analyses is presented in the tables just cited. For summaries of the same analyses in more conventional form, including both sums of squares and mean squares, see Taylor, Berry, and Block, *op. cit.*, pp. 13, 15, 17, 24–26.

[20]In other words, the probability is less than 1 in 10,000 of obtaining by chance a difference as large or larger than that between real groups and individuals reported in Table 1. Those unfamiliar with analysis of variance may wish to consult Quinn McNemar, *Psychological Statistics* (New York, 1955), pp. 243 ff.

responses is significant at the .001 level and that the interaction between the two primary variables is also significant.[21]

Table 2 shows that the mean number of responses produced by nominal groups was considerably larger than that produced by real groups on each of the three problems. The analysis of variance indicates that this superiority of nominal to real groups is significant at far beyond the .0001 level. The difference among the three problems in number of responses is again significant, but in this case the interaction does not even approach significance.

Table 2. Mean total number of responses to each problem by real groups and nominal groups.

	Tourists	Thumbs	Teachers	Mean of means
Real groups	38.4	41.3	32.6	37.5
Nominal groups	68.3	72.6	63.5	8.16
Mean of means	53.4	57.0	48.0	
Analysis of variance		d. f.	*F*	p
Real *vs.* nominal groups		1, 22	96.3	.0001
Among problems		2, 44	7.8	.005
Interaction		2, 44	.09	—

It seemed important to compare the performance of real and nominal groups not only in terms of the number of ideas produced but also in terms of the originality and quality of these ideas. For this purpose, additional analyses were undertaken.

A large proportion of the responses to any one of the problems was, of course, produced by more than one of the nominal or real groups, a small number of the ideas on each of the problems being suggested by nearly all of the twenty-four groups. On each problem, however, an appreciable number of suggestions was made by

[21]Interaction refers to the variation in the effect of one independent variable upon the dependent variable as a function of a second independent variable (see McNemar, *op. cit.*, pp. 283, 301–303).

only one of the twenty-four groups; these may be described as unique responses. The number of such unique ideas provides one satisfactory measure of the originality of the performance of a particular group.

In Table 3 are given the mean number of unique responses produced by real and nominal groups on each of the three problems. The superiority of the nominal to the real groups on this measure is significant at the .005 level. The difference among the three problems in mean number of unique responses is also significant, but the interaction is not.

Table 3. Mean numbers of unique responses to each problem.

	Tourists	Thumbs	Teachers	Mean of means
Real groups	7.5	17.7	7.3	10.8
Nominal groups	13.7	28.1	17.5	19.8
Mean of means	10.6	22.9	12.4	

Analysis of variance	d. f.	*F*	p
Real *vs.* nominal groups	1, 22	11.4	.005
Among problems	2, 44	42.1	.0001
Interaction	2, 44	1.29	—

Comparison of Table 3 with Table 2 suggests that the difference between nominal and real groups in number of unique responses might result simply from the difference between the two in total number of responses, in other words, that nominal groups might produce more unique responses just because they produce more responses. To test this hypothesis, the method of analysis of covariance was employed, with the number of responses to each problem as the control variable.[22] For each of the three problems a separate

[22]The analysis of covariance provides a method for determining whether there is any significant difference between nominal and real groups in number of unique ideas produced after allowance is made for the difference in total number of ideas produced, the number of unique ideas being correlated with the total number of ideas. For an introduction to analysis of covariance, see McNemar, *op. cit.*, pp. 343 ff.

analysis of variance was carried out and then extended to include an analysis of covariance.[23]

The separate analyses of variance for the three problems yielded results consistent with those of the over-all analysis of variance reported in Table 3. The superiority of nominal to real groups in number of unique responses is significant at the .02 level for the Thumbs Problem and at the .001 level for the Teachers Problem, although it just reaches the .06 level for the Tourists Problem. The analyses of covariance show that after adjustment for differences in total number of responses, there is no significant difference between real and nominal groups in number of unique responses on either the Tourists Problem or the Teachers Problem.[24] On the Thumbs Problem, however, the difference after adjustment is significant at the .02 level. Moreover, computation of the adjusted means shows that this difference actually favors the real groups. In other words, when allowance is made for difference in total number of responses produced, the real groups produced more unique responses than the nominal groups, but only on the Thumbs Problem.

Detailed examination of the 483 different suggestions for solution of the Tourists Problem and of the 513 different suggestions for solution of the Teachers Problem indicated that these suggestions differed in quality with respect to at least three dimensions: feasibility, effectiveness, and generality. Accordingly, five-step rating scales were constructed for use in measuring these three. The intention was to construct scales such that the successive steps on each scale would be subjectively equal, each step would be relatively unambiguous, and all five steps would actually be used in

[23]Initially it appeared appropriate to extend the analysis of variance reported in Table 3 to include an analysis of covariance by using the general procedure described by Lindquist, *op. cit.*, pp. 332–333. One of the assumptions, however, which would be involved in such an analysis for all three problems simultaneously would be that the regressions within the six subgroups are both linear and homogeneous (*ibid.*, p. 323). Inspection of the appropriate scatter diagrams suggested that the slope coefficients for the real and nominal groups were homogeneous on each of the three problems but that those for the Thumbs Problem differed considerably from those for the other two problems. Accordingly, it was decided to carry out separate analyses of covariance for the three problems.

[24]For the conventional tables summarizing these three analyses of variance and of covariance, see Taylor, Berry, and Block, *op. cit.*, pp. 18–19.

rating. After some pretesting and revision, three scales were obtained which were considered acceptable for use.

The 791 different responses made to the Thumbs Problem differed from those made to the other two problems in that they represented anticipated consequences instead of suggested steps for solution. For this reason only one of the three rating scales constructed for rating responses to the other two problems, namely, generality, appeared equally applicable in the case of the Thumbs Problem. For this problem, however, analogous to feasibility and effectiveness on the other problems were the dimensions of probability and significance, respectively. Accordingly, two additional rating scales were constructed by the same method to measure these latter variables.[25]

All three authors of the present report participated in the rating of the responses to the three problems. The first author rated the responses to the Tourists, Thumbs, and Teachers Problems on effectiveness, probability, and generality, respectively; the second author on generality, significance, and feasibility, respectively; and the third author on feasibility, generality, and effectiveness, respectively. Thus the responses to each problem were rated on three different scales by three different raters, presumably increasing the independence of the ratings of the three characteristics. Each rater employed a different scale for each of the three problems, thus presumably minimizing the possibility that a single idiosyncratic interpretation of any of the scales would occur for all three problems. Since the use of a single scale for a single problem required the making of 483 to 791 judgments, and hence a great deal of time, it was not considered worth while to have the ratings replicated. The mean rating for all responses to the Tourists Problem was 3.27 on feasibility, 1.97 on effectiveness, and 2.20 on generality; for all responses to the Thumbs Problem, 2.04 on probability, 1.88 on significance, and 2.21 on generality; for all responses to the Teachers Problem, 2.83 on feasibility, 2.69 on effectiveness, and 2.39 on generality.

The intercorrelations among the dimensions rated are given in

[25]For a copy of each of the five rating scales, see Taylor, Berry, and Block, *op. cit.*, pp. 40–42. Inspection of these scales provides the best available definition of each of the variables.

Table 4 for each of the three problems. Although all the correlations except one are significantly different from zero at the .01 level, the fact that all are relatively small suggests that the dimensions chosen for rating may be reasonably independent. It should be added, of course, that since the ratings being correlated were made by different raters, the obtained coefficients are probably somewhat lower than they would otherwise be because of the limited reliability of the ratings.

Table 4. Intercorrelations among ratings of responses to each of three problems.

	Feasibility *vs.* effectiveness	Feasibility *vs.* generality	Effectiveness *vs.* generality
Tourists	.25*	−.01	.27*
Teachers	.15*	.16*	.35*
	Probability *vs.* significance	**Probability *vs.* generality**	**Significance *vs.* generality**
Thumbs	.35*	.10*	.38*

*Significant at or beyond the .01 level.

One additional point concerning the procedure used in rating deserves emphasis. All ratings were made of the responses as they appeared on the master list for the given problem and without any knowledge of whether the response had been made by real or nominal groups. This was done, of course, to eliminate any possible tendency of any rater to bias his ratings to favor either real or nominal groups.

After the ratings were all recorded on the master lists, these lists were then used in conjunction with the matched data sheets to obtain scores on each of the three rated dimensions on each problem for each of the real and nominal groups. The score for each group for a given problem and a given dimension was simply the sum of the ratings on that dimension of the responses given by the group to that problem.

Table 5. Feasibility, effectiveness, and generality of responses to Tourists Problem.

	Feasibility	Effectiveness	Generality	Mean of means
Real groups	129.1	75.3	82.4	95.6
Nominal groups	215.1	133.0	141.1	163.1
Mean of means	172.1	104.1	111.8	
Analysis of variance		d. f.	*F*	p
Real *vs.* nominal groups		1, 22	43.5	.0001
Among measures		2, 44	199.2	.0001
Interaction		2, 44	9.3	.001

A comparison of the mean scores of real and nominal groups on each of the three rated dimensions is presented in Tables 5, 6, and 7 for the Tourists, Thumbs, and Teachers Problems, respectively. On each of the three dimensions for each of the three problems, the mean for the nominal groups is much larger than that for the real groups. The analyses of variance summarized in Tables 5, 6, and 7 show that this superiority of the nominal to the real groups is significant well beyond the .0001 level for each of the three

Table 6. Probability, significance, and generality of responses to Thumbs Problem.

	Probability	Significance	Generality	Mean of means
Real groups	80.3	78.7	92.1	83.7
Nominal groups	157.3	133.2	155.8	148.8
Mean of means	118.8	106.0	124.0	
Analysis of variance		d. f.	*F*	p
Real *vs.* nominal groups		1, 22	114.6	.0001
Among measures		2, 44	4.60	.05
Interaction		2, 44	1.49	—

Table 7. Feasibility, effectiveness, and generality of responses to Teachers Problem.

	Feasibility	Effectiveness	Generality	Mean of means
Real groups	92.6	90.4	78.5	87.2
Nominal groups	179.6	163.3	148.5	163.8
Mean of means	136.0	126.9	113.5	

Analysis of variance	d. f.	*F*	p
Real *vs.* nominal groups	1, 22	55.5	.0001
Among measures	2, 44	38.7	.0001
Interaction	2, 44	6.21	.005

problems. The differences among the three measures are also significant for all three problems, and the interaction is significant for only the Tourists and Teachers Problems; these latter findings, however, are of limited interest because of the nature of the measures involved.

Comparison of Tables 5, 6, and 7 with Table 2 suggests that, as was true of unique responses for two of the problems, the differences between nominal and real groups on each of the rated dimensions for each of the three problems might have resulted simply from the difference between such groups in total number of responses. The score of a nominal or real group on any given quality dimension depends both on the rating of each idea given and on the total number of ideas produced. If the average rating on such a dimension is the same for the nominal as for the real groups, then the difference in score on that dimension would reflect only the difference in total number of responses. To determine whether there was any significant difference between nominal and real groups on these measures weighted for quality after allowance is made for difference in total number of responses, an analysis of covariance was carried out for each of the three problems.[26]

[26]Inspection of the appropriate scatter diagrams indicated that the regressions of each of the three rated variables upon total responses not only were linear for both real and nominal groups but had essentially the same slope in all six subgroups.

After adjustment for differences in total number of responses, no significant differences remain between real and nominal groups on the three measures for either the Tourists or the Teachers Problem. After adjustment, however, the difference on the Thumbs Problem remains significant at the .03 level.[27] Computation of the adjusted means shows that this remaining difference still favors the nominal groups. Thus for the Thumbs, but not for the Tourists and Teachers Problems, there is a superiority of the nominal over the real groups on the three quality measures over and above that accounted for by a superiority in total number of responses.

DISCUSSION

The comparisons of primary interest in the several analyses are, of course, those between the performance of real groups and that of individuals or of nominal groups. The analyses, however, also involved other comparisons to which attention may best be given first.

Three different problems were employed in the experiment, the purpose being to give greater generality to the results; if only a single problem had been employed the question might well have been raised whether the results could be generalized beyond that problem. Significant differences were found among the three problems in mean total number of responses (Tables 1 and 2) and also in mean number of unique responses (Table 3). But these findings are of little interest because different problems would obviously be expected to produce different numbers of ideas and also different numbers of unique ideas.

On both the Tourists Problem (Table 5) and the Teachers Problem (Table 7), significant differences were obtained among the mean scores based upon differential weighting of the ideas produced with respect to feasibility, effectiveness, and generality, respectively. Similarly, significant differences were obtained for the Thumbs Problem (Table 6) among the mean scores on the

Accordingly, since the other assumptions involved in such a procedure (Lindquist, *op. cit.*, p. 323) appeared to be satisfied, the analyses of variance reported in Tables 5, 6, and 7 were extended to include the corresponding analyses of covariance (*ibid.*, p. 332–333).

[27]For more complete information concerning the results of the analyses of covariance, see Taylor, Berry, and Block, *op. cit.*, pp. 24–26.

three measures involving differential weighting with respect to probability, significance, and generality, respectively.[28] These findings, however, are also of little interest, since the differences between means on these measures depend directly upon the particular scales employed to rate the ideas. Changes in the rating scales would change the magnitude and even the direction of the differences between such means.

The first important finding was that on each of the three problems the mean total number of ideas produced by the twelve groups was considerably larger than the mean number produced by the forty-eight individuals, the difference being highly significant (Table 1). It is true that the interaction is significant, indicating that the difference between real group and individual performance does vary among the three problems. But on all three problems group performance is clearly superior to individual performance. Such group superiority may very well account for the widespread impression that group participation does facilitate production of ideas. The individual who compares his own performance working alone with that of a group in which he participates at another time may understandably conclude that group interaction stimulates creative thinking, whether or not this is in fact the case. Many of those participating in the groups in the present experiment made comments indicating that they believed such participation had been stimulating.

The comparison of group performance with individual performance does not, however, provide an adequate answer to the question: Does group participation when using brainstorming facilitate or inhibit creative thinking? To answer this question, the performance of the twelve real groups was compared with that of the twelve nominal groups on each of the three problems with respect to (a) mean total number of ideas produced, (b) mean number of unique ideas produced, and (c) the three measures which involved the weighting of the ideas with respect to quality. The results of these several analyses were both clear-cut and consistent.

[28]These analyses were carried out for a problem at a time rather than a measure at a time because the Thumbs Problem involved different measures from the other two.

The performance of the twelve real groups is markedly inferior to that of the twelve nominal groups both in terms of number of ideas produced (Table 2) and in terms of number of unique ideas produced (Table 3). Since in neither case was the interaction significant, these findings apply equally to all three problems. The mean scores of the real groups on the three weighted measures were also markedly inferior to those of the nominal groups for the Tourists (Table 5), Thumbs (Table 6), and Teachers Problem (Table 7). In the case of the Thumbs Problem, the fact that the interaction is not significant indicates that the same result was obtained with all three measures. In the case of both the Tourists and Teachers Problems, this interaction is significant. On both problems, however, although the magnitude of the difference varies, the real group mean scores on all three measures are quite inferior to those of nominal groups (Tables 5 and 7). In brief, the performance of the real groups is inferior to that of the nominal groups on all three problems with respect to each and all of the measures of performance employed.

To the extent that the results of the present experiment can be generalized, it must be concluded that group participation when using brainstorming *inhibits* creative thinking. What accounts for such inhibition? Although data are not available to provide an adequate answer, two suggestions may be made. In brainstorming strong emphasis is placed upon avoiding criticism both of one's own ideas and of the ideas of others. Nevertheless, it appears probable that the individual working in a group feels less free of possible criticism by others even when such criticism is not expressed at the time than does the individual working alone. To the extent that this is true, group participation is inhibiting. A second reason is that group participation may reduce the number of different ideas produced. A given number of individuals working in a group appear more likely to pursue the same train of thought—to have the same set or the same approach to the problem—than do the same number of individuals working alone. The greater the variety of set, train of thought, or approach, the greater would be the expected number of different ideas produced. To the extent that group participation reduces such variety, it inhibits production of ideas.

Certain supplementary analyses were carried out to aid in the interpretation of the major comparisons. To determine whether any significant differences between nominal and real groups would remain in number of unique responses after allowance had been made for difference in total number of responses, analyses of covariance were carried out for each of the three problems. These analyses showed that after adjustment with total number of responses as the control variable, the difference in number of unique responses is no longer significant for either the Tourists or the Teachers Problem. This finding may be interpreted to indicate that on these two problems the superiority of nominal over real groups is largely a matter of difference in number of responses and not, in addition, of any difference in the uniqueness or originality of the responses produced. In contrast, the analysis of covariance for the Thumbs Problem showed that, after adjustment for differences in total number of responses, the difference in number of unique responses not only is significant but actually favors the real groups. In other words, the degree of uniqueness of whatever responses were produced was somewhat greater for real than for nominal groups, but only on this one problem. To avoid any possible confusion at this point, it may be worth stating again the more important finding reported above that with respect to *number* of unique responses the nominal groups were superior to the real groups on all three of the problems.

A similar analysis of covariance was carried out for each of the three problems to determine whether differences between the nominal and real groups on the three quality measures also resulted simply from the difference in total number of responses. The results of such analyses showed that, after adjustment for differences in total number of responses, no significant differences between nominal and real groups remained on the three measures involving weighting for feasibility, effectiveness, and generality, respectively, for either the Tourists or the Teachers Problem. In contrast, after such adjustment for differences in total number of responses, significant differences favoring the nominal groups remain for the Thumbs Problem on the measures weighted for probability, significance, and generality of ideas, respectively. Taken together, these findings may be interpreted to mean that the

superiority of nominal over real groups on the three quality measures is largely a matter of difference in number of responses and only to a limited degree, if any, a matter of difference in the quality of the ideas produced. But it must be kept in mind that because the nominal groups produce more ideas they also produce more ideas which equal or exceed any given level of quality.

To what extent can the conclusion be generalized that group participation when using brainstorming inhibits creative thinking? Would different results be obtained if some other kind of problems were employed, if the subjects were highly trained in brainstorming, if more time were allowed for work on each problem, or if the size of group were increased? To answer these questions fully would require, of course, additional careful experimentation. That any or all such possible changes in conditions would be expected to alter essentially the conclusion reached is by no means clear.

Would the inferiority of real to nominal group performance disappear with some other kind of problem? Even more important, would real-group performance actually be superior to that of nominal groups? An affirmative answer to the first question appears doubtful and to the second even more improbable. On the basis of rather extensive pretesting, the three problems employed were selected from about twenty to be as appropriate as possible for use with brainstorming and also for use with Yale students. These problems appear to be generally similar to those which brainstorming is ordinarily accustomed to attack. Moreover, the question is not whether following the basic rules of brainstorming would be more effective with some other kind of problems; in such a case, the performance of nominal as well as real groups would be improved. Rather, the question is whether with some other kind of problem group participation when using brainstorming will facilitate, and not inhibit, performance. If such facilitation should be obtained in a carefully controlled experiment, a new question would arise as to the kinds of problems with which group participation is facilitating and with which it is inhibiting. Finally, it should be pointed out that the three problems were rather heterogeneous in character, thus increasing the generality of the results. The Thumbs Problem is clearly of a different type from the other two; the Teachers Problem appears to depend more heavily upon

previously acquired, directly relevant knowledge than does the Tourists Problem.

More training and experience in following the basic rules of brainstorming than that received by the present subjects might well be expected to facilitate production of ideas. But such training and experience would be expected to improve the performance of nominal as well as of real groups. Indeed, it appears probable that with more highly trained subjects essentially the same conclusion would be reached as in the present study.

The time limit of twelve minutes for each problem was chosen, on the basis of considerable pretesting, as one which would permit group members to express all ideas occurring to them within the work period and at the same time not result in excessive periods of silence for individual subjects.[29] In the actual experimental sessions, appreciable periods of silence appeared between responses near the end of the twelve minutes. Doubling or tripling the time allowed would, of course, increase the number of ideas produced by real groups, but it would also increase the number produced by nominal groups. It is possible that allowing more time would reduce the degree of superiority of nominal over real groups. It appears improbable, however, that such superiority would be eliminated and even more improbable that with any reasonable increase in time real group production of ideas would actually become superior to that of nominal groups. It is worth noting that the number of ideas per problem produced by real groups within twelve minutes compares favorably with the number reported for groups working considerably longer on other problems.[30] The mean number per problem varied from 32.6 for the Teachers to 41.3 for the Thumbs Problem (Table 2). Some individual groups, of course, produced a considerably larger number of ideas on each of these problems.

Osborn suggests that the optimum size for a brainstorming

[29]Question has been raised concerning the fact that the four individuals in a nominal group were each alloted twelve minutes per problem for a total of forty-eight minutes, whereas the real group was only alloted twelve minutes. This statement, of course, misses the point. What is relevant is the fact that the number of man-minutes devoted to a problem by the four members of a real group was forty-eight, precisely the same as for the four members of a nominal group.

[30]Osborn, *Applied Imagination.*

group is between five and ten.[31] The choice of four as the group size in the present experiment was based upon the fact that only about ninety-six men with appropriate experience in small discussion groups were available, together with the fact that N in the statistical analysis of this type of experiment is not the number of individual subjects but the number of groups. Increasing the size of the real groups to five or larger would probably increase the number of ideas produced, but it is also probable that an equal, if indeed not greater, increase would occur in the number of ideas produced by nominal groups of equal size. Hence increasing the size of the group would not be expected to alter the conclusion reached here.

The burden of proof would appear to be upon those who believe that with some change or changes in conditions group participation when using brainstorming will be found not to inhibit creative thinking, but instead to facilitate it.

It may be appropriate to emphasize in closing that the present experiment includes no evaluation of the basic rules of brainstorming—only an examination of the effects of group participation when using brainstorming. Such an evaluation would require experiments designed to compare the creativity of individuals or groups following the four basic rules with creativity of individuals or groups following some other procedure.

[31] *Ibid.*, p. 87.