# Faculty Working Papers

A TEST OF THE NOMINAL AND DELPHI TECHNIQUES
IN GROUPS OF DIFFERENT SIZES

David M. Boje, Graduate Student in Business
Administration, and J. Keith Murnighan,
Assistant Professor of Business Administration

#470

**College of Commerce and Business Administration**
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

March 1, 1978

A TEST OF THE NOMINAL AND DELPHI TECHNIQUES
IN GROUPS OF DIFFERENT SIZES

David M. Boje, Graduate Student in Business
Administration, and J. Keith Murnighan,
Assistant Professor of Business Administration

#470

Summary:

   This study investigated the effects of modified forms of the Delphi and
Nominal group decision making techniques on a set of four problems in groups
of sizes 3, 7, and 11. Data were collected for a series of three trials;
between trials Nominal group members received verbal feedback from their
group while Delphi group members received written feedback. Results were
compared to means obtained from randomly selected individual estimates. The
data suggested that the mean estimates of a set of individuals were somewhat
more accurate than those from Nominal groups. At the same time, members of
the Nominal and Delphi groups became more confident of their answers, sug-
gesting the possibility that groupthink (Janis, 1972) was prevalent. No
effects for different group sizes were found. Discussion suggested that
future research is necessary to clarify the findings within this area.

The use of committees and groups to solve problems is a common orga-
izational phenomenon. Early research on group dynamics (e.g., Cartwright
and Zander, 1968) has indicated that informal group meetings subject group
members to strong social pressures that may inhibit the creativity and
efficiency of any problem solving they attempt. Several more formal pro-
cedures, including the Delphi and Nominal techniques (Delbecq, Van de Ven,
Gustafson, 1975), have been devised to alleviate some of these pressures
and, hopefully, improve the efficiency of group problem solving. The
present study was designed to evaluate the accuracy of the solutions to
several tasks reached by groups of different sizes using different pro-
cedures. As is typical in the group problem solving literature, the results
were also compared to the solultions of a set of non-interacting individuals
to assess whether the group interaction led to improved accuracy.

The Delphi technique is one of the formal procedures that was devised
to remove the social pressures inherent in group problem solving. The
technique proceeds through several stages. First, the problem is stated
as clearly as possible. An open-ended questionnaire soliciting any and
all potential solutions is then sent to the participants. When the quest-
ionnaires have been returned, the responses are summarized, and reactions
to the proposed solutions are elicited, again by individual questionnaire.
If a clear consensus emerges from the responses to this second questionnaire,
the solution (i.e., the consensus) is adopted and final reports are sent to
the participants. If substantial differences of opinion exist, additional
questionnaires, each summarizing the results of the previous questionnaire
and asking for additional opinion, are necessary. Thus, this procedure

allows for no verbal or face-to-face contact, and, given the anonymity of the participants and their responses, removes the possibility for social pressure.

Previous research of a somewhat uncontrolled nature (Dalkey, 1972; Pfeiffer, 1968) indicated that, for Almanac problems (e.g., how many phones were there in Africa in 1967?), groups using the Delphi technique were more accurate than groups using no formal procedure (i.e., interacting groups). Gustafson, Shukla, Delbecq, and Walster (1973) asked face-to-face but non-interacting subjects to work on problems that required a probability estimate concerning whether an individual of a particular height or weight was a male or female. Their results indicate that: (1) the Delphi technique did not increase efficiency over either interacting or separated individuals; and (2) after feedback of the other group members' responses, second estimates were no better and sometimes worse than original estimates. In another study, Van de Ven (1974) asked subjects to generate a job description for dormitory counselors; the results indicated that the Delphi technique resulted in much richer descriptions than those generated by interacting groups.

Thus, the research evaluating the effectiveness of the Delphi technique is somewhat mixed. When there is no correct answer, the technique generates a large number of alternatives. For Almanac problems (Dalkey, 1972; Pfeiffer, 1968), the results are encouraging; for probability estimates (Gustafson et al., 1973) they are not. Thus, in order to evaluate the generality of the Delphi technique's accuracy, groups in the present study were presented with both Almanac problems and problems that required a probability estimate. In addition, subjects were asked to make successive estimates. Dalkey (1972)

indicates that accuracy increases over trials for the Almanac problems;

Gustafson et al., (1973) indicate that accuracy decreased over trials for

probability estimates.

In addition Dalkey (1972) has suggested that the accuracy of the Delphi

technique should improve with larger groups.  Since the addition of group

members cannot increase social pressure within the group and thus cannot

increase process losses (Steiner, 1972), larger groups should be more accurate

than small groups.  This study used groups with 3, 7, and 11 members to test

this hypothesis.

The Nominal technique was devised by Delbecq and Van de Ven (1971), and

incorporated some aspects of individual brainstorming (cf., Dunnette, Campbell,

and Jaastad, 1963) in face-to-face group meetings so that groups might gener-

ate many alternatives, and might also be committed to the solution by the

group members.  This process also involves several steps.  First, individuals

are presented with a clear statement of the problem by the leader.  Each

group member is then encouraged to generate as many alternative solutions

as possible, without discussion.  A round-robin presentation of solutions

follows, with each group member presenting a solution in turn.  Members are

encouraged to present solutions that build on solutions presented by other

group members.  After the leader has recorded all the potential solutions

and displayed them in full view of the group, they discuss each alternative,

focusing on clarification and evaluation.  Finally, the group votes on the

most appropriate solutions, and alternates between votes and short discussions

until they reach an obviously favored solution.  Unlike Delphi, the Nominal

technique involves face-to-face interaction; this has been emphasized as

the major advantage of the technique by Nominal group advocates (e.g.,

Delbecq et al., 1975) and the major disadvantage by Delphi advocates

(e.g., Tersine and Riggs, 1976). Research results which have compared

the two has found little difference in the number of alternative solutions

suggested (Van de Ven, 1974) and an advantage in accuracy for the Nominal

technique (Gustafson et al., 1973). The present study again pits one

technique against the other, with two different problem types, three

group sizes, and three trials for each problem.

## Method

### Subjects

324 undergraduate students enrolled in an introductory organizational

behavior course at a large midwestern university received credit toward a

course requirement for their participation in the study.

### Procedure

Subjects were told that they would be using one of several decision

making techniques that were analogous to those used by groups in organizations.

Each individual was given the four problems; all were encouraged to try to be

as accurate as possible in their estimates. In the Nominal process groups,

subjects were introduced to each other, were seated around a table, began

to work on the problems individually, and were asked to maintain silence as

they generated their answers. As in the Delphi conditions, group

members were asked for an estimate of the correct answer and one fact or

reason in support of the estimate. In the Nominal groups, individuals

presented their estimates in a round-robin fashion to the experimenter,

who recorded them on the blackboard. After a brief discussion of the
estimates, the group members were asked to make their second individual
estimate. The round-robin was again followed by discussion for the second
and third trials. Unlike the Nominal process proposed by Delbecq and
Van de Ven (1971), the estimates were not put to a vote. Instead, to
increase comparability to the other conditions, the group's estimate was
operationally defined as the mean of the individual responses at each trial.

In the Delphi and individual process conditions, subjects were seated
behind opaque partitions, restricting interpersonal contact. In the Delphi
condition, each person's individual estimate and reason or fact was copied
verbatim by the experimenter and two assistants and distributed to the
other group members as quickly as possible. In the individual condition,
subjects were told to reflect upon the problems and to think of additional
information that might be important in finding the solution to the problems.
In each case, the group estimate was, again, the mean of the individual
responses at each trial. Because of a scarcity of subjects, the "groups"
in the individual condition were randomly selected with replacement from a
pool of 70 subjects.

Before debriefing, each subject was asked to complete a brief questionnaire
concerning his/her reactions to the decision technique and the experimental
task.

Problems

Two subjective likelihood problems were adapted from Gustafson et al.,
(1973); two problems were also taken from the 1976 World Almanac. The two
subjective likelihood problems (weight and height) and the two Almanac
problems (Jupiter and dollars) were:

(1) <u>Weight</u>: The average weight of men is 154 pounds (69.9 kg). The average weight of women is 128 pounds (58.1 kg). Out of a random sample of 100 people all of whom are 150 pounds (68.0 kg) in weight, how many would be male? (Correct answer: 81.13); (2) <u>Height</u>: The average height of men is five feet nine inches (1.75 m). The average height of women is five feet four inches (1.63 m). Out of a random sample of 100 people all of whom are 68 inches tall (1.73 m), how many would be male? (Correct answer: 64.29); (3) <u>Jupiter</u>: The earth's moon has a diameter of 2,160 miles (3,476 km). The diameter of the sun is 864,000 miles (1,390,435 km). What is the diameter of the planet Jupiter at its equator? (Correct answer: 86,800 miles; 139,687 km); and (4) <u>Dollars</u>: The dollar bill is a piece of paper measuring 2 5/8" (6.67 cm) by 6 1/8" (15.56 cm) with a thickness of .0043" (.109 mm). New notes will stack 233 to an inch, if not compressed. How many dollar bills would be needed to weigh exactly one pound (.45 kg)? (Correct answer: 490).

Following Phillips and Edwards (1966), subjects responded to the subjective likelihood problems on a logarithmically calibrated scale of odds to reduce the potential conservatism effect. To control for possible order effects, problems were arranged in seven separate random orders; each order was used at most once within each treatment condition.

In addition to a numerical estimate and a fact or reason for each problem, subjects indicated the confidence they had in each of their answers on seven-point scales.

## Design

Three levels of group size (three, seven, and eleven), three types of decision techniques (Delphi, Nominal and Individual), four problems (weight,

height, Jupiter, and dollars), and three trials were examined in a 3 x 3 x 4 x 3 design. Problems and trials were repeated measures; group size and technique were between factors.

The dependent variables were: (1) A measure of group accuracy (the deviation of the group mean from the correct answer, standardized to allow comparisons among problems); (2) The mean confidence rating reported by individuals in each trial and for each problem. (3) Individual responses on ten questionnaire items (see Table 3 for a listing of the questions), which questioned the efficiency of the technique they used.

Due to an insufficient number of subjects, there were a maximum of seven groups in each of the technique/size conditions. In the Delphi condition, there were 7 three-person groups, 6 seven-person groups, and 6 eleven-person groups. In the Nominal conditions, there were 7, 7, and 5 groups, respectively. And in the individual conditions, there were 7, 7, and 6 groups.

## Results

### Accuracy

The mean estimates for each problem by the groups using different techniques are shown, for each of the three trials, in Table 1. Separate results are not depicted for the different group sizes because size yielded no significant effects on the accuracy of the group's estimates. The results of the size by techniques by problems by trials analysis of variance for the standardized accuracy deviation scores revealed one significant effect, the techniques by trials interaction: $\underline{F}(4,98) = 3.26$, $\underline{p} < .05$. Table 2 depicts the means of this interaction. The main ef-

fect for techniques approached standard significance levels, $F(2,49) =$ 2.25, $p < .12$, and the means depicted in Table 2 suggest that the Nominal technique over-estimated the correct responses more than the other two techniques. Post hoc tests on the interaction means using the Newman-Keuls procedure revealed no significant differences between the means. This failure to obtain statistical significance does not imply that there are no important differences between the techniques. The

Insert Tables 1 and 2 about here

data suggest that while the Delphi and Nominal groups moved toward greater overestimates over trials, the participants in the Individual condition moved toward greater accuracy. The prediction that trials will lead to increases in judgment accuracy with the use of either the Delphi or the Nominal technique was not supported. It should be noted that neither size nor problems yielded significant effects for accuracy.

These results might be questioned, however, because high overestimates and high underestimates in the same condition might yield a mean, for that cell, that would appear accurate. Thus, the variances of the standardized deviation scores for the cells in the techniques by trials interaction were compared to one another (see Table 3). The results clearly show that the Nominal technique yields considerably larger variances than either the Delphi or the Individual conditions. The random selection of individual estimates in the individual condition almost assured low variances in those conditions. Nevertheless, as Gustafson, et al., (1974) have pointed out, a technique that yields consistent estimates is desirable. The Delphi

and Individual technique were clearly more consistent than the Nominal technique, with the Individual technique having a slight advantage.

<hr>
Insert Table 3 about here.
<hr>

To further insure that over- and under-estimates did not cancel one another, an additional analysis, using the standardized absolute value of the difference between the group estimate and the correct answer, was run. The results yielded only one significant effect, for techniques: $F(2,49) = 4.04$, $p < .03$. Post hoc tests on the means (using the Newman-Keuls technique) revealed that the Delphi and Individual techniques were significantly ($p < .05$) more accurate than the Nominal technique, and that no other differences were significant. The Individual technique was slightly better than the Delphi technique. This analysis supports the previous conclusion, that the Individual technique yields more accurate estimates than either of the other two techniques, although it is not significantly better than the Delphi technique. The fact that the previously significant techniques by trials interaction was not significant in this analysis leads to some question about the effect of trials.

## Confidence

An analysis of variance for size by technique by problems by trials for an individual's confidence in his decision accuracy yielded significant main effects for techniques ($F(2,49) = 6.41$, $p < .004$), for problems ($F(3,147) = 239.75$, $p < .0001$), and for trials ($F(2,98) = 61.58$, $p < .0001$). Also significant were the problems by trials ($F(12,294) = 21.15$, $p < .001$) and the techniques by problems by trials interactions ($F(12,294) = 2.21$, $p < .02$).

Post hoc tests indicated that the participants' confidence was significantly greater in the subjective likelihood problems (weight and height) than the Almanac problems (Jupiter and dollars). Due to the large number of means involved in the technique by problems by trials interaction, it is not clear which means contributed most to the effect. The problems by trials interaction, however, appears to be due to considerably greater increases in confidence for the Almanac problems, especially for the dollars problem, than for the probability estimates. This in turn may be explained by the observation that participants conveyed more written and verbal information on these problems.

Table 4 displays the means within each technique over trials. Although this interaction only approached significance ($F(4,98) = 2.03$, $p < .10$), the table indicates that, in each condition, the participants' confidence increased over trials. Post hoc comparisons among the techniques yielded no significant differences, although participants in the Delphi groups started more confident and continued to remain more confident than the participants in the other conditions.

In both Nominal and Delphi groups, subjects attempted to influence others by suggesting they had expertise that was relevant to the Almanac problems (such as completing an astronomy course or handling money in a bank). Such statements within the group may also have contributed to an explanation for the increased confidence, even when the information conveyed might have been erroneous.

Insert Table 4 about here

There was no support for the prediction that verbal information in the Nominal technique would result in participants being more confident in this technique. In fact, the Delphi technique produced slightly (not significantly) more confidence than the other two techniques. Advocates of the Delphi procedure might suggest that this result can be explained by the absence of loss of face involved in the use of information provided by others.

## Questionnaire Items

The only significant effects on the questionnaire items were two main effects for group size and seven main effects for technique. Post hoc tests using the Newman-Keuls procedure indicated that the three-person groups felt more people would increase their group's accuracy, compared to the eleven-person groups. Also, people in the eleven-person groups felt significantly less free to contribute their ideas than people in the seven-person groups.

Table 5 displays the means and $F$-ratios for the significant effects for techniques. The results suggest that the Delphi technique is only somewhat superior in the eyes of the participants than working alone, and that the Nominal technique generally resulted in the most positive perceptions of effectiveness, satisfaction, and freedom.

Insert Table 5 about here

## Discussion and Conclusions

Several of the results in this study were very curious. First of all, group size had no effect on the accuracy of the group decisions. Even though group size has often been cited as a critical variable in the study of groups (e.g., Cummings, Huber, and Arendt, 1974; Davis, Laughlin, and

Komorita, 1976) size in this study affected neither accuracy nor confidence.
The fact that size did affect perceived freedom in contributing one's ideas
replicates earlier results (e.g., Hackman and Vidmar, 1970). Also somewhat
expected were the results indicating that the small groups felt additional
members would increase their group's accuracy. Size increases, however,
did not increase group accuracy. It may be that in groups that use structured
decision processes, like those employed in this study, size has little oppor-
tunity to affect the outcomes of the group process. With less structured
conditions, size may be more effectual. Other studies, possibly using
additional sets of problems, might test this hypothesis.

Another curious finding in this study is that, as the Nominal and
Delphi groups became more confident in their estimates (over trials), they
also became less accurate (at least for the raw accuracy scale). Thus, a
form of groupthink (Janis, 1972) may have been operating in these groups.
Greater influence may have been exerted by group members espousing the
least accurate estimates. Further research is necessary to substantiate
this suggestion, and might possibly tie such a notion to the recent results
on the group polarization phenomenon (Myers and Lamm, 1976).

The third curious finding in this study stems from a comparison of
the accuracy scores and subjects' perceptions of the techniques. The
Nominal technique yielded the least accurate answers, but the most positive
affective responses. It seems obvious that the interpersonal interaction
within the group led to more positive ratings on the questionnaire items,
and that isolation or working on the problems alone is much less pleasant.
The "misperceptions" by the Nominal group members further suggests the

presence of groupthink. Also, in terms of effectiveness, these data suggest that when accuracy is important, group members should not interact with one another, or even exchange information. In situations where accuracy is less important, and satisfaction is more important, interaction appears to be desirable. Future research might test these notions and pursue the possibility, not documented here, that certain conditions might lead to both positive affect _and_ accuracy.

Merely averaging a set of individual estimates led to estimates that were as good or better than either of the group techniques. Thus, while previous research was inconsistent in its support of one or the other technique, the present research calls both into question. The variety of methodologies among studies suggests that an unconsidered factor may be leading to spurious results in all of the research relating these techniques. More research is needed to investigate this possibility. An accumulation of research findings that would allow observers to hypothesize which of the underlying dimensions might be critical is one approach to resolve the problem. Another is to hypothesize what dimensions are most important, manipulate them, and hope that the results are a function of that particular factor. In either case, considerable future research appears warranted.

REFERENCES

Cartwright, D., and Zander, A. Group Dynamics. New York: Harper and Row, 1968.

Cummings, L. L., Huber, G. P., and Arendt, E. Effects of size and spatial arrangements on group decision making. Academy of Management Journal, 1974, 17, 460-475.

Dalkey, N. C. Studies in the Quality of Life: Delphi and Decision Making. Lexington, Mass.: Lexington Books, 1972.

Davis, J. H., Laughlin, P., and Komorita, S. S. Small group research in Social Psychology, 1968-1975. Annual Review of Pyschology, 1976, 24, 1101-1188.

Delbecq, A. L., and Van de Ven, A. H. A group process model for problem identification and program planning. Journal of Applied Behavioral Science, 1971, 7, 3.

Delbecq, A. L., Van de Ven, A. H., and Gustafson, D. H. Group Techniques for Program Planning: A Guide to Nominal and Delphi Processes. Glenview, Ill: Scott Foresman, 1975.

Dunnette, A. L., Campbell, J., and Jaastad, R. The effects of group participation on brainstorming effectiveness for two industrial samples, Journal of Applied Psychology, 1963, 47, 30-37.

Gustafson, D. H., Shukla, R. K., Delbecq, A., and Walster, G. W. A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, delphi groups, and nominal groups, Organizational Behavior and Human Performance, 1973, 9, 280-91.

Hackman, R. J., and Vidmar, N. Effects of size and task type on group performance and member reactions, Sociometry, 1970, 33, 37-54.

Janis, I. Victims of Groupthink. Boston: Houghton Mifflin, 1972.

Myers, D. G., and Lamm, H. The group polarization phenomenon. Pscyhological Bulletin, 1976, 83, 602-627.

Pfeiffer, J. New Look at Education. New York: Odyssey Press, 1968.

Phillips, L. D., and Edwards, W. Conservatism in a simple probability inference task, Journal of Experimental Psychology, 1966, 72, 346-54.

Steiner, I. Group Process and Productivity. New York: Academic Press, 1972.

Tersine, R. J., and Riggs, W. E. The Delphi technique: A long-range planning tool, Business Horizons, 1976, 19, 51-56.

Van de Ven, A. and Delbecq, A. L. Nominal versus interacting group processes for committee decision making, Academy of Management Journal, 1971, 14, 203-13.

Van de Ven, A. Group Decision Making and Effectiveness: An Experimental Study. Kent, Ohio: Kent State University, 1974.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

FOOTNOTE

1. These data were used to conduct a homogeneity of variance test;
   Hartley's $F_{max}$ (9.11) = 7.41, $p$ < .10. Because analysis of variance
   is robust with respect to non-homogeneity (Winer, 1962), this result
   indicates that interpretation of the previous result is not problematic.

TABLE 1

Mean Estimates for Each Problem by

Each Group Technique Over The Three Trials

| Problem (Correct Answer) | Technique | Trials 1 | 2 | 3 |
|---|---|---|---|---|
| Weight (81.13) | Nominal | 73.75 | 74.22 | 74.45 |
| | Delphi | 73.96 | 76.74 | 78.64 |
| | Individual | 73.21 | 69.53 | 72.97 |
| Height (64.29) | Nominal | 70.70 | 76.79 | 77.39 |
| | Delphi | 65.42 | 69.94 | 72.14 |
| | Individual | 67.90 | 69.41 | 70.51 |
| Jupiter (86,800) | Nominal | 154,000 | 191,400 | 198,600 |
| | Delphi | 95,990 | 155,500 | 173,100 |
| | Individual | 119,600 | 140,400 | 136,600 |
| Dollars (490) | Nominal | 949.4 | 1081.3 | 984.2 |
| | Delphi | 653.1 | 877.1 | 816.1 |
| | Individual | 942.1 | 973.2 | 1017.6 |

TABLE 2

Mean Standardized Accuracy Scores for the

Techniques x Trials Interaction

| Techniques | Trials 1 | 2 | 3 | Mean |
|---|---|---|---|---|
| Nominal | .163 | .270 | .180 | .205 |
| Delphi | -.186 | -.054 | -.015 | -.085 |
| Individual | .020 | -.198 | -.158 | -.112 |
| Mean | -.001 | -.006 | .003 | -- |

Note: The correct answers for the four problems, after being transformed to standardized scores, differed; the mean of the transformed scores was -.430.

TABLE 3

Variances of the Standardized Accuracy Scores

for the Techniques x Trials Interaction

| Techniques | Trials | | |
|------------|--------|------|------|
|            | 1      | 2    | 3    |
| Nominal    | .688   | .537 | .584 |
| Delphi     | .174   | .243 | .193 |
| Individual | .093   | .132 | .152 |

TABLE 4

Mean Confidence Ratings within the

Problem Solving Techniques Over Trials

| Techniques | Trials 1 | 2 | 3 | Mean |
|---|---|---|---|---|
| Nominal | 3.07 | 3.66 | 3.79 | 3.50 |
| Delphi | 3.52 | 3.95 | 4.21 | 3.89 |
| Individual | 3.05 | 3.33 | 3.46 | 3.28 |
| Mean | $3.21_a$ | $3.65_b$ | $3.82_b$ | -- |

TABLE 5

Mean Response and Summary of Analyses for the Questionnaire Items

| Questionnaire Item | Technique | | | $F_{(2,49)} =$ | $p <$ |
| --- | --- | --- | --- | --- | --- |
| | Delphi | Nominal | Individual | | |
| 1. How would you rate the procedure you used as a way of making decisions? (Very Poor--Excellent) | $2.88_{ab}$ | $3.32_a$ | $2.69_b$ | 3.38 | .05 |
| 2. How much did you enjoy working on this project? (Not at All--A Great Deal) | $2.64_a$ | $3.54_b$ | $2.56_a$ | 16.12 | .001 |
| 3. To what extent do you feel this decision technique was appropriate for these problems? (Not at All--A Great Deal) | $2.84_{ab}$ | $3.33_a$ | $2.69_b$ | 4.78 | .01 |
| 4. Do you feel you had enough time to answer these questions? (Too Little--Too Much) | $4.64_a$ | $4.74_a$ | 5.55 | 19.67 | .001 |
| 5. Do you feel greater accuracy would result if more people worked with you on this project? (Less Accurate--More Accurate) | $5.23_{ab}$ | $4.94_a$ | $5.69_b$ | 7.30 | .01 |
| 6. To what extent did you feel free to contribute your ideas? (Not at All--A Great Deal) | $4.44_a$ | $5.40_b$ | $4.09_a$ | 21.75 | .001 |
| 7. Do you feel others working on this problem influenced your response? (Not at All--A Great Deal) | $3.85_a$ | $4.72_b$ | $1.44_c$ | 86.12 | .001 |

NOTE: Means sharing a common subscript within the levels of each main effect are not significantly different from one another at the .05 level using the Newman-Keuls procedure.