TECHNOLOGICAL FORECASTING AND SOCIAL CHANGE 23, 89–94 (1983)

The Delphi Technique
An Experimental Evaluation

WALTER E. RIGGS

ABSTRACT

Much research has been devoted to the Delphi technique. However, very little substantive work has been done on the subject of Delphi accuracy. The purpose of this effort was to test the accuracy of Delphi vs the conference method in making long-range forecasts. College students were used to form Delphi and conference groups that predicted the point spreads of college football games far in advance of play. The results substantiate the claim that Delphi outperforms conference methods on the basis of accuracy for long-range forecasting.

For several decades organizations have attempted to amass the talents of groups of individuals in an effort to combine their individual skills and improve decision making. Group decisions are necessary when the scope of a problem is such that no one individual has sufficient expertise and knowledge to effect a solution. However, the traditional process of grouping these "experts" together has caused a number of counterproductive side effects.

The Delphi technique was developed at the RAND Corporation in 1950 by Dalkey and his associates to eliminate many of the negative effects related to the use of interacting groups for decision making (especially long-range forecasting) [3]. The purpose of the research reported here was to investigate the forecast accuracy of the Delphi technique vs the conference method under differing levels of group knowledge concerning the forecast objectives.

Van de Ven [14] has identified several major factors that inhibit the performance of interacting groups. Previous research substantiates the effects of these factors. Among them is the tendency of low-status members to "go along" with the opinions of high-status members in spite of contrary feelings [13]. Group pressures for conformity also negatively affect the participants [8]. Interacting groups have a tendency to "fall into a rut" and pursue a single train of thought for long periods [5]. Also, dominant personalities have a tendency to influence the group [1]. Often, compromise decisions are reached rather than consensus decisions when interacting groups are employed.

The Delphi technique was developed to overcome the problems of interacting groups. The participants utilize written responses. Face-to-face contact is not required. Therefore the technique offers significant advantages. Delphi prevents domination by high-status participants or strong personalities. The judgments of individuals hostile to one another

WALTER E. RIGGS is Assistant Professor of Management Department, Georgia State University, Atlanta, Georgia.

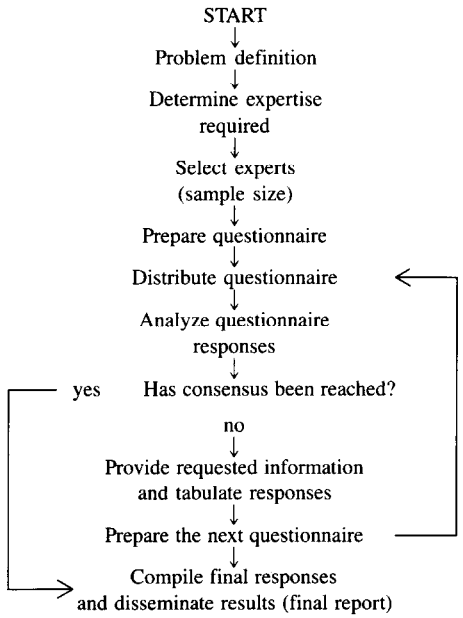
 0040-1625/83/010089–06$03.00

may also be aggregated [4]. It is particularly useful in soliciting information from participants who cannot physically come together.

The Delphi Process

In actual application the specific procedures of Delphi vary, however; the following example is typical. The forecast is initiated by a questionnaire that requests estimates of a set of numerical quantities, that is, dates by which specific technological events will occur, probabilities of occurrence by given dates, event desirability and feasibility, and the like. The results of the first round are summarized, and the median and interquartile range of the responses is computed and fed back to the respondents with a request to revise their first estimates where appropriate. On succeeding rounds, those individuals whose opinions deviate greatly from the majority, that is, are outside the interquartile range, are requested to give a reason for their extreme opinions. A collection of these reasons is presented to each participant along with a new median and interquartile range. They are then given another opportunity to reconsider or revise their earlier opinion or estimate. The process is continued in an iterative manner until a consensus is reached. Preliminary questionnaires are often used to select and develop the questions for which estimates will be obtained. Table 1 provides a flowchart of the typical Delphi procedure.

TABLE 1
A Flowchart of the Typical Delphi Process



There are several extensions and modifications of the Delphi technique. One modification is known as cross-impact analysis, which takes into consideration the impact of the occurrence of one event on a subsequent event when several events are interrelated. Usually the analysis will develop a series of conditional probabilities for events. Through an iterative process the technique insures the forecasts of interrelated events are consistent

with individual probabilities of occurrence. This procedure helps to insure the elimination of contradicting predictions [3].

Another modification is called SEER (System for Event Evaluation and Review). SEER modifies the process by establishing an initial list of forecasts that have been constructed through a series of interviews prior to the beginning of the Delphi process. It can result in reducing the number of questionnaire rounds, thus saving time. Also, participants can be asked to answer questions in only their area of expertise. This aids in eliminating possible distorted responses caused by lack of expertise in a particular area [7].

Previous Research

With regard to the original purpose of Delphi, long-range forecasting, it is difficult to measure the effectiveness of the technique in comparison with other available approaches. Much of the work that has been done with Delphi has assumed that the approach yields more acceptable results than direct confrontation techniques. However, the research that has been done leaves the issue of superiority open to question.

Research on the effectiveness of Delphi has been conducted utilizing two approaches. First, experimenters have attempted to judge how well the Delphi approach predicts events without comparing the approach with another technique. One approach utilized "almanac" data [2]. That is, participants were asked to predict events that had previously occurred but about which they were unaware. Sufficient information was provided for the prediction process. This experiment seemed to result in at least "good" predictions of what actually occurred. Another study evaluated the Delphi prediction process on the basis of convergence, reliability, and comparison with other studies [9]. The conclusion was that Delphi proved credible. A third experiment tested the ability of Delphi to overcome the "bandwagon" effect [6]. Participants were given fictitious information about the first round results, which in turn influenced their second-round opinions in the direction of the false information.

The second approach to assessing the Delphi technique is to compare the process directly with other methods of forecasting. A recent study compared Delphi with two other direct confrontation methods, the first in which a group leader makes a decision with the aid of group members, and the second in which the group opinion is achieved by majority vote [10]. Delphi was determined to be the least accurate of the techniques evaluated. Another recent study evaluated Delphi vs face-to-face methods on the basis of ability to gain consensus and strength of judgments [11]. Delphi proved to be superior on both bases of comparison. A third study compared Delphi with a committee approach in short-term forecasting [12]. Delphi was more accurate, but not at the stated level of statistical significance.

It is clear from the research above that the superiority of Delphi over conventional face-to-face forecasting techniques on the basis of accuracy has not been clearly proven. This experiment attempts to contribute to the resolution of the question of accuracy.

The Experimental Method

The experiment reported in this paper compared the performance of a modified Delphi and the direct confrontation method by proceeding in the following fashion. The events chosen were two college football games whose point spreads were to be simultaneously predicted approximately four weeks in advance. The first game was an intense rivalry about which the participants had intimate knowledge through the media (termed a

high information game). The second game was one in which the participants had some but not a vast quantity of information (termed a low information game). The events were selected because of the quantitative nature of the available parameter, the relatively large volume of information about the four teams available through the media, and the time frame that was sufficiently long for them to be considered surrogates for long-range events (while also being short enough to be easily dealt with experimentally). This selection also permitted the effectiveness of the techniques to be tested in both high and low knowledge level situations.

The participants in the experiment were junior- and senior-level undergraduate business students. The choice of these students was based on the relatively high degree of knowledge they possessed concerning factors affecting the events through heavy exposure to the media. Also, the large number of participants afforded was a significant factor in this selection.

Several undergraduate classes of students were chosen to participate. Each class was divided into halves, with each half being physically separated. The halves were then broken into groups of four to five participants. Class division and group membership were randomly assigned. Each of the groups in the first half of the class was asked to conduct a discussion resulting in prediction of the point spread for each of the two games. These were, in effect, direct confrontation groups.

The second half of each class constituted the Delphi groups. Each member of a Delphi group was told not to interact with other participants in any way. They were seated so that none of the participants in a group faced anyone else. The members of each of the groups were asked to record their estimate of the point spread for each of the two games. The estimates were collected and the mean point spread computed for each Delphi group. The participants were then asked to revise their original estimate given the mean computed for the first round. This revision process constituted the second round of the Delphi. The mean of this revision was then used as the prediction for the group.

Members of both the direct confrontation and Delphi groups were provided with information concerning the four teams involved. Specifically, the previous records and scores were supplied as well as information on injuries to key players.

Experimental Results

A total of 16 groups participated in the experiment, 8 conference groups and 8 Delphi groups. After the completion of the two games, the absolute differences between the predicted and actual point spreads were computed. A two-way analysis of variance with an interaction effect was then employed to determine whether there was a significant difference between the Delphi and conference forecasts. Table 2 contains the results of that analysis of variance.

A significant difference was obviously expected in the accuracy of prediction for the high vs the low information game. This difference was substantiated with an F value of 113.95. A significant interaction effect was not expected between games and forecasting methods. This was also borne out with an F value of 3.67 that is not significant at the .05 level. However, at the .05 level a significant difference was found between the forecasting methods with an F value of 8.89. (The critical F value for 1 and 28 degrees of freedom at the .05 level is 4.20.) The R^2 value for this experiment was 0.82.

TABLE 2
ANOVA Table for Forecast Method and Game Effects

Source	Sum Square	Degrees of Freedom	Mean Square	F
Game	1067.80	1	1067.8	113.95
Forecast method	83.37	1	83.37	8.89
Interaction	34.37	1	34.37	3.67
Error	262.47	28	9.37	
Total	1448.01	31		

Conclusions

For this particular experiment the analysis-of-variance results clearly indicate a singificant difference between the conference method and the modified two-round Delphi. The mean absolute forecast errors (MAFEs) for the high information game were 2.2 and 5.6 for Delphi and conference methods, respectively. For the low information game the MAFEs were 13.9 for Delphi and 17.0 for conference. The results therefore appear to support the contention that Delphi is superior to the conference method for long-range forecasting in both high and low information environments.

However, these results should be viewed with caution. The Delphi technique was modified for this particular experiment. The participants were not physically dispersed and therefore had knowledge of, but no verbal interaction with, the other members in their group. Also, only two rounds were employed in the Delphi (although one would think that a third round would only result in an improvement). Finally, the student participants could not be truly classified as "experts" in football predictions. However, the students did possess much information and a high level of interest concerning the games. Overall, in spite of the obvious limitations of the study, the evidence lends credibility to the statement that Delphi procedures are superior to conference methods for long-range forecasting.

References

1. Chung, K., and Ferris, M., An Inquiry of the Nominal Process, *Academy of Management Journal* 14:520–524 (1971).
2. Dalkey, N., Comparison of Group Judgment Techniques with Short-Range Predictions and Almanac Questions, Rand Corporation, R–678, New York, 1971.
3. Dalkey, N., and Helmer, O. An Experimental Application of the Delphi Method to the Use of Experts, *Management Science* 9:458–467 (1963).
4. Delbecq, A. Van de Ven, A., and Gustafson, D., *Group Techniques for Program Planning*, Scott, Foresman, Glenview, Ill.: 1975.
5. Dunnette, M., Campbell, J., and Jaastad, K., The Effect of Group Participation on Brainstorming Effectiveness for Two Industrial Samples, *Journal of Applied Psychology* 47:308–318 (1963).
6. Francis, A., An Experimental Analysis of a Delphi Technique: The Effect of Majority and High Confidence–Low Confidence Expert Opinion on Group Consensus, Unpublished Ph.D. dissertation, The Pennsylvania State University, 1977.
7. Fusfeld, A., and Foster, R., The Delphi Technique: Survey and Comment, *Business Horizons* 14:63–74 (1971).
8. Hoffman, L., Group Problem Solving. In *Advances in Experimental Social Psychology*, L. Berkowitz (ed.), Academic Press, New York, 1965.
9. Lee, J., A Methodological Evaluation of Delphi Forecast Procedures and Products in a Field Study Situation, Unpublished Ph.D. dissertation, University of Minnesota, 1977.

10. Miner, F., The Effectiveness of Problem-Centered Leadership, Nominal Leadership, and the Delphi Process in a High Quality–High Acceptance Problem, Unpublished Ph.D. dissertation, University of Minnesota, 1976.
11. Penfield, G. The Relative Efficacy of Varying Applications of Face to Face Interaction Versus Delphi in Developing Consensus about Relative Priority Among Goals in Student Affairs, Unpublished Ed.D. dissertation, University of Cincinnati, 1975.
12. Sack, J., A Test of the Applicability of the Delphi Method of Forecasting as an Aid to Planning in a Commercial Banking Institution, Unpublished D.B.A. dissertation, Arizona State University, 1974.
13. Torrance, E., Group Decision Making and Disagreement, *Social Forces* 35:314–318 (1957).
14. Van de Ven, A., *Group Decision Making and Effectiveness*, Kent State University Press, Kent, Ohio, 1974.

Received 18 March 1982