1992

The Effects of Type of Evaluation and Task on Individual and Group Brainstorming Performance

Jerry Gerasimoss Vasilias
Loyola University Chicago

The Effects of Type of Evaluation and Task On Individual and Group Brainstorming Performance

by

Jerry Gerasimoss Vasilias

A Thesis Submitted to the Faculty of the Graduate School of Loyola University Chicago in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

January

1992

ACKNOWLEDGEMENTS

The author wishes to express thanks and appreciation to the members of the thesis committee, R. Scott Tindale Ph.D., and Linda Heath Ph.D. for their guidance, patience and support. Special thanks go out to Dr. Tindale, who served as committee chair, who read and gave constructive feedback on countless drafts, who was available for improptu consultations, and who kept me on track.

I am also indebted to all the people who assisted in the pilot testing and data collection stages of this project, especially Paul Mullen and Linda Thomas, who gave selflessly of themselves and of their time, and who made me laugh.

Special thanks go to my family and my adopted sister Tina for believing in me. Lastly, I thank and am forever indebted to my pal Layla.

VITA

The author, Jerry Gerasimoss Vasilias, was born on August 3, 1964, in Cephalonia, Greece. He emigrated to the United States in 1970. Once in the United States, he completed his grammar school education at DeWitt Clinton Public School in Chicago. He entered Stephen Tynn Mather Public High School in 1978.

In August, 1982, Mr. Vasilias entered Loyola University Chicago. He graduated magna cum laude with Bachelor's Degrees in Psychology and Philosophy in January, 1987.

In 1988, he entered the Applied Social Psychology program at Loyola. There, he has been involved in various projects relating to group dynamics, diversity and quality of life issues, and attitude and behavior change interventions to safeguard against the threat of acquired immune deficiency syndrome.

Mr. Vasilias has worked as an Evaluation Coordinator / Research Assistant for the Department of Family Practice at Rush Presbyterian St. Luke's Hospital in 1989, as an Assistant Research Analyst at Leo Burnett Advertising Agency in 1990, and as a Biofeedback Therapist at Diamond Headache Clinic since 1986 where presently he is the Acting Director of Biofeedback Services.

TABLE OF CONTENTS

LIST OF TABLES

CONTENTS OF APPENDICES

CHAPTER I

INTRODUCTION

In the early 1950's Osborn claimed that with brainstorming rules "the average person can think up twice as many ideas when working with a group that when working alone (p. 229)." Traylor, Berry, and Block (1958) tested this supposition by comparing four member groups to subjects individually instructed to brainstorm. They antithetically concluded that "group participation, when brainstorming, inhibits creative thinking (p. 43)." In their study, Traylor et al. found that nominal groups (the averaged output of a group of subjects who worked individually) produced a greater number of ideas than did interacting groups, a finding which has since been well replicated (Diehl & Stroebe, 1991a, 1991b, 1987; Mullen, Johnson, & Salas, 1991; Bond & Van Leeuwen, 1991; Bouchard, Drauden & Barasaloux, 1974; Bouchard & Hare; 1970).

There have been several alternative explanations posed for why interacting groups produce less than nominal groups. First, many have hypothesized that there are social psychological inhibiting mechanisms that are engaged when in the presence of others, or when one is a member of a group (Mullen et al., 1991); for example, drive arousal (Geen & Bushman, 1987) self attention (Carver & Scheier, 1981), and

fear of evaluation (Diehl & Stroebe, 1987; Amabile, 1979; Collaros & Anderson, 1969). Secondly, procedural mechanisms such as production blocking have been seen as responsible (Stroebe & Diehl, 1991a, 1991b; Diehl & Stroebe, 1987). Lastly, economic mechanisms, have been seen as responsible for often there is an intentional withdrawal of effort or noninvolvement on a task; for example social loafing (Latane, Williams, & Harkins, 1979).

This proposal focused on two perspectives found within these explanatory mechanisms; social loafing and fear of evaluation (i.e., one of the explanations offered by the classic brainstorming literature). These two approaches were chosen, because a review of the literature revealed that they made divergent hypotheses and arrived at inconsistent findings regarding brainstorming performance, even though both used the same variables. This study was carried out to test whether this discrepancy was due to differences in how each paradigm defined, manipulated, and operationalized the shared variables evaluation, group, and task.

CHAPTER II

REVIEW OF THE LITERATURE

Classic Brainstorming Research

According to the classic brainstorming paradigm interacting groups produce less than nominal groups because individual group members are often afraid of being negatively evaluated and criticized by other group members (Diehl & Stroebe, 1987; Magnin & Harris, 1981; Cottrell, Wack, Sekarek, & Rittle, 1968). Thus, the fear of public mocking, criticism, and possibly humiliation, leads many group members to censor their ideas, and not produce as much output (Collaros & Anderson, 1969). Individuals and nominal groups escape the stifling effects of evaluation apprehension because they are not allowed to interact with other group members.

Osborn's brainstorming rules were forged for the purpose of dismantling this fear of evaluation in interacting groups. The rules include: prohibition of criticism, encouragement to be free wheeling -- to produce zany, unusual ideas, a call to produce as many ideas as possible, and to combine ideas to create new ones. Although these brainstorming techniques should diminish or alleviate the fear of criticism, fear of evaluation is believed to persist simply because people expect that they will be evaluated when they are with others (Mullen

et al., 1991; Diehl & Stroebe, 1987; Magnin & Harris, 1980).

There is much evidence to show that this fear of evaluation can be heightened and can have an effect on performance. Early support for the evaluation apprehension explanation came from Collaros and Anderson (1969) who found that interacting groups' productivity could be lessened significantly the more individual group members perceived other members to be experts. Besides offering less ideas, subjects who were made to believe that their groups consisted of at least one or more experts claimed that they: 1) felt reluctant in offering ideas for fear of criticism, 2) tended to censor their ideas, and 3) sensed disapproval from other group members. As such, group members felt inhibited by the presence of the presumed more knowledgeable members, and subsequently contributed fewer ideas to the brainstorming task.

Amabile (1979) was also able to show that threats of criticism and negative evaluation could also impact deleteriously an interacting groups' work on an art activity. Specifically, subjects that were told that the art projects they would be making would be critically appraised by graduate art students created designs that were judged to be significantly lower on creativity than designs of subjects in no evaluation control groups.

Though the work of Amabile (1979), and Collaros and Anderson (1969) provided evidence for the evaluation

apprehension hypothesis, both studies were limited by the fact that they only used interacting groups. Magnin and Harris (1980) heeded this criticism and included nominal groups as a second level of the groups factor. Likewise, the researchers experimented with different levels of evaluation to ascertain which types of evaluation threats would have the greatest effects on the different groups' productivity. Yet, there were no differences between immediate and delayed evaluation, nor between relevant and irrelevant evaluation, primarily because of methodological limitations that masked and suppressed any possible effects: low n (only 6 groups experimental cell), and ineffective experimental manipulations. As expected nominal group out performed real interacting groups.

Diehl and Stroebe (1987) reexamined the notion of fear of evaluation as an explanation for the disparity between interacting and nominal groups. In their study, experimental nominal group subjects were told either: 1) that they would be watched and judged by raters behind a one way mirror, 2) that they would be videotaped, or 3) nothing about evaluation. The results showed that there was a significant drop in productivity under the combined evaluation conditions. This finding supported the evaluation apprehension hypothesis that the threat of being judged has inhibiting results.

In a follow-up study Diehl and Stroebe had interacting groups and nominal groups exposed to different levels of

evaluation. It was hypothesized that the mean number of ideas produced on a brainstorming task for the experimental nominal groups would approximate the number of ideas generated by the interacting groups in the control evaluation condition. The results indicated a main effect for group (nominal groups outproduced interacting groups) and for evaluation (more ideas were produced in the low rather than the high evaluation conditions), but the interaction was not significant. The researchers stated that "although there is a tendency for the evaluation manipulation to have a greater effect on nominal groups this interaction did not approach an acceptable level of significance (p 505)." However, sample size was small -- only four groups per cell were used, so lack of power may have prevented the researchers from finding a significant interaction.

Social Loafing

Whereas the classic brainstorming paradigm views evaluation as decreasing productivity, the social loafing paradigm predicts that individual level evaluation will increase performance. Social loafing researchers have proposed that when participants work together on any task where their outputs are pooled, loafing is a likely consequence because individual outputs are "lost in the crowd," and participants realize that they can receive neither credit nor blame for their individual performance (Latane, Williams, & Harkins, 1979). Many researchers have asserted

that when participants are made to believe that their individual outputs can and will be evaluated, the loafing effect will be eliminated (Szymanski & Harkins, 1988, 1987; Harkins & Jackson, 1985; Harkins & Petty, 1982).

Harkins and Jackson (1985) hypothesized that the greater the identifiability of individuals working within groups, the more their performance will be enhanced, for identifiability connotes that evaluation will follow. In their experiment, Harkins and Jackson used a box that had four dividers (this provided for the identification of individual output), or no dividers (this provided pooled data). As predicted, the identifiable condition increased the number of ideas generated in a brainstorming task as compared to the pooled data condition. This experiment similar to Harkins and Petty's (1982), and Harkins' (1987), found that the actual presence of an external evaluator was not required to increase output. All that was needed was the belief that external evaluation was possible.

Szymanski and Harkins (1987) did use a physically present form of external evaluation, an experimenter, to increase identifiability. They told their group subjects that "to ensure confidentiality, on our research team only the experimenter will know how many uses each person generates... the experimenter will count the number of uses you produce and then add them to those generated by the other participants (p.893)." The analysis of the number of ideas generated for

the brainstorming task revealed that the potential for experimenter evaluation significantly increased group performance.

Having shown that evaluation can have an effect with maximizing tasks, Harkins and Szymanski (1988) attempted to find out how evaluation affects subjects working with optimizing tasks (Steiner, 1972). As such, subjects worked on a vigilance task which required them to press a button when they saw a dot flash on a T.V. monitor. External evaluation was achieved by telling subjects to select a code from a set of random numbers which was to be entered at the computer. Then, that code would be added automatically to the number of dots they detected. Thus, they were told, the experimenter and themselves would be the only ones who would know their code numbers and output. Subjects who heard this speech made significantly fewer errors than did control subjects who were told that they would be the only ones who would know their output.

Comparison and Contrast of the Two Paradigms.

The results from these two paradigms seem to lead to a contradictory conclusion: Evaluation both increases and decreases individuals' performance in groups. But how can this be? One possible explanation involves the basic differences between the two paradigms in terms of how each has operationalized and manipulated evaluation.

Social loafing researchers have interpreted and

operationalized evaluation as a performance enhancing device. Specifically, social loafing researchers predicted that if individuals are led to believe that their output will be identifiable and individually monitored they will not loaf. Harkins et al. (1980) wrote:

> The results (social loafing) are easily explained by a minimizing strategy where participants are motivated to work only as hard as necessary to gain credit for a good performance or to avoid blame for a bad one. When the experimenter was unable to monitor individual outputs directly, performers sloughed off (p.464).

All the social loafing papers reviewed here have defined and manipulated evaluation in what can be referred to as the "monitoring" sense of the word. The monitoring evaluation manipulation was instilled through several techniques. For instance, Harkins and Jackson (1985) used a box with either four dividers or no dividers. Szymanski and Harkins (1987, 1988) told their subjects that a select number of researchers would know the amount of work they produced. These strategies effectively curtailed participants from being passive or idle when working in groups.

Contrastingly, classic brainstorming researchers have construed and manipulated evaluation in the critical and judgmental sense of the word. Several evaluation apprehension researchers achieved this manipulation by telling subjects that expert judges were sitting behind one-way mirrors listening to the ideas they produced and appraising the worth

of these ideas (Magnin & Harris, 1980; Diehl & Stroebe, 1987). Thus, evaluation, for this paradigm, has been construed and operationalized in a manner geared towards appraising the worth or merit of the overall performance.

Besides defining and using evaluation differently, a close examination of the types of tasks used in the two paradigms reveals further differences in definition and usage. Though both paradigms claim to have used creativity tasks, each has used tasks that can be considered distinct and different from the other. Specifically, Szymanski and Harkins (1987), and Bartis, Szymanski, and Harkins (1988) asked subjects to ponder how many different uses they could come up for the common knife. This task does not seem to be very stimulating, challenging, or even thought-provoking; this type of task will here be referred to as a mundane idea generation task. The brainstorming research too has had subjects work on creativity tasks, and the tasks used by this paradigm could be said to inspire, more truly, creativity. For instance, some of the tasks used in this paradigm have been: what practical benefits or difficulties would arise if people had an extra thumb on each hand (Bouchard & Hare, 1970; Magnin & Harris, 1980); how can a person of average ability achieve fame and immortality though he does not possess any particular talent (Collaros & Anderson, 1969); how can the life quality be improved in the suburbs (Diehl & Stroebe, 1987). These type of tasks will be referred to here as an interesting idea

generation tasks. In short, the two paradigms did not give their participants the same type of tasks on which to work.

Lastly, the two strains of literature seem to have different interpretations of what constituted a group. Specifically, in Harkins and Szymanski's (1987; 1988) studies subjects were individually run, and the output of four subjects was summed to attain what was called "group" level data. If one accepts that individuals loaf less when they are alone, it seems that what was attained from these two studies should have been referred to as summed individual level data, or nominal groups data. Besides using only nominal groups data, recent social loafing research has been sparse in providing individual and interacting groups data in a single experiment. Providing appropriate and accurate comparative level data is an essential next step.

Purpose

The purpose of this study was to better understand the inconsistency between the social loafing and brainstorming literatures concerning evaluation, task and group. Because the review of the literature had suggested that the way these terms were operationalized could provide a reason for the inconsistent findings, each term has been manipulated once more in an attempt to tease out the effects due to each paradigm's specifications. The different evaluation conditions used were: 1) no evaluation, 2) monitoring in correspondence with the social loafing definition of the term

and 3) critical evaluation in correspondence with the classic brainstorming paradigm's definition of the word. Likewise, because it is believed that each model used different types of tasks, all participants worked on two tasks: 1) an interesting idea generation task and a 2) mundane idea generation task. Finally, interacting groups, nominal groups, and individuals were used to be able to assess the effects of the aforementioned variables on different group types. The dependent variables for this study were, 1) the number of non-overlapping ideas each individual or group generated, 2) and the creativity of these ideas.

Hypotheses

If the social loafing operationalization is indeed performance enhancing and the classic brainstorming operationalization is performance decreasing, then it was hypothesized that subjects in monitoring evaluation conditions performance would be better that subjects experiencing critical evaluation.

Likewise, it was expected that individuals and nominal groups would perform better than the interacting groups in the no-evaluation control conditions. In fact, individuals and nominal groups performance was expected to vary depending on the evaluation conditions they experienced. Specifically, individuals and nominal groups were expected to have the best performance showings during the monitoring conditions, and their worst during the critical evaluation condition.

Interacting groups performance, on the other hand, was expected to be at a consistent low rate across conditions because of the inhibitory forces and uneasiness active in such groups where individuals are asked to interact and work with others they just met.

A three way interaction was also expected: the critical evaluation manipulation was expected to affect individuals and nominal groups most severely on their work with the interesting idea generation task.

CHAPTER III

METHOD

Subjects

Two hundred seventy undergraduate Loyola psychology students volunteered as participants in this study. All subjects received research credit for their participation. Upon arriving to the site of the experiment, subjects were randomly assigned to the different experimental conditions.

Design and Materials

Design. This was a 3 x 3 x 2 factorial design. Each subject experienced one of three levels of evaluation (no-evaluation, monitoring, or critical evaluation), and of group (individuals, nominal or interacting groups) and worked on two tasks (an interesting idea generation task and mundane idea generation task).

Tasks. The two tasks used were: How can a person of average ability achieve fame and immortality though he/she does not possess any particular talents; generate as many ideas as you possibly can -- the interesting idea generation task, and, generate as many uses as you possibly can for the common knife -- the mundane idea generation task. Order of appearance for the two tasks was counterbalanced to avoid any effects due to order.

Evaluation. To operationalize the different evaluation

conditions a 2" (length) x 1" (width) x 3" (height) box was used. Subjects in the no-evaluation condition had this box appear in front of them filled with several folded sheets of paper. Once given the tasks, they were instructed to write down each idea they generated, fold it several times, and deposit it in the container with the other folded sheets. To promote no-evaluation/no identifiability the researcher collected the ideas generated for each task by putting all the ideas in the box into a common envelope.

To induce the monitoring evaluation condition the experimenter gave each of the deliberating subjects an empty box, same as the one described earlier, to fill with their ideas. To further enhance identifiability, after each task, the researcher collected each person's data and put it in a separate envelope earmarked with that person's experimentally assigned code number.

To induce the critical evaluation condition the experimenter told subjects that "psychologists who study creativity in these type of settings were going to critically evaluate and appraise the worth of the ideas they produce." And as was done with subjects in the monitoring evaluation condition, each subject experiencing this evaluation condition filled their own individual box, and the output they generated was collected individually and their experimental code number was put on the front of the envelope.

Group. Subjects worked on the two tasks either

individually or in three member groups. There were two varieties of three member groups: nominal and interacting groups. Subjects in nominal groups sat at the same table and worked on the tasks but were told they could not communicate to one another. Subjects in interacting groups were first asked to introduce themselves to each other. Then, during the brainstorming task, they were asked to verbalize each idea they came up with before writing it down and depositing it in their box. Interactiveness was defined in this way to allow inter-member evaluation to occur while minimizing the potential blocking effects present in normal group interaction (Diehl & Stroebe, 1991a, 1991b, 1987).

Brainstorming Rules. While working on the tasks, subjects were: instructed that criticism was prohibited, encouraged to create unusual ideas, to combine ideas whenever possible -- one's own if working in a non-interactive mode, to produce as many ideas as possible. Appendix A contains the brainstorming rules, and all other instruments used in this study.

Upon finishing with the tasks, subjects were asked to fill out a brief two page questionnaire containing an anxiety measure and three manipulation check questions.

State Trait Anxiety Index. This self-report anxiety measure has two sections: a trait and a state measurement of anxiety. Only Form Y-1, the state part of the State Trait Anxiety Index (STAI) (Spielberger, Gorsuch, Lushene, Vagg, &

Jacobs, 1983), was used for of interest was people's perceptions of presently felt state anxiety resulting from the evaluation condition experienced.

Next, because there were differences expected between the two tasks, subjects were asked which task they found to be more challenging, and secondly, which they thought was more interesting to work on.

Lastly, subjects were asked to assess the effectiveness of the identifiability component of the evaluation manipulation by indicating how easily they thought the experimenter could identify their work from that of other subjects by placing a slash, "/", through a line that had "1 -- Can easily identify" and "7 -- Cannot easily identify" as anchorpoints.

Procedures

Upon the subjects' arrival, the experimenter randomly chose to run a specific evaluation and group condition. Subjects were next seated and told that the purpose of this experiment was to assess how brainstorming techniques affects productivity. The experimenter then handed out and read aloud a sheet delineating the brainstorming rules that were to be used. After, the first task was handed out. Subjects were told to keep in mind and use the brainstorming rules just discussed with this first task. They had 10 minutes to work on the task. Once completed the experimenter collected the data that was generated according to the appropriate means,

and read the brainstorming rules aloud a second time and then distributed the second task. Subjects were again asked to work on this task for 10 minutes while keeping the just read brainstorming rules in mind. Upon completion and collection of this task subjects were asked to fill out a short questionnaire complete with the manipulation checks and the STAI anxiety measure. After, subjects were debriefed verbally and given a written summary of the project. Lastly, they were asked to please refrain from talking about this study with other students.

CHAPTER IV

RESULTS

Manipulation Checks. A 3 x 3 analysis of variance of the effectiveness of the identifiability component of the evaluation manipulation revealed there to be a main effect for evaluation ($F(2, 233) = 14.1$, $p < .000$). A Student Newman-Keuls (SNK) showed that subjects in the no evaluation conditions were less likely to believe that the experimenter could identify their work ($M = 2.83$) than subjects in the monitoring ($M = 1.52$) or the critical evaluation conditions ($M = 1.82$). There were no statistically reliable mean differences between the monitoring and the critical evaluation conditions.

Next, the manipulation concerning subjects perceptions of differences between the tasks was assessed. As expected, subjects thought the immortality task was more interesting (X^2 (1, N = 270) = 4.8, $p = .028$), and more challenging to work on (X^2 (1, N = 270) = 102, $p < .000$), than the knife task. Table 1 contains the exact number of subjects that found each task challenging and interesting.

Table 1

Summary Table of the Number of Subjects that Found Each Task Interesting and Challenging.

	Task 1 "Knife Task"	Task 2 "Immortality Task"
Which task was more challenging to work on?	52_a	218_b
Which task was more interesting to work on?	117_a	153_b

Note: Different subscript letters indicate a significant difference ($p < .05$) between the two tasks on the different dimensions.

Lastly, subjects' STAI scores were analyzed with a 3 x 3 analysis of variance to ascertain if there were mean differences in anxiety due to type of evaluation and group condition experienced. An analysis of variance on the STAI scores revealed a main effect for evaluation ($F(2, 256) = 5.4$, $p < .005$), but, the pattern of the means attained was not in the expected direction. It was found that subjects in the critical evaluation conditions had the highest mean score overall ($M = 39.5$) and were slightly more relaxed than subjects experiencing no-evaluation ($M = 36.5$) and significantly more at ease than participants in the monitoring group conditions ($M = 34$). There was no significant group, or group by evaluation condition effect.

Assessing Quantitative Performance. To determine consistently the number of ideas each individual and/or group generated per task, rules were developed to guide the tallying process. First, based on past research, only non-repetitive, non-redundant, and non-overlapping ideas were counted. Thus, if an individual responded that a common knife can be "used to throw at a target" and "as a dart," only the first suggestion was counted because both ideas are essentially the same. Second, suggestions could share a common verb and not be considered repetitive or redundant, only if the verb was used toward creating suggestions with different purposes or ends; so, if "cutting vegetables" and "cutting one's wrists" were offered by an individual both ideas would be accepted for the

aim or end or each is entirely different and distinct. Third, if umbrella, general, and vague ideas were offered along with explicit, definite, and specific ones, the broad suggestions were omitted from the count. However, if they were the only ones tendered they were counted. Specifically, if for the immortality task, an individual wrote "to break a world record" and then proceeded to elaborate on ways this could be done, the aforementioned broad idea was not included, but the specifics that flowed from the enumeration process were. Lastly, any ideas that were judged to be irrelevant or illegible ideas were eliminated. Irrelevant ideas were those that failed to abide the instructions given at the beginning of each task. For instance, for the knife task some subjects offered mistakenly descriptors like "shiny," "sharp," "pointy," instead of uses, and these ideas were excluded from the count.

To assess the reliability of these guidelines the researcher and a research assistant tallied separately the number of ideas a sample of 24 subjects generated and correlated their scores. A correlation of .98 confirmed that the rules were highly reliable and were used without modification on the remaining participants' output.

After every participants' output had been quantified, mean individual member performance scores for the nominal and interacting groups were calculated. The mean was taken for nominal and interacting groups because prior research had revealed that when comparing the work of individuals who did not participate in group interaction with individuals who had, interacting group members' work tended to be highly correlated (Myers, DiCecco, & Lorch, 1981).

A 3 (type of evaluation) x 3 (type of group) x 2 (type of task) repeated measures analysis using the BMDP.2V statistical package was then done on the quantitative estimates. The omnibus repeated measures analysis of variance appears in Appendix B.

The analysis revealed there to be a significant main effect for evaluation ($F(2, 141) = 7.87$, $p = .0006$). Post hoc SNK analysis comparing each subjects averaged performance across the two tasks -- an average was taken because the task by evaluation interaction effects was not significant, revealed that under monitoring evaluation conditions subjects generated significantly more ideas than did subjects in the no evaluation conditions or in the critical evaluation conditions. The means and standard deviations for each type of evaluation appear in Table 2.

Table 2

Summary Table of Mean and Standard Deviations for Each Evaluation Condition Across Task for the Quantitative Estimate.

Evaluation		
No Evaluation	Monitoring Evaluation	Critical Evaluation
M	M	M
12.8_a (4.2)	14.5_b (3.9)	11.6_c (4.4)

Note: Different subscript letters indicate a significant difference ($p < .05$) between the different evaluation condition means. Subjects experiencing critical evaluation generated the least number of ideas overall.

Likewise, the analysis revealed a main effect for task ($F(1, 141) = 70.76$, p .0000). As expected, subjects generated a higher average number of ideas for the knife task ($M = 14.9$, $SD = 5.5$), than for the immortality task ($M = 11$, $SD = 4.4$).

Also, there was a marginally significant effect for group ($F(2, 141) = 2.14$, p .12). A posteriori SNK tests revealed that though nominal groups generated more ideas ($M = 13.8$) than did the interacting groups ($M = 11.7$), this was not statistically significant. Table 3 contains the means and standard deviations for each group.

Table 3

Summary Table of Mean and Standard Deviations for Each Group Condition Across Task for the Quantitative Estimate.

Group		
Individuals	Nominal Groups	Interacting Groups
M	M	M
13.1 (4.4)	13.8 (3.8)	11.6 (4.4)

The group by evaluation, and the group by evaluation by task interactions were not significant.

Assessing Qualitative Performance. The qualitative dependent variable, creativity of ideas, was operationalized in a quantitative manner to be able to derive a creativity score for each individual and/or group. First, all the aforementioned non-redundant, non-overlapping ideas were tallied and a frequency distribution was developed so that the number of times an idea was offered determined whether it was to be considered more or less a creative idea. Thus, the more frequent ideas were defined as less creative. There were 466 different ideas, with frequencies ranging from 1 to 169, generated for the knife task, and 864 distinct ideas were produced for the fame and immortality task, with frequencies ranging from 1-69. Next, to eliminate the skewed distributions associated with each task, the frequency counts were transformed to natural logarithms. Thus, the log transformed scores were summed and divided by the number of ideas presented by each individual/ group to control for differences in the number of ideas generated. Finally, a 3 x 3 x 2 repeated measures analysis of variance using the BMDP.2V statistical package was done. The complete analysis of variance for this dependent variable appears in Appendix B.

The analysis revealed a main effect for task ($F(1, 141) = 315$, $p = .0000$), with subjects' ideas being rated

significantly more creative and unique for the fame and immortality task (*M* = 2.25, *SD* =.5) than for the knife task (*M* = 3.41, *SD* =.72).

Likewise there was a main effect for evaluation (*F*(2, 141) = 5.78, *p* = .004). Comparisons using SNK on the averaged performance across both tasks showed that subjects in the no evaluation (*M* = 2.8) and monitoring evaluation conditions (*M* = 2.7) produced ideas that were considered significantly more creative than those generated by subjects in critical evaluation conditions (*M* = 3.03). The means and standard deviations associated with each evaluation condition are in Table 4.

Table 4

Summary Table of Mean and Standard Deviations for Each Evaluation Condition Across Tasks for the Qualitative Estimate.

Evaluation		
No Evaluation	Monitoring Evaluation	Critical Evaluation
M	M	M
2.8_a (.46)	2.7_a (.48)	3.03_b (.49)

Note: Different subscript letters indicate a significant difference ($p < .05$) between the different evaluation condition means. Subjects in the critical conditions produced ideas considered less creative those of subjects in the other evaluation conditions.

Lastly, there was a main effect for group ($F(2, 141) = 3.26$, $p = .04$). The SNK test revealed that nominal groups' averaged performance across both tasks ($M = 2.8$) was not statistically different from that of non-grouped individuals ($M = 2.9$), yet, it was significantly different from that of the interacting groups' ($M = 2.65$). According to these results, interacting groups had generated, unexpectedly, the most creative ideas. Table 5 contains the different means and standard deviations associated with the different group conditions.

Table 5

Summary Table of Mean and Standard Deviations for Each Group Condition Across Tasks for the Qualitative Estimate.

Group		
Individual	Nominal Groups	Interacting Groups
M	M	M
2.9_a (.54)	2.8_a (.34)	2.65_b (.43)

Note: Different subscript letters indicate a significant difference ($p < .05$) between the different group means.

Similar to the quantitative variable, although there were were higher order interactions expected, none reached significance.

CHAPTER VI

Discussion

The findings from this experiment provided some support for the premise that the inconsistent findings between the classic brainstorming and social loafing paradigms regarding evaluation and the other variables was due to the divergent operational definitions of the terms. Results showed that subjects performed better under the social loafing paradigm's monitoring operationalization of evaluation as opposed to the classic brainstorming paradigm's inhibition inducing manipulation. Specifically, when subjects were made to feel that they would be evaluated, in the sense that their work would be identifiable over that of others and that they would be held accountable for it, they generated more ideas than when they were told this and that the worth of the ideas they generated would be critically appraised.

Results concerning the quality of performance also supported this hypothesis. By using the newly developed objective measuring device to assess the quality of ideas, it was found that the type of evaluation people experienced impacted on the creativity and originality of ideas they offered. As expected, more inventive and imaginative ideas

were produced under no evaluation or monitoring evaluation conditions as opposed to critical conditions. Thus, the present results showed that an environment that fostered fear and apprehension was deleterious to the quantity and quality of people's work.

The findings did reveal a difference between the tasks. The immortality task, a prototype for the type of task used by the classic brainstorming paradigm, was thought to be more interesting and more challenging to work on than the knife task, the social loafing prototype. Likewise, subjects produced fewer ideas for it than they did for the knife task. With regards to the qualitative estimate, the ideas generated for the immortality task were more creative than those generated for the knife task.

Looking at the main effect for group across evaluation it was found that the means for the quantitative variable were in the desired direction with nominal groups outperforming interacting groups.

The findings for the qualitative estimate revealed that there were mean differences for the different group conditions, but the direction of the results was unexpected. Specifically, and antithetically, the creativity mean scores for interacting groups were lowest overall indicating that they had produced the most creative ideas. Though these findings are in unison with researchers like Graham (1977) who argue that decrements in brainstorming groups' performance are

often compensated by increases in the quality of the ideas generated, empirical support has been remote for this supposition. According to Diehl and Stroebe (1987, 1991a, 1991b) there is another explanation for these unexpected results.

> The findings for quality of ideas appear to be heavily dependent on the type of measure used: In all studies that assessed total quality (the sum of the quality ratings of the ideas produced by a given subject or group) nominal groups performed better than did real groups did. No consistent pattern emerged for the other measures. Among those studies, findings were not only inconsistent between studies but even within the same study, if several topics, subject groups, or experimental conditions had been used (p 497, 1987).

Hence, the results attained might be exclusive to the measuring technique used. Clearly, the logical next step research-wise would be to assess quality using several methods including the objective device used here to assess directly whether quality will vary depending on how output will be measured.

Besides the unexpected group effect findings, there were several hypotheses that were not supported by the data, but had they been that would have made a stronger case for the operational definition hypothesis argued for. For instance, though it was expected that nominal groups would outperform, qualitatively and quantitatively, interacting groups in the no-evaluation conditions, we found no evidence of this.

Lastly, though it was hypothesized that subjects performance would vary depending on what type of task they

were working on, and the type of evaluation and group condition they were experiencing, this three way interaction was not significant.

Though not all the hypotheses made were supported by the findings, this study was successful for several reasons. First, by examining the variables it did, this project has come to a better understanding of what can be, and what is not responsible for interacting groups' typically lackluster output. Specifically, although evaluation was found to have strong effects on performance, it did not interact with the group variable, and as such it could not adequately explain the differences between nominal and interacting groups. Thus, the most viable and plausible explanation for why productivity losses exist in interacting groups is not evaluation, but what Diehl and Stroebe have dubbed, production blocking (1991a, 1991b, 1987).

Secondly, the results provided some strong evidence that the inconsistency between the two social psychological paradigms was due to some degree to opposing definitions and administrations of commonly shared terms. So, this project has brought a heightened awareness of the semantical differences and nuances associated with commonly used words in psychology. Thus, when using a popularly cited and used psychological terms like evaluation, group, or task, one cannot expect unilateral understanding and consensus on that term's meaning by those using it; evaluation, task, and group

will mean different things to the different researchers studying and using the terms.

However, when future research in this area is done, there are things that should be done differently. First, though this project's power was substantial over that of those reviewed in this paper, future work in this area can begin with a power analysis to ascertain the exact number of people that will be needed to find a moderate effects for all variables being manipulated.

Another thing that could be done differently with future research in this area is the way the interacting groups' condition is to be manipulated, for it is plausible that the mediocre quantitative and unexpected qualitative group results were due to a lacking interacting group manipulation. Specifically, during such conditions subjects were not interacting, so much as they were vocalizing and blurting out ideas. Perhaps future research can experiment with creating a more truly interactive group condition where subjects will be required to communicate to one another to be able to complete a given task. At the same time, researchers will need to develop safeguards to control against the threat of production blocking from being active, as was done in the present case.

Likewise, different response modes need to be manipulated and assessed. Though this experiment had subjects vocalize and then write down their ideas many researchers in this area

forgo the written modality and have their subjects tape recorded. Though there has been research suggesting that productivity losses are heavier under taped recorded sessions (Mullen et al., 1991), there has also been research suggesting that response mode does not make a difference (Bond & Van Leeuwen, 1991). It seems that more research needs to be done with this variable especially since many who use the brainstorming techniques, like advertising companies running focus groups, are using audio as well as video recorders.

Yet another variable that can be examined is whether the presence of the experimenter has any effect on performance. Mullen et al (1991) found that there were heavier productivity losses when researchers were present. In the present study, the experimenter was present in every evaluation and group condition. Future experimentation could assess whether the presence or absence of the experimenter in certain evaluation conditions impacts the performance of subjects in any way.

Lastly, the empirical research done in this area has used undergraduates to arrive at the findings attained, yet brainstorming techniques are used predominately by non-undergraduates, specifically, by those in industrial and business settings, i.e. advertising. As such, future research should consider using subjects from within the very arenas that seem to be using the brainstorming groups and techniques the most to confirm or disconfirm what has been established with student samples.

References

Amabile, T. (1979). Effects of external evaluation on artistic creativity. Journal of Personality and Social Psychology. 37 (2), 221-233.

Bartis, S., Szymanski, K., & Harkins, G. S. (1988). Evaluation and performance: a two edged knife. Personality and Social Psychology Bulletin. 14, (2), 242-251.

Bond, F. C. & Van Leeuwen, M. (1991). Can a part be greater than the whole? On the relationship between primary and meta-analytic evidence. Basic and Applied Psychology. 12, (1), 33-40.

Bouchard, T. J. (1969). A comparison of two group brainstorming procedures. Journal of Applied Psychology. 56, 418-421.

Bouchard, T. J., Drauden, G., & Barasaloux, J. (1974). A comparison of individual, subgroup, and total group methods for problem solving. Journal of Applied Psychology. 59 (2) 26-227.

Bouchard, T. J., & Hare, M. (1970). Size, performance, and potential in brainstorming groups. Journal of Applied Psychology. 54 (1), 51-55.

Carver,C. S., & Scheier, M. F. (1981). The self-attention induced feedback loop and social facilitation. Journal of Experimental Social Psychology. 17, 545-568

Collaros, P., & Anderson, R. L. (1969). Effects of perceived

expertness upon the creativity of members of brainstorming groups. Journal of Applied Psychology. 53 (3), 153-163.

Cottrell, N. B., Wack, D.L., Sekarak, G. J., & Rittle, R. M. (1968). Social facilitation of dominant responses by the presence of others. Journal of Personality and Social Psychology. 9, 245-250.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: toward the solution of a riddle. Journal of Personality and Social Psychology. 53 (3), 497-509.

Diehl, M., & Stroebe, W. (1991). Productivity loss in idea-generating groups: Tracking down the blocking effect. Journal of Personality and Social Psychology. 61 (3), 392-403.

Geen, R. G., & Bushman, B. J. (1987). Drive Theory: Effects of socially engendered arousal. In B. Mullen & G. R, Goethals (Eds.) Theories of Group Behavior (p 89-109). New York: Speinger-Verlag.

Graham, W.K. (1977). Acceptance of ideas generated through individual and group brainstorming. Journal of Social Psychology. 101, 231-234.

Harkins S. (1987). Social loafing and social facilitation. Journal of Experimental Psychology. 23, 1-18.

Harkins, S., & Szymanski, K. (1989). Social loafing and group evaluation. Journal of Personality and Social Psychology.

56 (6), 934-941.

Harkins, S., & Szymanski, K. (1988). Social loafing and self-evaluation with an objective standard. Journal of Experimental Social Psychology. 24, 354-365.

Harkins, S., & Jackson, M. J. (1985). The role of evaluation in eliminating social loafing. Personality and Social Psychology Bulletin. 11 (4), 457-465.

Harkins, S., & Petty, E. R. (1982). Effects of task difficulty and task uniqueness on social loafing. Journal of Personality and Social Psychology. 43 (6) 1214-1229.

Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: the causes and consequences of social loafing. Journal of Personality and Social Psychology. 37, 822-832.

Magnin, K. B., & Harris, J. R. (1980). Effects of anticipated evaluation on individual brainstorming performance. Journal of Applied Psychology. 65 (2), 219-225.

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A Meta-analytic integration. Basic and Applied Social Psychology. 12 (1), 3-23.

Myers, L. J., DiCecco, V. J., & Lorch, F. R. (1981). Group dynamics and individual performances: Pseudo-group and quasi-f analyses. Journal of Personality and Social Psychology. 40 (1), 86-98.

Spielberger, D. C., Gorsuch, L. R., Lushene, R., Vagg, R. P., & Jacobs, A. G. (1983). State Trait Anxiety Inventory

Manual. Palo Alto: Consulting Psychologists Press.

Szymanski, K., & Harkins, S. (1987). Social loafing and self-evaluation with a social standard. Journal of Personality and Social Psychology. 53 (5), 891-897.

Steiner, T. (1972). Group process and productivity. New York: Academic Press.

Stroebe, W., Diehl, M. (1991). You can't beat good experiments with correlational evidence: Mullen, Johnson, and Salas's meta-analytic misinterpretations. Basic and Applied Social Psychology. 12 (1), 25-32.

Traylor, D. W., Berry, P.C., & Block, C. H. (1958). Does group participation when brainstorming facilitate or inhibit creative thinking? Administrative Science Quarterly. 3, 23-47.

APPENDIX A

EXPERIMENT # 34

Thank you for volunteering to participate in our research project. The purpose of this experiment is to assess how brainstorming techniques affect productivity.

Please know that all of the information that we collect today is confidential. This means that it will be seen only by qualified researchers and will be used for research purposes only.

Finally, should you decide at any point to discontinue your participation on our project, for whatever reason, please feel free to do so. Though we do not expect that this will happen, we want you to know that you are free to leave the study at any point without incurring any penalty.

Please feel free to ask any questions. Once again, thank you for participating on our project.

Sincerely,

Jerry Vasilias

I have read the above and understand it.

_____ _____

Signature Date

_____

Instructor

^^

Generate as many uses as you possibly can for the common knife.

^^

^^

How can a person of average ability achieve fame and immortality, though he/she does not possess any particular talent?

^^

^^

BRAINSTORMING RULES

^^

Brainstorming rules are often used when new, unique, original, or creative ideas are desired. Please read the four brainstorming rules below. You will be asked to keep these rules in mind when working on two tasks.

1. Criticism is ruled out. Adverse judgement of ideas must be withheld. No one in an interacting group should criticize anyone else's ideas. For individuals in the non-interacting conditions, we ask that you do not criticize any idea that comes to you. Instead we ask that you write down every idea that you think of.

2. Freewheeling is welcome. The wilder the idea, the better. Do not be afraid to write down anything that comes to mind; the "farther - out" the idea, the better.

3. Quantity is what is of interest. Thus, try to come up with as many ideas as you can.

4. Combine and improve on already written ideas. Individuals within interacting groups should try to improve and build upon other people's ideas. Individuals working alone or in non-interacting conditions should try to change and modify suggestions they have made. In short, do not be afraid to combine and improve on already stated ideas.

#_____

SELF EVALUATION QUESTIONNAIRE

A number of statements which people have used to describe themselves are given below. Read each statement and then indicate how you feel **right now**, that is, **at this moment**, by circling one of the numbers following each statement.

	Not at all 1	Somewhat 2	Moderately so 3	Very much so 4
1.	I feel calm	1	2	3	4
2.	I feel secure	1	2	3	4
3.	I am tense	1	2	3	4
4.	I feel strained	1	2	3	4
5.	I feel at ease	1	2	3	4
6.	I feel upset	1	2	3	4
7.	I am presently worrying over possible misfortunes	1	2	3	4
8.	I am satisfied	1	2	3	4
9.	I feel frightened	1	2	3	4
10.	I feel comfortable	1	2	3	4
11.	I feel self-confident	1	2	3	4
12.	I feel nervous	1	2	3	4
13.	I feel jittery	1	2	3	4
14.	I feel indecisive	1	2	3	4
15.	I am relaxed	1	2	3	4
16.	I feel content	1	2	3	4
17.	I am worried	1	2	3	4
18.	I feel confused	1	2	3	4
19.	I feel steady	1	2	3	4
20.	I feel pleasant	1	2	3	4

1. Which of the two tasks did you find more interesting to work on?

_____ KNIFE TASK _____ ACHIEVING GREATNESS TASK

2. Which of the two tasks would you say was more challenging to work on?

_____ KNIFE TASK _____ ACHIEVING GREATNESS TASK

3. Indicate to what degree you believe the experimenter can identify the work that you individually produced from that of anyone else by placing a " / " through the line below. For instance, if you feel the experimenter can easily identify your work from someone elses the " / " should be placed towards the left end of the scale. The more you feel this to be true the closer your " / " should appear towards the left. If you feel the researcher cannot easily identify your work the " / " should appear toward the right end of the scale. Again, the stronger you feel this to be true the closer your " / " should appear towards the right end of the scale.

---

1 — Can easily identify my work

7 — Cannot easily identify my work

Some Information About the Study

Within groups research there is a much replicated finding: when individuals work on a task as a group, they not to generate as much total output. There are two prominent and contrasting explanations for this phenomenon: 1). social loafing (SL) (Latane, Williams, & Harkins, 1979) and 2). evaluation apprehension as proposed by the classic brainstorming research (EA) (Magnin & Harris, 1981). SL theory hypothesizes that when working in groups, individuals can easily loaf (and do) because the researcher can not detect what each individual member generates. The EA literature holds that groups produce less because individual members within the groups are afraid of being criticized by other group members, so they censor their ideas, and produce less.

After a careful review of both paradigms it was noticed that each approach had: 1). defined and manipulated the term evaluation differently from the other (EA has used a critical definition SL has used a monitoring definition); 2). used different types of tasks (EA has used creativity tasks SL has used tasks that can be called "pseudo-creativity"; 3). different interpretations of what constituted group level data (SL summed individual level data to attain groups level data, EA used interacting groups data).

The purpose of this study you just participated in was to observe if and how the different uses of evaluation, tasks, and groups, affects brainstorming performance. Some of you were in groups while others worked alone on the different tasks. Likewise, some people experienced no evaluation, some experienced a monitoring evaluation condition, others experienced a critical evaluation condition.

If you have any further questions about the study, please feel free to leave a message at the psychology department for Jerry Vasilias. If you would like more information about these areas of research, the references listed below would be a good place to start.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: toward the solution of a riddle. Journal of Personality and Social Psychology. 53 (3), 153-163.

Szymanski, K., & Harkins, S. (1987). Social loafing and self-evaluation with a social standard. Journal of Personality and Social Psychology. 53 (5) 891-897.

APPENDIX B

Omnibus Repeated Measures Analysis of Variance for the Quantity of Ideas Generated.

Source	SS	DF	MS	F	PROB
EVAL	532	2	266	7.9	.0006
GROUP	144	2	72	2.1	.1212
EG	184	4	46	1.4	.2495
1 ERROR	4766	141	34		
TASK	859	1	859	70.8	.0000
TE	8	2	4	.3	.7204
TG	24	2	12	1.0	.3734
TEG	47	4	11.7	1.0	.4262
2 ERROR	1712	141	12		

Omnibus Repeated Measures Analysis of Variance for the Quality of Ideas Generated.

Source	SS	DF	MS	F	PROB
EVAL	5	2	2.6	5.8	.0039
GROUP	2.8	2	1.4	3.3	.0414
EG	1	4	0.25	0.6	.6830
1 ERROR	62	141	0.44		
TASK	81	1	81	315.3	.0000
TE	0.5	2	0.3	1.0	.3610
TG	0.3	2	0.2	0.6	.5451
TEG	1.5	4	0.4	1.5	.2148
2 ERROR	36.4	141	0.3		

Summary Table of Mean and Standard Deviations for Each Evaluation Condition for Each Task for Each Dependent Variable.

	Evaluation		
	No Evaluation	Monitoring Evaluation	Critical Evaluation
	M	M	M
Quantity			
Task 1	15.1 (5.7)	16.1 (5.2)	13.1 (4.6)
Task 2	10.9 (4.1)	12.7 (4.5)	9.4 (4)
Quality			
Task 1	3.4 (.47)	3.3 (.5)	3.6 (.46)
Task 2	2.1 (.64)	2.2 (.75)	2.5 (.69)

Summary Table of Mean and Standard Deviations for Each Group Condition for Each Task for Each Dependent Variable.

	Group		
	Individuals	Nominal Groups	Interacting Groups
Quantity			
Task 1	15.1 (5.7)	16.2 (5.2)	13.1 (4.6)
Task 2	11.1 (4.5)	11.5 (3.2)	10.3 (5.2)
Quality			
Task 1	3.46 (.55)	3.38 (.37)	3.30 (.4)
Task 2	2.34 (.82)	2.24 (.4)	2.01 (.57)

APPROVAL SHEET

The thesis submitted by Jerry Gerasimoss Vasilias has been read and approved by the following committee:

Dr. R. Scott Tindale, Director
Associate Professor,
Department of Psychology
Loyola University Chicago.

Dr. Linda Heath
Professor,
Department of Psychology
Loyola University Chicago.

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirement for the degree of the master of arts.

12/12/91
Date

R. Scott Tindale
Director's Signature