

The Delphi as a Naturalistic Evaluation Tool*

Gail V. Barrington

Gail V. Barrington & Associates

Calgary

RÉSUMÉ

Cet article passe en revue l'évolution de la méthode Delphi et les premiers usages qui en ont été fait. Les avantages de son utilisation sont exposés, et ses désavantages explorés. Enfin, il présente l'utilisation actuelle de la méthode Delphi et fait quelques suggestions destinées à en accroître l'usage.

ABSTRACT

This article reviews the evolution of the Delphi method and its early uses. The advantages of its use are outlined and the traditional disadvantages explored. Current uses of the Delphi are presented, and some suggestions advanced for its further use.

The increasing interest in the use of qualitative methodologies to evaluate programs has resulted from the legitimization of non-statistical approaches by such researchers as Guba and Lincoln (1981). Their elaboration of the naturalistic paradigm, in contrast to the scientific one, has encouraged evaluators to examine a variety of normative techniques which may have been overlooked before. Although the Delphi method was dismissed as "unscientific" by several writers in the seventies (Weaver, 1970; Sackman, 1974; Brooks, 1979), it should be reevaluated for its usefulness as a qualitative tool for the more creative evaluators of the late 1980s.

This article provides a brief overview of the evolution of the Delphi method, examines the advantages and disadvantages of its use as a naturalistic evaluation tool, proposes the use of a Delphi as one methodology in a sequential evaluation model, presents a current example of Delphi use in an evaluative setting, and advances some pointers to consider when designing a Delphi.

Definition

The Delphi method uses a series of questionnaires, open-ended at first, more focused at each subsequent stage, which allows a group of individuals, known as panel members, to probe a complex issue anonymously over a period of time.

The particular type of Delphi which is explored in this paper is known as a policy Delphi which usually generates a great deal of written comment in response to the open-ended questions posed at the first stage, or Round 1, as it is sometimes called. These comments are then sorted, analyzed, and summarized by the monitor team conducting the study, and are sent back

* A Draft Version of this Paper was Presented at Evaluation '86, The Canadian Evaluation Society Annual Conference, April 30, 1986, Banff, Alberta

in Round 2 for panelists to quantify their responses from Round 1 according to some rating scale such as relative importance, desirability, feasibility, etc., and sometimes further comment is solicited. In Round 3 rated responses and new comments are returned for final judgements and final remarks. The results are analyzed and final outcomes returned to panel members as well as to those who commissioned the study.

The policy Delphi is usually seen as one process among a variety of activities related to the policy or issue area under review, and may grow out of other data collection components, or may feed into other types of analysis or committee work. The outcome of a policy Delphi can never be predicted as opinions may converge or diverge. In either case, issues are identified and opinions probed to a degree not possible in any single-staged data collection process. In addition, participants may experience a degree of stimulation from the interactive dialogue which occurs in the Delphi that is both instructive and challenging.

Evolution

The Delphi had a rather ominous beginning in the early 1950s at the Rand Corporation. A study entitled "Project Delphi" was conducted by Helmer and Dalkey to poll experts about possible industrial targets which the Soviets might select and to estimate the number of atomic bombs required to reduce these targets by a prescribed amount.

In the sixties, the Delphi was used for technological forecasting in the aerospace and electronics industries. Later, in the seventies, its use broadened into that of a flexible planning tool used by government as well as industry. The growing need to incorporate subjective or judgemental information into the analysis of complex issues led to its use in education, health, urban renewal, transportation, environment, and management science. In 1969 the policy Delphi was developed by Turoff (Linstone and Turoff, 1975) in order to generate and identify opposing views on policy resolution issues.

However, charges began to mount that the Delphi method was unreliable, unscientific, and misleading which tended to curtail its use. The strongest critic was Sackman (1974) who, also under the auspices of the Rand Corporation, published a scathing attack on the value of the Delphi method. By his use of the scientific paradigm to judge the methodology, it failed miserably. But reexamined in a naturalistic context, the Delphi remains a vibrant and useful qualitative tool.

Advantages

The Delphi has a number of advantages for the naturalistic evaluator.

1. Polls Multiple Realities

Unlike the scientific inquirer who views phenomena as discrete subsystems, Guba and Lincoln (1981:57) suggests that the naturalistic inquirer acknowledges the multiplicity of participant perceptions and values with which he works. The Delphi allows the study monitor to select and poll a

panel which represents all desired stakeholder groups. Because the interaction is anonymous, the emotional and psychological dynamics of group interaction are minimized. Participants respond from within their own context, completing their questionnaire at their leisure and in their own words, thus conveying strongly their own world view.

A naturalistic evaluator who is seeking rich descriptive data from a variety of viewpoints will find, therefore, that the Delphi is an appropriate tool. The final report has a texture of authenticity which may be absent in more sanitized studies which do not provide for extensive individual response. This real-life graininess tends to enhance credibility with those who commissioned the study and may thereby increase the likelihood of implementation of study recommendations.

2. Inquirer and Subjects Interact

While the scientific inquirer takes great pains to divorce himself from any contact with his subject, the naturalistic inquirer acknowledges the inevitability of subject-inquirer interaction and uses it to advantage. The Delphi monitor can elicit discussion on a specific issue, zooming in with as fine a focus as is required to attain illumination. The panelists' responses are manipulated in a variety of ways — edited, reorganized, regrouped, and classified, particularly in Round 1. In Rounds 2 and 3 panelists can react to monitor interpretation, they can make their own judgements, they can reconsider and reevaluate their own and others' judgements. This interactive and cyclical process can turn into a shared journey of discovery.

3. Focuses on Differences as Well as Similarities

As the goal of scientific inquiry tends to be generalizability of outcomes, the methodologies employed tend to focus on similarities. However, naturalistic inquirers acknowledge both multiple realities and context, and therefore they tend to search for differences as well as similarities. While early Delphis tended to force convergence of opinion (Weaver, 1970), more recent varieties, such as the policy Delphi (Linstone and Turoff, 1975) avoid convergence as a necessary outcome altogether.

Disadvantages

Over the years, a number of disadvantages of the Delphi have been proposed. Besides Weaver's criticism of its forced convergence and lack of clarity regarding participants' reasons for making judgements, Brooks (1979) elaborated the hazards of panel selection. But the expansive approach of the policy Delphi responds to these charges. Divergent views are recorded and rationales for judgements are solicited. In an evaluation context it is unlikely that panelists will be uniformed or inbred as they are program stakeholders representing as wide a variety of views as possible.

The greatest disadvantage of the Delphi, as Sackman pointed out, was lack of rigor. He felt that test design, sampling, use of experts, and interpretation of findings were unreliable and unscientific. However, he limited his comments to the traditional Delphi and ignored such later varieties as the policy Delphi. Sackman was refuted by Coates (1975) who stressed that the

Delphi's role was not to provide high reliability consensus data but rather to cajole participants into challenging their own assumptions about complex issues.

According to Guba (1986) rigor is an inappropriate criterion for a naturalistic evaluator. Rather, a naturalistic study should develop an internal validity of its own based on feedback from study participants. Trustworthiness and authenticity become more appropriate in the naturalistic paradigm. A Delphi provides several opportunities for perception checking through its iterative process so that revisions can be made to the evaluator's construction of reality as the study progresses.

For management of the Delphi data, which tend to be voluminous, content analysis techniques can be employed such as unitizing and coding responses, sorting, categorizing, and summarizing. In addition, techniques suggested by Holsti (1969:113) such as frequency analysis, scaling, and contingency analysis can be employed.

Use can be made of an auditor to review decisions made by the monitor team to offset any tendency toward unconscious bias. Team members can debrief each other at the end of a day's work by probing for inconsistencies and demanding rationales for decisions made. Finally, the Delphi can be used as one of a sequence of evaluative activities; the outcomes of a Delphi can provide support for other, possibly more traditional, data collection techniques. If all these safeguards do not quell the sceptics, one can conclude with Guba and Lincoln (1981:27), "Whatever degree of apparent objectivity may be lost is more than compensated for by the continuously emerging insights that naturalistic methods produce."

The Delphi In A Sequential Evaluation Model

Delphis in the past have been combined with a variety of other non-traditional methodologies such as scenario writing, trend analysis, and simulation forecasting. However, for evaluation purposes, it can be beneficial to use the Delphi in conjunction with such traditional methodologies as quantitative analysis and interviewing. The complementarity of different methodologies, according to Greene (1985:10-11), can cause each set of results to enrich and clarify those which have come before in an emergent discovery process.

A Delphi could be employed at a number of different phases in an evaluation study. It could be used initially to identify issues and poll stakeholder opinions. In this way, the evaluator could ascertain the breadth of the spectrum with which he was working before moving on to other methodologies tailored to these preliminary findings.

Another use of the Delphi, in an evaluative context, is towards the end of the data collection process where fact finding or corroboration is minimized in favour of a focus on problem solving related to key outstanding issues. The following model outlines such a sequential data collection approach.

Sequential Program Evaluation Model Employing A Delphi With Other Methodologies

Stage	Methodology	Suggested Subjects
Stage 1	Survey Questionnaire	Users and service-givers
Stage 2	Quantitative Analysis Development of Interview Format	
Stage 3	Interviews	Key stakeholders
Stage 4	Content Analysis Identification of Key Issues	
Stage 5	Delphi	Selected stakeholders including users and service-givers
Stage 6	Drafting of Study Recommendations	

The results of this Delphi might, if appropriate, be fed into some sort of committee process if further crystallization were necessary prior to the recommendation phase, or they might lead to further issue-specific interviews. The results of such a study design would be rich in detail and grounded in reality, ensuring greater acceptance of study recommendations and increasing the likelihood of their implementation. The whole process is flexible according to the evaluator's needs. Obviously time is a critical factor in methodology selection.

Generally a Delphi can take up to six months to complete, which includes a lot of "down time" waiting for returns. It has been suggested that a more controlled time frame would result from the use of electronic mail rather than the traditional print-and-post format. With the proliferation of micro-computers and the ease with which networking can be accomplished, there is no reason why a Delphi could not be conducted electronically in a very efficient manner. Further, content analysis of responses could be done on a computer using both key word selection in a data base system and simple editing devices available in any word processing package. The era of cutting and pasting is definitely coming to a close.

A Current Evaluation Delphi

The author is currently acting as external evaluator for a five-year educational experiment being conducted in three senior high schools in a major Western Canadian city. The study has been guided by a series of seven eval-

uative goals developed by a steering committee involving an annual cycle of data collection activities which include the following:

1. Questionnaires to students, teachers and administrators which measure attitudes and perceptions regarding the project.
2. On-site observation on a regular basis in each of the schools.
3. Curriculum evaluations conducted by provincial curriculum consultants.
4. Sequential interviews of a 5-10% sample of students, 50-100% of teachers, and 50-100% of administrators based on questionnaire results and interview results of each of the groups in turn.
5. Academic achievement evaluation which compares final grades of project students with control groups.

Each year a formative evaluation report is prepared based on the data collected by the above-mentioned methodologies and, at the end of the fifth year, a summative evaluation report will be prepared.

However, it was felt that the restrictions placed on this study by stakeholder groups' requirements, not to mention the sheer size of the project, exclude a great deal of useful information of a more idiosyncratic nature. Further, each of the three schools involved in the project has developed its own project model and, not surprisingly, believes that its approach is both unique and the most effective. There has been little communication among school staffs, although from the evaluator's perspective they share many commonalities. In addition, it appeared that the multitude of ideas, procedures and materials generated by these staffs over the five-year period would languish once the project was over. Another school district attempting a similar project would have to start again.

And so a Delphi was designed for inclusion in the final year's evaluative activities with the twin goals of asking some open-ended evaluative questions and collecting suggestions and materials which might be included in a planning guide for the future. The questions asked will be similar in nature to the following:

1. What major benefits have accrued to your school as a result of the project?
2. If you were doing this project over again, what things would you like to see done differently?
3. What essential components would another school need to implement a similar program?

It is anticipated (as much as one can in any Delphi) that along with a great deal of one-of-a-kind data, some common threads will become visible, not only to the evaluator, but to the panel members from each school staff. Some sharing should take place, particularly with regard to Question 2 which may be instructive and stimulating for panel members. It is hoped that Question 3 will elicit materials, procedural tips and suggestions to

form a data bank containing some of the fugitive kinds of information which can be vital to a fast start-up in another school. The iterative nature of the Delphi will allow time for reflection, discussion and revision as panelists and team members move toward a more sophisticated construction of reality. The process should be exciting; the product worthwhile in its own right and enriching to final summative activities.

Pointers When Planning A Delphi

Because the outcomes of a Delphi are unpredictable, the following points should be kept in mind to minimize risk:

1. Limit discussion to two or three key issues.
2. Select panelists with care, ensuring representativeness and commitment.
3. Control timeframe tightly, using telephone reminders for late submissions, rewards for early ones.
4. Limit generation of data by requesting panelists to condense comments (e.g., specify response length).
5. Minimize monitor bias by setting clear ground rules in advance, by having more than one monitor, and by involving a study auditor.

It has been shown that the Delphi has value as an investigative tool for the naturalistic evaluator. Its flexible, emergent design allows the polling of the realities of a number of stakeholder representatives while minimizing group dynamics. Its multi-staged structure encourages reflection upon complex issues and allows both monitors and panelists to interact over time. Traditional arguments about the lack of rigor of the Delphi have been largely dispelled by a shift from the scientific to the naturalistic paradigm and by the use of the Delphi in a sequential model of data collection. Surely this valuable tool deserves a place in the repertoire of today's program evaluator.

References

- Barrington, G.V. (1981) "The impact of environmental forces on Alberta Community Colleges 1980-1990" Unpublished doctoral dissertation, Department of Educational Administration, University of Alberta, Edmonton
- Barrington, G.V. (1986) "Evaluating English as a Second Language: A Naturalistic Model" *TESL Canada Journal* vol. 3, no. 2
- Coates, J.J. (1975) "In defense of Delphi: A review of *Delphi Assessment, Expert Opinion, Forecasting, and Group Process* by H. Sackman" *Technological Forecasting and Social Change* 7:193-194
- Greene, J.C. (1985) "Validity, management, and utilization in mixed-methods evaluation designs" Paper presented at the Joint Meeting of the Evaluation Research Society, the Evaluation Network, and the Canadian Evaluation Society, Toronto

- Guba, E. (1986) Personal correspondence with the author, Bloomington, Indiana, May 6
- Guba, E. & Lincoln, Y.S. (1981) *Effective Evaluation: Improving the Usefulness of the Evaluation Results Through Responsive and Naturalistic Approaches* San Francisco: Jossey-Bass
- Holsti, O.R. (1969) "Introduction to aspects of inference from content data" In G. Gerbner, O.R. Holsti, K. Krippendorff, W.J. Paisley, & P.J. Stone (eds.) *The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques* New York: John Wiley and Sons
- Linstone, H.A., & Turoff, M. (eds.) (1975) *The Delphi Method: Techniques and Applications* Reading, Massachusetts: Addison-Wesley
- Sackman, H. (1974) *Delphi Assessment, Expert Opinion, Forecasting, and Group Process* Rand Report R-1283-PR Santa Monica, California: Rand Report Abstracts, Spring
- Turoff, M. (1970) "The design of a Policy Delphi" *Technological Forecasting and Social Change* 2:149-171
- Turoff, M. (1975) "The Policy Delphi" In H.A. Linstone & M. Turoff (eds.) *The Delphi Method: Techniques and Applications* Reading, Massachusetts: Addison-Wesley
- Weaver, W.T. (1970a) *The Delphi Method: Background and Critique* Report I (Working Draft) Syracuse: Educational Policy Research Center
- Weaver, W.T. (1970b) *The Delphi Method: Research Design and Analysis of Results* Report II (Working Draft) Syracuse: Educational Policy Research Center