December 1998

In GSS Research How Many Groups per Treatment Condition Are Enough?

Jerry Fjermestad
New Jersey Institute of Technology

In GSS Research How Many Groups per Treatment Condition Are Enough?

Jerry Fjermestad
School of Management and New Jersey Center for Multimedia Research
New Jersey Institute of Technology

Abstract

This study presents, summarizes, and analyzes the results of 178 Groups Support Systems (GSS) experiments based on categories of groups per treatment condition. The results suggest that for studies designed with less than 7 groups per condition there are significantly lower probabilities of finding significant differences. Task type and GSS type further moderate these results.

Introduction

We study groups because in general, groups are better than individuals at understanding the problem, they make better decisions than individuals, they are better at catching errors, and so on. However, groups are not flawless, groups may make riskier decisions than need be, social pressures may force the group into groupthink (Janis, 1972), and groups may require a very long time to develop a solution.

These and other process gains and losses are of great interest to organizations. Obviously, if we can minimize the losses, greater productivity can be achieved. But, how do we study these groups? We can use case studies, laboratory studies, surveys, and field studies (Benbasat, 1989; Cash, 1989). Zmud, Olsen, and Hauser, (1989) suggest that case studies are most useful for theory building and that controlled experiments both in the laboratory and field for theory testing. In this regard, DeSanctis (1989) suggests that the design and use of systems that support small groups might best be studied under laboratory conditions if the technology is novel, where large numbers of groups are required for statistical analysis or where the uniqueness of naturally occurring groups precludes replication.

DeSanctis (1989) also suggests that there several difficulties in studying human groups:

- Approaches that groups use to solve problems vary enormously.
- It is exceeding difficult to evaluate the process and outcomes of information exchange in groups.
- To apply statistical inference methods to empirical data where the group is the unit of analysis requires an extremely large sample sizes to detect significant effects.

DeSanctis goes on to say that if 80 groups are needed for sufficient power in a statistical test, then 400 subjects must be recruited for the study if each group is have an average of five people in it.

Furthermore, Braoudl and Orlikowski (1989) report that statistical power is important to IS researches who are conducting statistical inference testing. Power is the probability that a statistical test will correctly reject a false null hypothesis. The results of their study show that on the average, MIS research is substantially low in term of statistical power. This is something that can be control so as to ensure that the statistical tests have sufficient power to detect the phenomena under study. Baroudl and Orlikowski (1989) recommend several ways to increase statistical power, one of which is to increase the sample size- groups per treatment condition in GSS research.

Research Problem

Given that the group is the unit of analysis (DeSanctis, 1989; Zigurs, 1993) in GSS research, have GSS researchers used enough groups per treatment condition to observe the effects that they are looking for? From an analysis of the empirical literature what range of groups per treatment condition lead to the highest probability of success? Does task type and GSS type (GSS or CMC) moderate success?

Method

Fjermestad and Hiltz (1997) presented a four factor framework consisting of contextual, intervening, adaptation, and outcome factors and mapped 140 studies to it. Since then, an additional 38 studies have been added making a total of 178 experiments. A database was created (Fjermestad and Seah (1997) which consists of 1391 independent/dependent variable crosses. For the purposes of this paper the following elements of that database will be used:

- Groups per treatment condition: The number of groups in an experimental condition 5 categories: 3 groups or less (1, 2, 2-3, 3, and experiments that had at least one condition with 3 or less, i.e. 3,4,4,5); 4 to 6 groups (4, 5, 6, 4-5, 5-6, etc.); 7 to 10 groups (7, 8, 9, 7-10, etc.); 11 groups and up (11, 12, 15, 36, etc); and not reported.
- Effect: 4 categories of results (0- no effects; 1- positive effects i.e. GSS > FtF; 2- negative effects i.e. FtF > GSS; 3- no measures (no hypotheses were tested for a particular independent variable); 4- other effects i.e. interaction effects. Categories 3 and 4 are not shown.

Analysis of variances (ANOVA) were run on the scaled data (Risenthal & Risnow, 1984) where no effect, positive effect, and negative effect were scaled as 1, 2, and 3, respectively.

Results

Overall Analysis: Table 1 presents the frequency (counts) of hypotheses, percentages of "positive," "no effects," and "negative effects," Chi-Square, and ANOVA.

As can be seen from Table 1a, 16.7% of the hypotheses were "positive" (GSS > FtF). The Chi-Square analysis indicates that categories of groups per treatment condition are significantly different. This is confirmed by the ANOVA on the scaled counts. A Duncan's multiple range test (DMRT) shows that 4 to 6 groups per treatment condition category is significantly different from the other two. There are no significant differences between the 7 to 10 group and 11 and up categories.

Task Type: From Table 1b, it can be seen that the percentage of "positive effects" is higher for groups performing decision making tasks than for idea generation tasks across all groups per treatment conditions. A Chi-Square analysis and ANOVA confirms this observation. Furthermore, DMRT results further support these observations.

Table 1c continues the task type analysis across the three groups per treatment conditions. The Chi-Square analysis shows that the distributions of hypothesis counts are significantly different between groups per treatment condition. This is also confirmed by the ANOVA and by subsequent DMRT results.

Task Type and GSS: Tables 1d and 1e show the counts, Chi-Square, and ANOVA by GSS type and Task type. For GSS groups the Chi-Square and ANOVA indicated that there are significant relationships between the groups per treatment condition category and the percentages of effects. DMRT shows that for 4 to 6 groups there is a significant difference between task type. Although the Chi-Square test indicates a significant between task type the ANOVA and DMRT do not for the 7 to 10 groups per treatment condition category.

For CMC groups (Table 1e) the cell counts are shown, but because there are less than 5 in 1/3 of the cell the Chi-Squares are not valid.

Post-hoc Analysis

Are any of the percentages of "positive effects" significantly different? These post-hoc questions can be computed by testing the difference between proportions and using the Z statistic (Hayes, 1973).

Question	Z Statistic
Is the % "positive effects" significantly different between 4 to 6 and 7 to 10 groups per treatment condition (14.4% vs. 19.5%, 1a)? No!	Z= 1.45, p > .1
Between task type (15.2% vs. 19.9% ,1b)? No!	Z=1.57, p > .1
Between task for 4 to 6 groups (9.1% vs. 21.8%, 1c)? Yes!	Z=2.42, p=.01
Between task for 7 to 10 groups (32.9% vs. 12.9%, 1c)? Yes!	Z=2.98, p=.005
Between 4 to 6 and 7 to 10 by idea generation task (9.1% vs. 32.9%, 1c)? Yes!	Z=4.03, p=.0001
Between task for 4 to 6 groups (9.1% vs. 14.8%, 1d)? No!	Z= 1.10, p > .1
Between 4 to 6 and 7 to 10 by idea generation task (9.1% vs. 37.0%, 1d)? Yes!	Z=3.99, P=0001

Conclusions

The results suggest that studies with less than 7 groups per treatment condition have significantly more "no effect" results and consequently a lower probability of "positive results." Task and GSS type further moderate this effect.

Groups performing idea generation tasks have significantly less "positive effects" than group performing decision making task when there are 6 or less or 11 or more groups per treatment condition. Conversely, when there are 7 to 10 groups per treatment condition distribution of effects between idea generation and decision making tasks are significantly different. Groups working on idea generation tasks have significantly more "positive effects" based on the Z statistic. From observation, groups working on decision-making tasks have twice as many "negative effects."

The highest level of "positive effects" (37.0%, 7 to 10 groups) were observed for GSS groups performing idea generation tasks. CMC groups appear to have more "positive" and "negative" effects when working on decision-making tasks in comparison to idea generation tasks.

It is difficult to explain these observations. However, a reported finding of "no effect" may stem from a low power or a true "no effect" situation (Medler, Schneider, and Schneider, 1981). Cohen's (1988) opening sentence "the power of a statistical test is the probability that it will yield statistically significant results," suggests that GSS research needs to re-think how to evaluate the current research. Hays (1973) suggests that Beta will be ordinarily small for a large Alpha, thus by setting Alpha to be larger, the experiments will have more power. Then based on the data presented here, it might make for some controversy, but for experiments that have less than seven groups per treatment condition, call them exploratory studies and set Alpha to 0.1. This may increase the likelihood of "positive effects." The statistical test will become more powerful, and in principle, it is more costly to make the mistake of overlooking a true departure from the Null hypothesis, but not as costly to reject the Null hypothesis falsely (Hays, 1973; Braoudl and Orlikowski 1989).

Table 1. Groups per Treatment Condition

Percentages exclude No Measures and Other Effects

(Unit of Measurement is the Hypothesis)

(Unit of Measurement)	Task Type	Total	No	Positive	Negative	Percent	Percent	Percent
	or		**Effects**	**Effects**	**Effects**	**Positive**	**No**	**Negative**
	Experimental		**GSS = FtF**	**GSS > FtF**	**FtF > GSS**	**Effects**	**Effects**	**Effects**
	Design					**GSS > FtF**	**GSS = FtF**	**FtF > GSS**
Table 1a Groups per Treatment Condition								
3 or less		50	35	6	9	12.0	70.0	18.0
4 to 6		269	208	39	22	14.4	76.2	9.7
7 to 10		200	125	39	36	19.5	62.5	18.0
11 and up		341	213	60	68	17.6	62.5	19.9
Total		860	581	144	135	16.7	67.6	15.7
Chi-Square = 22.06, df= 4, p=0.001; F = 11.04, df= (2,809), p = 0.0001								
Table 1b Task Type Across All Treatment Conditions								
	Idea Generation	323	253	49	21	15.2	Chi-=0.77, df= 2, p = 0.001 F= 31.89, df= (1,643), p = 0.0001	
	Decision Making	321	194	64	63	19.9		
Table 1c Groups per Treatment Condition								
4 to 6 groups	Idea Generation	143	126	13	4	9.1	Chi= 21.12, df= 2, p =0.001 F = 22.86, df= (1,220), p = 0.0001	
	Decision Making	78	49	17	12	21.8		
7 to 10 groups	Idea Generation	76	44	25	7	32.9	Chi=12.16, df= 2, p= 0.002 F= 1.04, df= (1,145), p = 0.3101	
	Decision Making	70	43	9	18	12.9		
11groups and up	Idea Generation	85	70	10	5	11.8	Chi= 16.82, df=2 , p= 0.001 F = 16.37, df= (1.231), p=0.001	
	Decision Making	147	83	33	31	22.4		
Table 1d Groups per Treatment Condition for GSS Groups Only by Task Type								
4 to 6 groups	Idea Generation	143	126	13	4	9.1	Chi= 12.47, df= 2, p =0.002 F = 12.81, df = (1,203), p = 0.0004	
	Decision Making	61	43	9	9	14.8		
7 to 10 groups	Idea Generation	54	31	20	3	37.0	Chi= 15.63, df= 2, p =0.001 F = 0.95, df =(1,116), p = 0.3308	
	Decision Making	63	40	7	16	11.1		

Table 1e Groups per Treatment Condition for CMC Groups Only by Task Type								
4 to 6 groups	Idea Generation	None						
	Decision Making	17	6	8	3	47.1		
7 to 10 groups	Idea Generation	23	13	5	4	22.3	Chi-Square not valid	
	Decision Making	7	3	2	2	28.6		
11 groups and up	Idea Generation	15	12	2	1	16.7	Chi-Square not valid	
	Decision Making	56	31	13	12	23.2		

References

Baroudl, J.L., and Orlikowsi, W.J., "The Problem of Satistical Power in MIS Research," *MIS Quarterly*, March, 1989, 87-106.

Cash, J.J. and Lawrence, P.R., (Eds.) The Information Systems Research Challenge: Qualitative Research Methods, Volume 1, Harvard Business School, Boston, 1989.

Cohen, J., Statistical Power Analysis for the Behavioral Sciences, Lawrence Erlbaum, Associates, Hillsdale, JY, 1988.

DeSanctis, G., "Small Group Research in Information Systems: Theory and Method," in The Information Systems Research Challenge: Experimental Research Methods, Volume 2, Benbasat, I. (Ed.) Harvard Business School, Boston, 1989

Fjermestad, J., and Hiltz, S.R., "Experimental Studies of *Group* Decision Support Systems: An Assessment of Variables Studied and Methodology," *Proceedings of the 30th Hawaii International Conference on System Sciences*, II, 1997, 45-65

Fjermestad, J. and Seah, T.S., "Using an Executive Information System to Analyze Group Support Systems: A Preliminary Implementation**,"** *Proceedings of the Third Americas Conference*, August 1997, 563-565

Janis, I.L., Victims of Groupthink, Houghton-Mifflin, Boston, 1972

Jessup, L.M., "Group Decision Support Systems: A Need for Behavioral Research," *International Journal of Small Group Research*, September 1987, 139-158

Hayes, W.L., Statistics for the Social Sciences, Holt Rinehart, and Winston, New York, 1973.

Medler, J.F., Schneider, P.R., and Schneider, A.L., Statistical Power Analysis and Experimental Field Research, Some Examples from the National Juvenile Restitution Evaluation, Evaluation Review, 5,6, 1981, 834-850.

Poole, M.S., Seibold, D.R., and McPhee, R.D., "Group Decision-Making as a Structurational Process," *Quarterly Journal of Speech*, 71, 1985, 74-102

Rosenthal, R., and Rosnow, R.L., Essentials of Behavioral Research Methods and Data Analysis, McGraw-Hill, New York, 1984.

Zigurs, I., "Methodological and Measurement Issues in Group Support Systems Research," in Group Support Systems New Perspectives, Jessup, L.M. and Valacich, J.S. (Eds.), Macmillan Publishing Company, New York, 1993

Zmud, R.W., Olson, M.H., and Hauser, R., "Field Experimentation in MIS Research", in The Information Systems Research Challenge: Experimental Research Methods, Volume 2, Benbasat, I. (Ed.) Harvard Business School, Boston, 1989.