THE DELPHI METHOD: A REVIEW OF METHODOLOGY AND AN APPLICATION IN THE EVALUATION OF A HIGHER EDUCATION PROGRAM

Alice L. Boberg
Sheryl A. Morris-Khoo
University of Calgary
Calgary, Alberta

Abstract: The first section of this article details the history of the Delphi method, explains the method, and describes when it should be used. The advantages and disadvantages of the method are reviewed. The second section discusses the application of a Delphi method in a review of a higher education program.

Résumé: La première partie de cet article détaille l'histoire de la méthode Delphi, donne une explication de la méthode, et décrit la façon de l'employer. Les avantages et les désavantages de la méthode sont examinés. L'application de la méthode Delphi est adressée dans la seconde partie dans le contexte d'un programme d'éducation supérieure.

The Delphi method was developed in 1953 by Norman Dalkey of the RAND Corporation and Olaf Helmer of the Institute for the Future. In the original study it was used to obtain reliable consensus of a group of military defense experts on the issue of an atomic bomb attack. The purpose of the Delphi method was to elicit perceptions or judgments held by 'experts' knowledgeable in a specialized area. The opinions were then refined through subsequent reviews with the eventual outcome being a converging consensus about a given subject. The rationale for this technique was the age-old adage 'Two heads are better than one' (Dalkey, 1969, p.6). The original method included follow-up interviews with each respondent. After its first application in 1953, the use of the method was limited until 1964 when it was tested as a forecasting method. Since the mid-1960s numerous Delphi studies have been conducted, although rarely following the original format. A 'standard Delphi format' evolved, usually consisting of four rounds of anonymous inquiry administered through mailed questionnaires for the purpose of increased

refinement and specificity (Cookson, 1986, p.5). Another type of Delphi, the Policy Delphi, is quite different from the conventional Delphi and was developed in 1969. Rather than seeking consensus Policy Delphi encourages not only convergent but divergent responses, and is useful when exploring and arguing alternatives. To date many disciplines including government, corporate, educational, and medical institutions have used some form of the Delphi (Barrington, 1986; Cookson, 1986; Helmer, 1983; Hencley & Yates, 1974; Jillson, 1975; Linstone & Turoff, 1975; Somers, Baker & Isbell, 1984). The ERIC listings from January 1983 to June 1989 report 131 articles on the Delphi technique.

WHAT EXACTLY IS THE DELPHI METHOD?

The 'standard Delphi method' or usual format consists of a series of questionnaires (called 'rounds') about some important question or problem, which are usually mailed to a sample of individuals who are considered 'experts' in a particular area. The anonymous responses in each round are analyzed and summarized by the individual(s) conducting the study, and are returned to the respondents in subsequent rounds for further consideration and response. The eventual outcome is the blending of diverse opinions or judgments, frequently into distinct majority and minority groupings (Anderson, Ball, Murphy & Associates, 1975; Barrington, 1986; Hencley & Yates, 1974; Sackman, 1975).

Many variations of the Delphi method have been used, including using fewer or more than four rounds, changes in the question format, and variations in the structure of the expert group. Other variants worth mentioning include: (1) using an extra, preliminary round that is open-ended and intended to help delineate the subject matter of inquiry; (2) a mini-Delphi where participants are brought together and secretly write their opinions on an issue which are then tabulated and opened for discussion and debate. This is followed by a secret ballot vote and the median of the responses is accepted as the group's consensus; (3) ranking objects or ideas; and (4) procedures have been refined to use computers to generate, conduct, and tabulate Delphi questionnaires (Helmer, 1983; Hencley & Yates, 1974).

WHEN SHOULD THE DELPHI METHOD BE USED?

Linstone (1978) identifies two circumstances where the Delphi technique is appropriate to use. The first is when "the problem does not lend itself to precise analytical techniques but can benefit from subjective

judgments on a collective basis"(p. 275). The second circumstance is when "individuals who need to interact *cannot* be brought together in a face-to-face exchange because of time or cost constraints" (Linstone, p. 275). Dalkey (1969) and Uhl (1983) believe that this method is also useful when you *would not* want to bring a group together because of individual or subgroup dominance.

The following are possible applications for the Delphi method:

- assessment of present conditions or issues,
- market, societal, and long-term technological forecasts,
- normative forecasts (formulation of goals and policies),
- the comparative evaluation of alternative action programs (planning), and
- management decision making in the private and public sectors (Helmer, 1983; Hencley & Yates, 1974).

ADVANTAGES AND DISADVANTAGES

The Delphi method has been described as a useful tool by proponents of the method, but it has also come under heavy criticism by other researchers. Some advantages of the Delphi method include: (1) individuality because of the isolated generation of ideas; (2) freedom of expression because of the anonymous nature of the responses; (3) deep exploration and elicitation of fine perceptions and distinctions because of the repeated rounds of feedback; (4) the provision of equal consideration of contributions of all the respondents; (5) minimal interference from psychological distractions of open-forum (face-to-face) discussions that avoids direct confrontation, group think, or group dominance; and (6) the efficiency of involving a large number of people in formulating group opinion without meetings, which is particularly important when participants are geographically isolated (Chaney, 1987; Cookson, 1986; Dalkey, 1969; Helmer, 1983; Hencley & Yates, 1974; Linstone & Turoff, 1975; Shortridge et al., 1988; Somers, Baker & Isbell, 1984; Uhl, 1983).

The criticisms lodged against the Delphi method include: (1) its tendency to produce manipulated consensus; (2) the lack of rigor as found in other social science research methods, including the lack of reporting of statistical significance, questionable validity and test reliability; (3) the considerable labor required, including mailings, tabulations of voluminous amounts of data, record keeping; and (4) the length of time required by this method because of the necessity to wait for questionnaires to be

returned (Anderson, Ball, Murphy & Associates, 1975; Barrington, 1986; Sackman, 1975). Proponents of the method argue that rigor is an inappropriate criterion for naturalistic inquiry. The trustworthiness and authenticity developed through perception-checking opportunities inherent in the method are appropriate.

BACKGROUND TO AND PURPOSE OF THE STUDY

The Department of Educational Policy and Administrative Studies in cooperation with the Faculty of Continuing Education of the University of Calgary offers a Master of Education degree and a professional diploma in Adult and Community Education (A/CE). The A/CE program has been operational since 1984 and a review was requested by the program's Advisory Committee in 1989.

The committee recommended that a Delphi method be used for the review to ascertain perceptions held about the program by relevant stakeholder groups, including both the faculty and students (which was essentially an assessment of the present conditions or issues). The Delphi method was chosen because of the need to acknowledge the multiplicity of perceptions and values and yet determine any group consensus. Some members of the population were not on campus during the study so the mailed questionnaire format was both cost- and time-effective.

DELPHI METHODOLOGY PUT INTO PRACTICE

One of the first steps in a Delphi design is to decide who will conduct the study. It was the decision of the committee to have a 'neutral' professor and graduate student conduct the study, both of whom were *not* involved in the program and had background experience in program evaluation. A six-month time frame was identified.

The next two steps are the exploration and identification of the subject under review, and the selection of panel(s) of 'experts' (Adilman, 1984; Cookson, 1986; Uhl, 1983). In this case, the committee indicated their desire to review perceptions held about the program by both students and professors. Although these participants are not 'experts' in the common definition of the word, they are qualified to make judgments about the program because of their involvement with the program. Administrative records were reviewed by a Department secretary to identify the population of students and professors involved with the

program; due to the limited numbers, the entire population was sur-
veyed rather than selecting a sample. The population consisted of 47
students and 16 faculty, for a total of 63. Participants were assigned a
number which appeared on response forms to ensure anonymity and to
facilitate follow-up as required, and the secretary was the only person
who had access to the list of participants and code numbers.

Three 'rounds' of questionnaires were used due to the short time frame.
It was decided to use an open-ended first round to evoke all concerns and
opinions regarding the A/CE program. The questions were:

1. I feel the strengths of the program are . . .
2. I feel the weaknesses of the program are . . .
3. What would you recommend for future changes?

The information generated in Round One was used to create the
questionnaire for Round Two. Round Three consisted of reorganizing
the items of Round Two. Examples are presented and discussed in the
'Results' section.

A cover letter, in addition to the questionnaire, was sent with each
round. The first round cover letter included a description of the Delphi
method, why we were reviewing the program, directions for question-
naire completion, assurance of anonymity for each participant and that
responses would be reported on a group basis only, and the due date for
return of the questionnaire. The second and third round cover letters
thanked the respondents for their participation, explained the format of
the questionnaire for the particular round, gave directions for comple-
tion and a due date, and stressed anonymity and group-reporting of
results. Following recommendations found in the literature, each letter
was personalized, hand-signed by the researchers, and a return postage-
paid envelope provided (Adilman, 1984; Altschuld & Lower, 1984;
Anderson, Ball, Murphy & Associates, 1975; Baumgartner & Heberlein,
1984; Dillman, Dillman & Makela, 1984; Lockhart, 1984; McKillip,
1984; Russo, 1984; Sudman & Bradburn, 1984).

DISCUSSION

The methodology presented is replicable to other programs and settings;
however, the results are not discussed in detail because they are setting-
specific. Examples of the results are presented for the purpose of sup-
porting the discussion.

Table 1
Response Rates

	Round One N	Round Two N	Round Three N
Students	39/47	35/38	29/34
Professors	15/16	10/12	10/10
Total	**54/63 (86%)**	**45/50 (90%)**	**39 /44 (89%)**

Round One was sent to the entire population of 63, and due to attrition, Round Two was distributed to 50 respondents (79% of N) and Round Three to 44 respondents (70% of N). The attrition rate was greatest from Round One to Round Two. Reasons for attrition included the inability to contact the participant (because of change of address, or sabbatical leave), non-response, or request for withdrawal from the study. It should be noted that students who had completed the program had a higher rate of attrition than students presently in the program.

The response rates were very substantial. In Round One 86% of the questionnaires mailed were returned, 90% in Round Two, and 89% in Round Three (see Table 1) . In the student group it was found that response rates were highest for students currently in the program.

The substantial response rates were partly due to rigorous follow-up conducted immediately after questionnaire return deadlines. The secretary utilized follow-up telephone contact for local participants, and a letter follow-up for out-of-town participants who had not responded. Table 2 shows the type and number of follow-ups conducted for each round. The percentages noted indicate the success of the follow-up effort. For example, in Round One telephone calls were made to 13 non-respondents, and this effort resulted in 69% returning their questionnaires.

Table 2
Type of Follow-ups for Each Round

Type of Follow-up	Round One	Round Two	Round Three
Telephone	13 (69%)	5 (60%)	8 (75%)
Letter	7 (71%)	7 (57%)	4 (50%)
Total	**20**	**12**	**12**

In this study telephone follow-up was slightly more effective in securing a response (returned survey) than letter follow-up. However, the two follow-up methods were used for different locales of participants.

ROUND ONE

The first round was an open-ended questionnaire asking respondents to comment on the strengths and weaknesses of the program, as well as their recommendations for change. Round One produced voluminous amounts of data. The data was comprised of statements made by one individual at a time and statements which were expressed by numerous individuals. For example, strengths of the program cited by more than 15 respondents included the short, intensive classes during the summer; the visiting professors; and the climate of group supportiveness. Responses to the questions of weaknesses and recommendations tended to be more individual, covering more diverse topics.

One of the disadvantages of the method includes the difficulty of working with large amounts of data. The authors chose to analyze the data independent of each other at first; reviewing the responses, categorizing them by themes, and calculating frequency analysis. These preliminary results were then reviewed by the authors together. This was done to ensure equity of interpretation of the qualitative data and to avoid bias.

We identified four main themes in the data from Question One (strengths of the program) and Question Two (weaknesses of the program). They were: program structure issues, coursework issues, issues about the professors (cited by students), and issues about the students (cited by professors). The recommendations noted in Question Three often stemmed from the responses of the first two questions (see Table 3).

Table 3
Categories

I.	Program Structure	- Strengths	- 14 statements
		- Weaknesses	- 22 statements
II.	Coursework	- Strengths	- 9 statements
		- Weaknesses	- 14 statements
III.	Professors	- Strengths	- 6 statements
		- Weaknesses	- 11 statements
IV.	Students	- Strengths	- 2 statements
		- Weaknesses	- 5 statements
V.	Recommendations		- 27 statements

Examples are:

- Program Structure (strength) - facilitation of studying while continuing to work; the diversity of student backgrounds/ disciplines;
- Program Structure (weakness) - no time for critical reflection during Summer Institute; management of pre-session materials;   ·
- Coursework (strength) - the integration of assignments with work-related situations; degree of learner-determined content;
- Coursework (weakness) - not enough emphasis on program design and delivery; not enough emphasis on city education;
- Professors (strength) - those professors that treat students as adult learners; the varied expertise of professors;
- Professors (weakness) - professors who do not demonstrate internalization of the principles of Adult Education; professors who don't model the behaviors taught;
- Students (strength) - motivated students; mature/experienced students;
- Students (weakness) - students who continue to work during the Summer Institute and are not prepared for class;
- Recommendations - develop a clear statement about the nature, goals and objectives of the program; better organization and delivery of pre-session materials.

ROUND TWO

Statements for the second questionnaire were created and refined from what the participants had said in Round One, frequently using exact jargon or wording. This was done to enhance the validity of the data. Participants were asked to vote on whether they agreed, disagreed, or had no opinion on each statement. Due to the nature of the responses in the categories of Students and Professors, the professors received a set of items on students, and the students received a set of items on professors. Both groups received the items on program structure, coursework, and recommendations.

The second round resulted in consensus on 42 of the 110 statements (see Table 4). Consensus was assessed when the majority of votes were either for agreement, disagreement, or no opinion.

We had two items which some participants felt were ambiguous. They were: (a) establish a sounder theoretical base in courses and (b) establish

more emphasis on practical content (from the Recommendations section). Rather than change the items in Round Three, which could possibly influence the participant's vote, we chose to identify those items and ask the participants to clarify their decisions if necessary in the comments section.

ROUND THREE

The means for each statement in Round Two were computed. The statements were then reorganized into groups under each category; those statements with majority of agreement, majority of disagreement, majority of no opinion, and no consensus. The respondents were once again asked to vote on each statement. Final data analysis involved computing the voting scores.

As expected with the Delphi method, Round Three results showed a movement towards greater consensus among participants (see Table Four). The largest shifts towards consensus were in the categories of Professors (weaknesses) and Program Structure (weaknesses), 55% and 41% respectively. A criticism of the Delphi method is the question of manipulated consensus. As an example, the quality of instruction, was noted as a Professor weakness in Round One. In Round Two, 12 respondents agreed that this statement was true, 18 disagreed that the quality of instruction was a weakness, and 2 had no opinion. In Round Three,

Table 4
Second Round Consensus

Category	Round Two	Round Three	% of shift
Program Structure			
- strengths	8/1	11/14	22%
- weaknesses	4/22	13/22	41%
Coursework			
- strengths	4/9	7/9	34%
- weaknesses	8/14	9/14	7%
Professors			
- strengths	5/6	5/6	0%
- weaknesses	1/11	7/11	55%
Students			
- strengths	2/2	2/2	0%
- weaknesses	2/5	2/5	0%
Recommendations	8/27	12/27	14%
Total	**42/110**	**68/110**	

only 2 respondents agreed, 25 disagreed, and again 2 had no opinion. The shift in consensus to disagreeing that the duality of instruction was a professor weakness is clearly evident in this example. The question remains whether this shift was due to 'manipulative' forces by seeing how others voted, or simply a shift in opinion after reflecting on the statement.

Although we had shifts towards greater consensus in some categories, we also had three categories in which the voting remained the same, thus no difference in consensus between the rounds.

A report on the findings of the Delphi study and our recommendations for further program review was submitted to the Advisory Committee in June 1989. Following a meeting held to discuss the report, a criticism expressed by the committee faculty was that students did not have the expertise, to make judgements about the program, particularly on the topics of course workload and curriculum content. This directly opposes the committee's wish to ascertain perceptions about the program from both stakeholder groups.

Some of the weaknesses identified in the study have been addressed; especially registering and receiving pre-session assignments for out-of-town students. Some weaknesses are not easily addressed. For example, the students indicated that the four-week session does not allow time for critical reflection but the students are full-time employees that either get released time or have to use vacation time to attend. Extending the time would only make it more difficult for students to attend. Another example is the issue of course content. Are adult students experts who can contribute to course development or are faculty the experts who should decide course content? These are some of the issues that need further study.

CONCLUSION

The purpose of our study was to identify the perceptions held about the A/CE program by both students and professors associated with the program. Because we were assessing the present state of the program from various stakeholder groups the Delphi method was an appropriate choice of tool. This method allowed us to examine the subjective judgments of a large group, considering the contributions of all the participants of the study. The final round was sent to 70% of the population, with 89% responding which we felt was sufficient to be deemed representative opinion of the population.

We did three rounds in total, with the first round being open-ended which was valuable in encouraging freedom of expression thus stimulating a 'catharsis' of opinions and issues without restriction of topic. The mailed questionnaire process encouraged individuality of thought, accessibility to out-of-town participants, avoided possible 'group think' or group dominance, and protected the student group from fear of repercussions. As noted in our results, the students currently in the program were committed to the survey process.

We found the six-month time frame adequate for three rounds of questionnaires. We did not have lengthy delays waiting for questionnaires to be returned because of the rigorous follow-up, which also was beneficial in increasing the response rate. The method was labor-intensive, particularly the analysis of Round One data and creation of the Round Two questionnaire. A solution to this problem could have been to originally sample from the population, but we knew that we would lose some participants through attrition, and the more representation of the total population we had, the better our ability to make generalizations. So we chose to risk working with a large volume of data rather than a small sample of people.

It was interesting that even with isolated generation of ideas in Round One, we had some topics or issues raised by more than fifteen people. This fact lent some reliability to the process. Using exact wording or jargon in formulating the second round questionnaire enhanced the validity of the data. That is to say, *we* did not generate the data, the participants did, and responded to it in subsequent rounds. The second and third rounds allowed the participants to fine-tune their perceptions. This is what the opponents of the method refer to as 'manipulation.' In our study there were shifts in the degree of consensus on certain items, but no change in the voting on other items.

This study has shed light on the program's strengths, weaknesses, and recommendations for change. We would recommend the use of the Delphi method as a *preliminary step* in the evaluation of higher education programs.

REFERENCES

Adilman, A. (1984, March). *Needs assessment in family planning: A guide for program development.* Edmonton, AB: Alberta Social Services and Community Health.

Altschuld, J.W., & Lower, M.A. (1984). Improving mailed questionnaires: Analysis of a 96 percent return rate. In D.C. Lockhart (Ed.), *New directions for program evaluation,* No. 21 (pp. 5–18). San Francisco, CA: Jossey-Bass.

Anderson, S., Ball, S., Murphy, R., & Associates. (1975). *Encyclopedia of educational evaluation.* San Francisco, CA: Jossey-Bass.

Barrington, G.V. (1986). The delphi as a naturalistic evaluation tool. *The Canadian Journal of Program Evaluation, 1,* 81–88.

Baumgartner, R.M., & Heberlein, T.A. (1984). Recent research on mailed questionnaire response rates. In D.C. Lockhart (Ed.), *New directions for program evaluation,* No. 21 (pp. 65–76). San Francisco, CA: Jossey-Bass.

Boberg, A., & Morris-Khoo, S. (1989). *Program review of the adult and community education (A/CE) program.* Unpublished manuscript, University of Calgary, Calgary, Alberta.

Chaney, H.S. (1987). Needs assessment: A delphi approach. *Journal of Nursing Staff Development, 3,* 48–53.

Cookson, P.S. (1986). Charting the unknown: Delphi and policy delphi strategies for international cooperation. *International Journal of Lifelong Education, 5,* 3–13.

Dalkey, N.C. (1969). *The delphi method: An experimental study of group opinion.* Santa Monica, CA: RAND.

Dillman, D.A., Dillman, J.J., & Makela, C.J. (1984). The importance of adhering to details of the total design method (TDM) for mail surveys. In D.C. Lockhart (Ed.), *New directions for program evaluation,* No. 21 (pp. 49–64). San Francisco, CA: Jossey-Bass.

Helmer, O. (1983). *Looking forward: A guide to futures research.* Newbury Park, CA: Sage.

Hencley, S.P., & Yates, J.D. (Eds.). (1974). *Futurism in education methodologies.* Berkeley, CA: McCuthan.

Jillson, I.A. (1975). The national drug abuse policy delphi: Progress report and findings to date. In H.A. Linstone & M. Turoff(Eds.), *The delphi method: Techniques and application* (pp. 124–159). Reading, MA: Addison-Wesley.

Linstone, H.A. (1978). The delphi technique. In J. Fowles (Ed.), *Handbook of futures research* (pp. 273–300). Westport, CT: Greenwood.

Linstone, H.A., & Turoff, M. (Eds.). (1975). *The delphi method: Techniques and applications*, Reading, MA: Addison-Wesley.

Lockhart, D.C. (1984). The stages of mailed questionnaire returning behavior. In D.C. Lockhart (Ed.), *New directions for program evaluation,* No. 21 (pp.89–98). San Francisco, CA: Jossey-Bass.

McKillip, J. (1984). Applying attitude theories to the return of mailed questionnaires. In D.C. Lockhart (Ed.), *New directions for program evaluation,* No. 21 (pp.77–87). San Francisco, CA: Jossey-Bass.

Russo, J.R. (1984). Using mailed questionnaires in negotiation consultation. In D.C. Lockhart (Ed.), *New directions for program evaluation,* No. 21 (pp. 19–31). San Francisco, CA: Jossey-Bass.

Sackman, H. (1975). *Delphi critique.* Lexington, MA: Lexington.

Shortridge, L., Doswell, W., Evans, M., Levin, R., Millor, G., & Carter, E. (1988). Preliminary report rounds I & II of the 1988 delphi survey of research priorities for New York state. *The Journal of the New York State Nurses Association, 19,* 22–23.

Somers, K., Baker, G., & Isbell, C. (1984). How to use the delphi technique to forecast training needs. *Performance and Instruction Journal, 23,* 26–28.

Sudman, S., & Bradburn, N. (1984). Improving mailed questionnaire design. In D.C. Lockhart (Ed.), *New directions for program evaluation,* No. 21, (pp. 33–47). San Francisco, CA: Jossey-Bass.

Uhl, N.P. (1983). Using the delphi technique in institutional planning. In N.P. Uhl (Ed.), *New directions for institutional research,* No. 37 (pp. 81–94). San Francisco, CA: Jossey-Bass.