

IV.C. Experiments in Delphi Methodology*

M. SCHEIBE, M. SKUTSCH, and J. SCHOFER

Introduction

The emphasis in the Delphi literature to date has been on results rather than on methodology and evaluation of design features. The other articles in this chapter do address the latter aspects. Still, quite a number of issues remain unsolved, particularly those concerned with the details of the internal structure of the Delphi. For example, the way in which subjective evaluation is measured may affect the final output of the Delphi. A number of variables enter here. Ostrom and Upshaw [1] have noted that the range of the scale provided has a marked effect on judgment. Persons playing the role of judges who estimated themselves as "relatively harsh" assigned average "sentences" of four years to "criminals" when presented with a one-to-five-year scale, and twenty-one years when presented with a 1 to 25-year scale. The difficulties involved with the selection of a suitable scale range can be solved by the employment of an abstract scale rather than one representing, for example, hard dollars or years. An abstract scale allows relative measures to be made. Abstract scales are particularly suited to the measurement of values, as for example in the development of goal weights to represent relative priorities for goal attainment.

A number of psychological scaling techniques which result in abstract scales are available. This study reports on the comparison of several scaling techniques which were tested in the context of an experimental Goals Delphi.

Another issue is that of the effects of feedback input, which form the sole means of internal group communications in the Delphi process. It is important to the design of Goals Delphis to determine the nature and strength of the feedback influence. In the experiment reported below, the impact of feedback was identified by providing participants with modified feedback data. The resulting shifts of opinion were then used as measures of feedback effectiveness.

Methods for the measurement of consensus are also considered and a redefinition of the endpoint of a Delphi is offered. Instead of consensus, the stability of group opinion is measured. This allows much more information to be derived from the Delphi, and in particular, preserves opinion distributions that achieve a multimodal consensus.

A number of Delphi studies have used high/low self-ratings of participant confidence. Evidence of the value of such confidence ratings in improving the results of the Delphi is somewhat limited, except under certain conditions of group composition [4]. In this study, the use of high/low self-confidence ratings is again evaluated, and the influence of a number of other personal descriptive variables is tested.

* This study was supported by the Urban Systems Engineering Center, Northwestern University, NSF Grant GU-3851.

Other design features include the application of short turn-around times using a computerized system for supporting inter-round analysis of the Delphi data. Although Turoff (Chapter V, C) has used a more complex, interactive computer system for this purpose, a simpler program is used here merely to accelerate accounting tasks.

Description of Procedure

The objective of this experimental Delphi study was the development and weighting of a hierarchy of goals and objectives for use in evaluating a number of hypothetical transportation facility alternatives. In the terminology suggested by Wachs and Schofer [2], goals are long-term horizontal aims derived directly from unwritten community values; *objectives* are specific, directional, measurable ends which relate directly to goals. Previous experiments by Rutherford, *et al.* [3] had indicated that Goals Delphis should be initiated by the development of objectives rather than goals, for the tendency toward upward drift in generality can be minimized if the Delphi participants are first asked to work at the more specific level. The development of goals, once the objectives have been defined, can be accomplished with much greater ease than can the reverse process.

The flow chart of Fig. 1 illustrates graphically the process by which the Goals Delphi and the design experiments were carried out. First, the initial list of objectives was generated. The process administrators presented the hypothetical transportation situation to the participants by means of a verbal description (Appendix I of this article), and a map (Appendix II of this article). The participants were then given a set of five blank 3 X 5 cards and asked to list no more than five objectives which they felt were applicable in the hypothetical situation. In all, seventy-seven objectives were submitted.

To derive goals from the list of objectives, and to eliminate the overlaps between them, a grouping procedure was followed. The process administrators first rejected those objectives that were exact duplicates of others, and assigned the remaining ones to sets. Each set represented objectives tending toward a common goal. Nine major goals were established, and these were "named" appropriately. Statements which were not strictly objectives were left out of the grouping process. The complete list of goals and objectives is given in Table 1. These goals were then returned to the participants for their evaluation. They were given the opportunity to add new objectives, and several were received and incorporated into the goal set.

Following the development of the goals hierarchy, a decision was made (largely because of time constraints) to concentrate attention at the goal level. The objectives, therefore, were not included in the weighting procedure. Objectives, however, were at all times appended to the goals related to each, so that participants would always be aware of the specific meaning of each goal. Participants were first asked to do a simple ranking of goals. As -discussed elsewhere in this paper, one of the purposes of the Delphi was to compare different scaling methods.

Participants were therefore asked to follow the ranking with a rating analysis. Nine-point Likert scales were used (0 = unimportant, 9 = very important). This type of

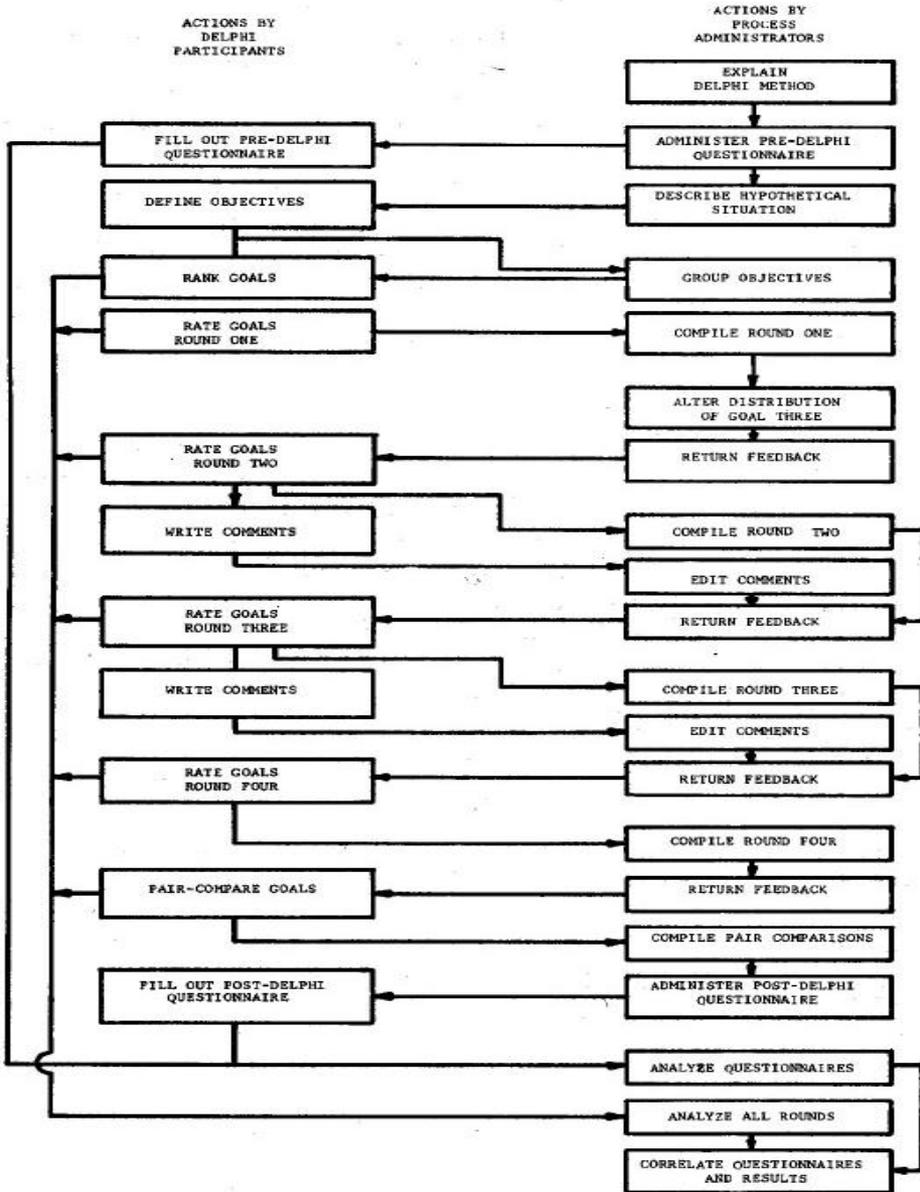


Fig. 1. The Goals Delphi experimental design.

**Table 1: Goals and Objectives Suggested for the
Hypothetical Transportation Scenario**

GOAL I:

Minimize the adverse environmental impacts of the system.

Objectives:

Minimize air, noise, and water pollution.

Minimize negative illumination and vibration effects.

GOAL II:

Preserve the recreational and social environment.

Objectives:

Provide compensation for parkland removed.

Minimize the amount of parkland taken.

Minimize the demolition of historic buildings.

Minimize the amount of non-urban land required for the new facility.

Minimize the adverse social consequences of the transportation facility location by providing adequate compensation to families and businesses (and employees).

Minimize the amount of urban land required for the new facility.

Minimize the number of residences relocated.

Minimize the disruption of existing neighborhoods, people and businesses.

GOAL III:

Minimize operating and construction costs.

Objective:

Minimize operating and construction costs.

GOAL IV:

Maximize mobility and accessibility for the urban area residents.

Objectives:

Minimize travel time.

Minimize monetary travel cost.

Reduce congestion

Locate the facility to increase mobility and accessibility in the area.

Provide sufficient mobility for all members of society.

Table 1 (cont.)

GOAL V:

Design the facility to operate as safely as possible.

Objective:

Design the facility to operate as safely as possible.

GOAL VI:

Coordinate the transportation system with land use development.

Objectives:

Maximize the flexibility of the facility so as not to hinder possible further urban development in the area.

Encourage the desired land use development pattern as stated in the land use plan.

GOAL VII:

Design the facility so that its aesthetic appeal is in harmony with the environment.

Objectives:

Use transport facilities to highlight the character of the city to increase awareness, interest, and participation by the facility users.

Design the facility to be visually pleasing to the surrounding community.

GOAL VIII:

Maximize positive economic benefits for the entire area.

Objective:

Maximize positive economic benefits for the entire area.

GOAL IX:

Minimize the adverse impacts occurring during construction.

Objective:

Minimize the adverse impacts occurring during construction.

scale was felt to be easily understood by the participants. In addition, when the ends are anchored adjectively, as in semantic differential scales, this scale is commonly found to have interval properties. Using the computer program developed for this study, the results of this first round were analyzed. The program was processed using a remote terminal; goal weights served as inputs, and histograms and various distributional statistics were produced as outputs. Frequency distributions of scores for each goal prepared by the computer were presented to the participants; along with the mean for each distribution.

Participants were asked to once again rate each goal on a nine-point scale, using the information from the previous round as feedback. In addition, those participants whose score on any goal was significantly distant from the group mean value for that goal were asked to write a few words explaining the reasons behind their positions. These statements were edited and returned in the next round. This procedure continued for a total of four rounds. The results are given in Fig. 2 and 3, which show the histograms produced in the first and the final rounds.

After the fourth weighting round, the participants were asked to perform a pair-comparison rating of all the goals. This was done to compare this scaling method with the nine-point rating scale and the ranking methods.

The initial development of the goals was accomplished during one two-hour class period. The rank ordering of the goals and the first three weighting rounds were conducted in a second two-hour period two days after the first. The fourth round took place an additional five days later, while the pair comparisons were made a week after the first weighting round.

An Experiment on the Effect of Feedback

There has been quite a bit written about the uses of feedback in the Delphi technique. Most of this, such as the work of Dalkey, Brown, and Cochran [4] at Rand, has concentrated on the effects of different types of feedback, such as written statements and various statistical measures. The effects of this feedback, particularly in the almanac-type Delphis, have been measured by comparing the accuracy of the opinions of a group given a certain feedback with that of a group given different feedback or no feedback at all. In Policy and Value Delphis the effect of feedback is evaluated by measuring the degree of consensus which is reached and the speed with which it is reached.

There seems to be very little in the literature, however, which examines the round-by-round effect of feedback or investigates the manner in which the feedback affects the distribution of scores in a particular round. In this study, it was decided to investigate these aspects of feedback, since the kind and amount of feedback used in the Delphi may be an important variable in its results. A greater understanding of the impacts of feedback might lead to better Delphi design. The method employed was to

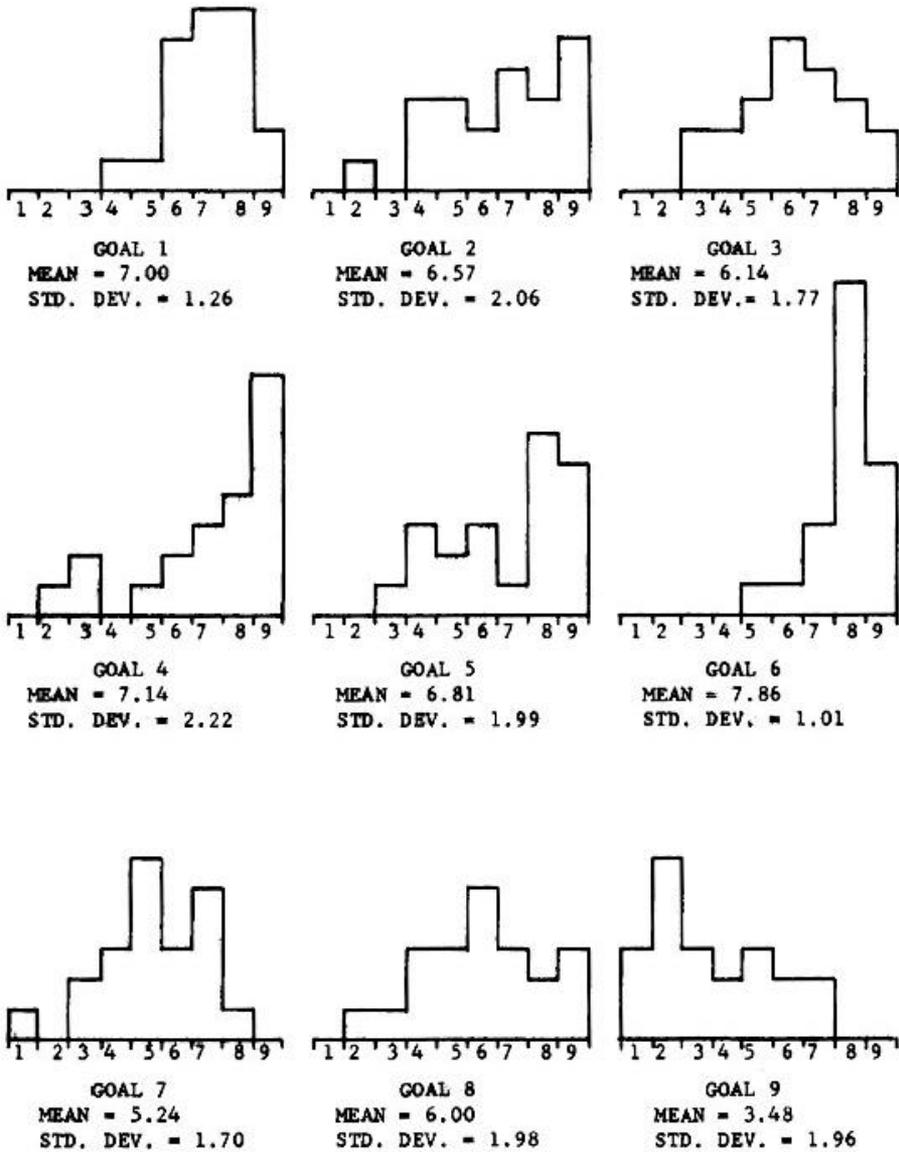


Fig. 2. Results of round one rating analysis.

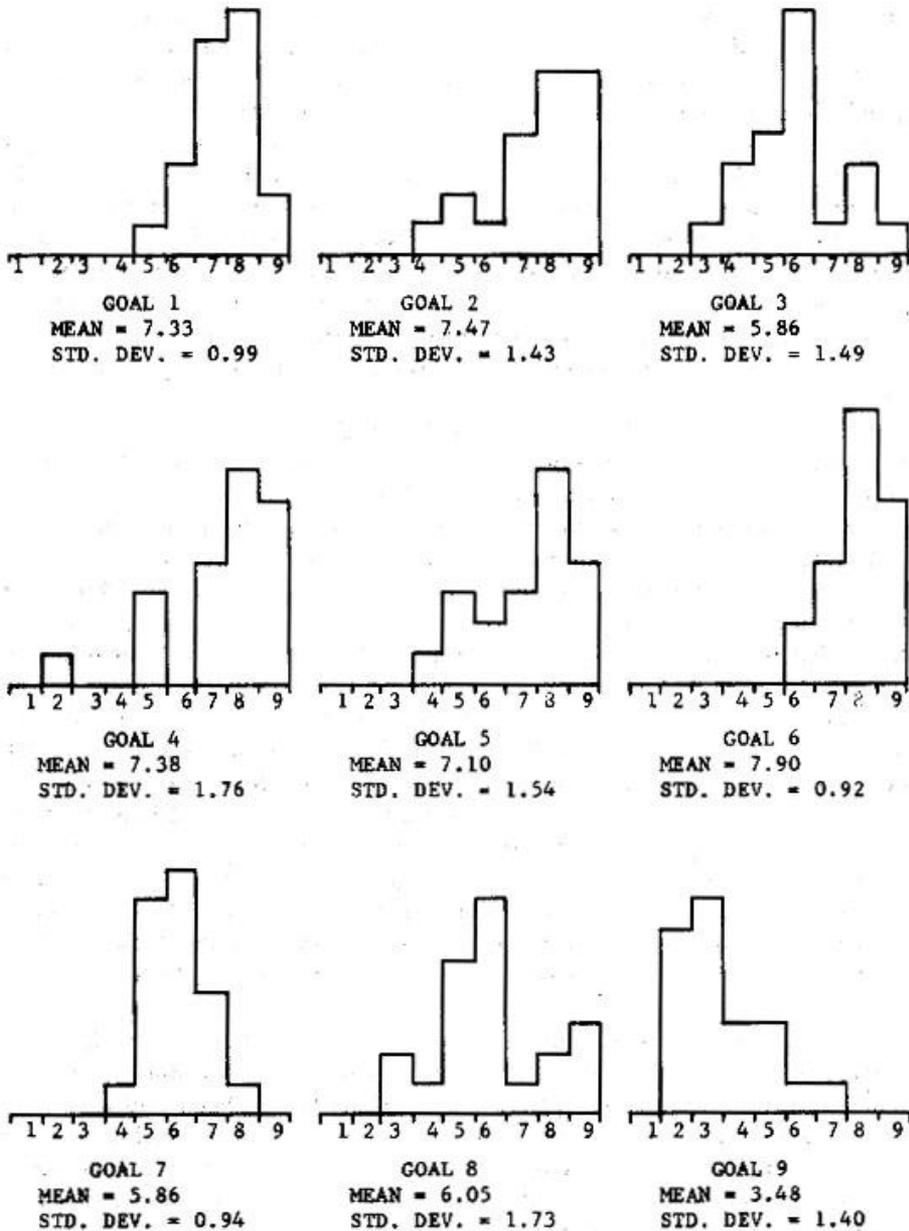


Fig. 3. Results of round four rating analysis.

provide participants with false feedback data, and then to observe the effect of this on the distribution of priority-weight scores.

Two types of feedback were used in this study. The first was a graphical representation of the distribution of scores together with a listing of the mean of the distribution. In addition, in later rounds edited, anonymous comments by the participants concerning the importance of the various goals were distributed. During the experiment on feedback, one goal was chosen and the distribution was altered by the administrators so as to change markedly the position of the mean. Since this was done after the first weighting round, no written feedback accompanied the altered distribution. The goal chosen for this test was Number 3 ("Minimize the operating and construction costs"). This goal was chosen because it appeared to have a good consensus after the first iteration. In addition, it was judged to be substantively important. It was felt that most participants would be very surprised by the altered distribution.

The second-round distribution showed that the feedback had had an effect, since a number of persons shifted their positions away from the true mean. By the third round, the distribution was once again similar to what it had been in the first round, although the distribution was shifted slightly to the lower end of the scale and remained that way permanently, showing residual effects of the gerrymandering. Figure 4 shows the actual distributions and the altered feedback used.

In attempting to explain the reasons behind these changes, the following hypothesis is offered. Upon seeing the first round of feedback information, the respondents had three options: they could ignore the feedback and keep their votes constant; they could rebel against the feedback and move their votes to the right, in the interest of moving the group mean closer to their true desire; or they could acknowledge the feedback and move their votes nearer the false mean. If they had followed either of the first two options, it would indicate that the feedback was not effective in changing individual attitudes. That the third option was in fact taken, however, indicates that the feedback did have an effect on the participants.

The third round, as a result of the feedback of the second round, also shows the effect of feedback. The second-round distribution showed that participants were attempting to increase the priority-for Goal 3, although with respect to their true initial opinions, they were actually decreasing this priority. It seems likely that many respondents, upon seeing this, felt that the group was moving closer to their original position and they decided to return to their first-round vote, since it no longer appeared that this position would be far distant from the mean value of the group.

This experiment suggests that the respondents are, in fact, sensitive to the feedback of distributions of scores from the group as a whole. These results seem to indicate that most respondents are both interested in the opinions of the other members of the group and desirous of moving closer to the perceived consensus.

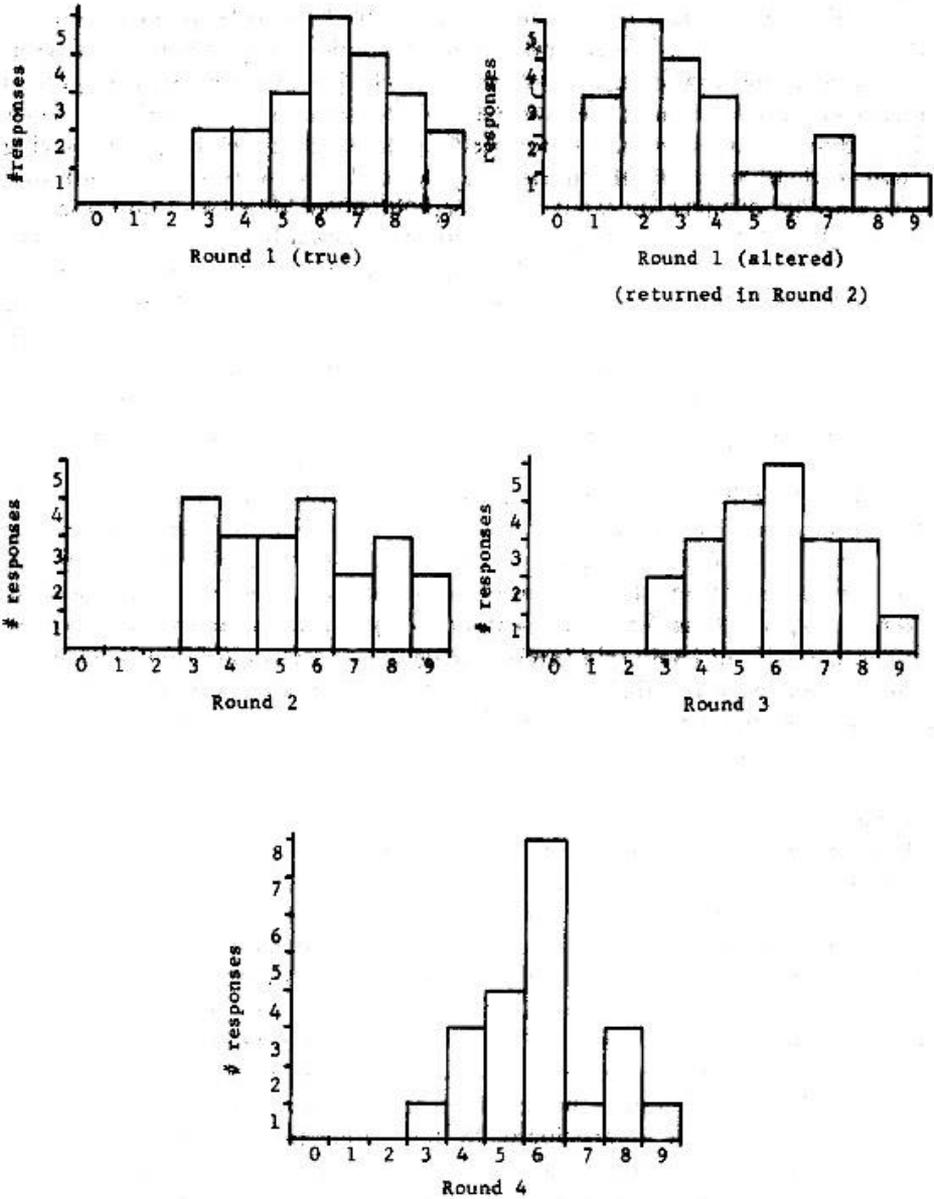


Fig. 4. Distributions generated for Goal 3.

Comparison of Scaling Techniques

With the exception of Dalkey and Rourke [5], there is little discussion in the literature of the different methods of scaling which could be used in a Delphi. The two most common methods which are used are simple ranking and a Likert-type rating scale. Even when these methods are used, there have seldom been attempts to ensure that the scales developed are, in fact, interval scales.

The necessity of having an interval scale is seldom emphasized in Delphis. There is the suspicion that on some occasions the scales derived are ordinal scales. An ordinal scale merely shows the rank order of terms on the scale, and no statement can be made concerning the distance between quantities.

Presumably, the primary reasons for using a Delphi, especially when comparing policies or measuring values, include the determination of not only which policies are considered most important, but also the degree to which each policy is preferred over the other possibilities. In order to assure that this can be determined, an interval scale must be obtained.

In this study, three methods which usually yield interval scales were tested. These methods were simple ranking, a rating-scale method, and pair comparisons. The purpose in trying three scales was to determine if all three methods yielded approximately equivalent interval scales. If this is found to be the case, then in future designs any one of these scales could be used. In this situation, it would probably be wisest to choose that scaling method which was considered easiest to perform by the participants in the Delphi. In this study it was found that the rating-scale method was considered by the participants as the most comfortable to perform. The limitation of the pair-comparison method is that it is time consuming. For example, to apply this method to a set of ten objectives, each participant must make forty-five judgments. The ranking method is fairly easy for a small number of goals, but becomes increasingly difficult as the number of goals increases, for it essentially requires the participant to order the entire list of items in his mind. In addition, many participants felt uncomfortable performing this method because they were prevented from giving two goals an equal ranking (i.e., forced ranking). While this dilemma might possibly have encouraged more thought concerning underlying priorities, it was felt that the frustration caused had a negative effect on the end result. The rating-scale method was found to be quick, easy to comprehend, and psychologically comforting. The participant's task is easy, since he must rate only one item at a time. The problem that remained was to determine whether such a scaling procedure would yield a scale, with interval properties.

In this experiment it was found that each of the three methods yielded somewhat different scales. Using the Law of Comparative judgment [6], scale values for each goal for each round were derived. These values were then translated onto a scale from one to nine. Graphical representations of these scales are given in Fig. 5. Because of the presence of feedback, the four rating rounds are not independent. Each one depends on those previous to it. The scales derived in each successive round should not be identical, for if the scale remains constant from round to round, the justification for using an iterative approach is lost. In addition, because of the order in which the scales were developed, the ranking scale can only be compared with the first-round rating scale and the pair-comparison scale can

only be compared with the fourth round rating scale. Because of four rounds of feedback between them, the ranking scale and pair-comparison scale should be compared only cautiously, and should not be expected to be identical.

The interscale comparison shown in Fig. 5 is not especially encouraging. The pair-comparison method is known to produce interval scales, and the similarity in results of this approach and the round-four, rating results is not strong. The scales produced by ranking and round-one rating are, however, not too different from each other. It is possible to interpret the progression of rating scales from round one to round four as a movement in the direction of the pair-comparison scale. This experiment did not pursue further weighting rounds, but, as discussed below, major changes in weights beyond round four do not seem likely. In addition, later pair-comparison responses might differ from that shown in Fig. 5. Given the complexity of the pair-comparison method for participants, however, it may not be unreasonable to accept cautiously the results of simple rating methods as fair approximations to an interval scale.

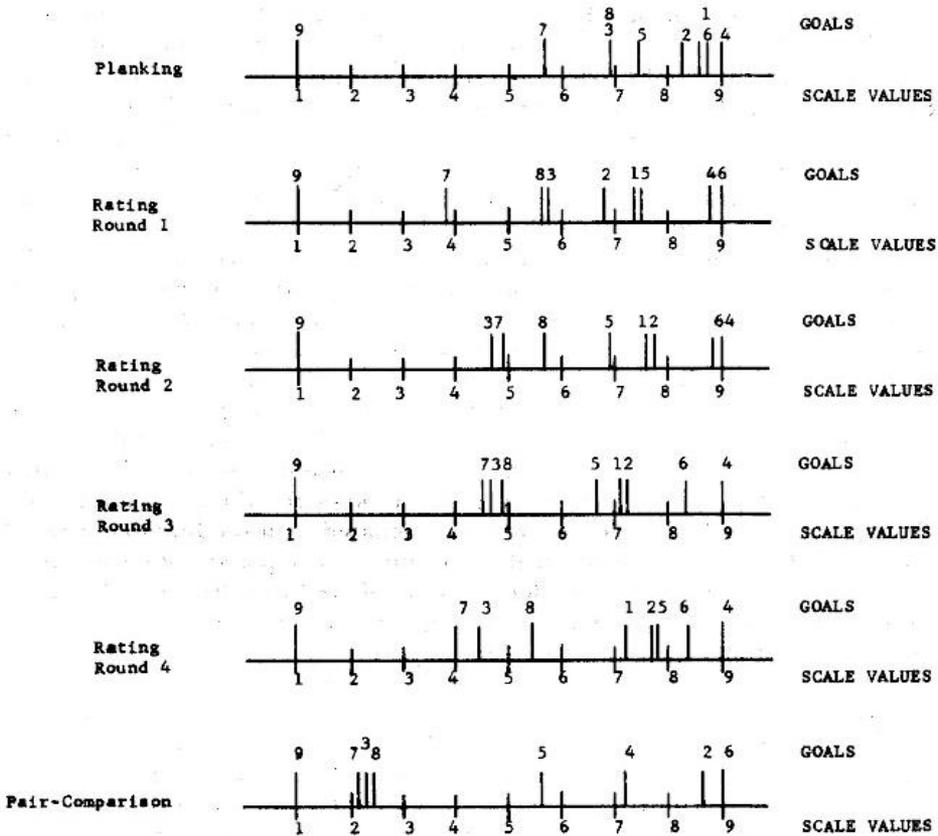


Fig. 5. Comparisons of goal weights on various categorical judgment scales.

The Effect of Personal Variables on Participant Behavior

Dalkey, Brown, and others have considered and used the confidence of participants in their responses to reach more accurate estimates of quantitative phenomena in Delphi exercises. Working with almanac-type data not available to the participants, they found that by selecting for inclusion in the feedback only those responses considered "highly confident" by their proponents, a slightly superior result was achieved [7]. Later, they found that in situations in which relative confidence was measured and in which the "highly confident" group was reasonably large; a definitely superior result could be expected [4].

Studies in the psychology of small groups, however, indicate that highly confident persons should be less influenced by group pressure than those with less confidence, and therefore it would be expected that highly confident individuals move less toward consensus than do others in the Delphi context. Later, Dalkey et al. [4] showed that "over consensus" may occur, and the ratio of average error to standard deviation may actually increase, if consensus is forced too quickly. In order to reach some greater understanding of theory and observation, therefore, several simple hypotheses were tested. It was felt that confidence might involve more than simply confidence in individual answers, and therefore a selection of variables representing various aspects of personal confidence were sought, as well as high/low confidence in each response. Appendix III of this article shows the questionnaire issued to all participants before the Delphi. The variables measured were as follows:

Pre-Delphi Survey Question Number	Independent Variable
1	A. Perceived academic standing relative to other participants
2	B. At-oneness with other participants
3	C. Familiarity with other participants
4	D. Confidence of own capacity in the subject area of the Delphi
5 and 6	E. Experience in the field
7	F. Expectation of the Delphi process (confidence in Delphi)

Each of these confidence variables was then correlated against dependent variables describing the amount of movement actually made by each participant toward the center of the distribution. It was, of course, not possible in this value-judgment Delphi to test accuracy as well, as was done in Dalkey's experiments.

The dependent variables were summed for every participant over each of his nine responses, and were as follows:

- (a) Total amount of change from round one to round four
- (b) Total number of monotonic (as opposed to oscillating) changes made
- (c) Amount of change made in second round
- (d) Number of responses exactly on the mode in round three
- (e) Number of responses within three places of the mode in round three

It was hypothesized that high confidence would be associated with small amounts of total change, monotonic rather than oscillating change, and low confidence with a high degree of change in round two and a high conformity to the consensus in round three.

The response with regard to confidence in individual answers (measured by high/low ratings of confidence on the Delphi answer sheets) was correlated significantly, negatively, but not very highly with the amount of change in the second round, although not with the total change between rounds one and four.

Percent of "highly confident" answers, however, was cross-correlated positively with perceived academic status, although this was not significantly connected with either movement variable. Amount of change in the second round was also just correlated (positively) with the "at-oneness" variable, although there was no relationship at all between "at-oneness" and percent highly confident. These represented the only significant correlations found.

The evidence for the effect of confidence on the tendency to converge is somewhat sketchy. The only conclusions that can be drawn from the experience is that the initial surprise on being confronted with some distribution of group opinion may to some extent cause the less confident members who believe that they associate with the rest of the group to move toward the center of opinion, but that this tendency is certainly not an overwhelming one.

At the end of the Delphi a second questionnaire (Appendix IV of this article) was used to determine whether the kind of feedback provided had any conscious effect on movement in the Delphi. The variables were as follows:

Post-Delphi Survey Question Number	Independent Variable
1	G. Satisfaction with the results
2, 3, and 4	H. Success of feedback in the learning process
5 and 6	I. Frustration with communication levels
7	J. Optimism for Delphi
8	K. Feeling of being rushed

These variables were correlated with the same dependent variables. Both optimism for the future- of Delphi and satisfaction with the, process correlated significantly and positively with the number of monotonic changes made, perhaps indicating that those people who were not caused to change their opinion radically were in better spirits after the Delphi than the others. However, the success-of-feedback variable was strongly and negatively correlated with *the* propensity to conform to the mode, in round three: In other words, those. who did conform to the visible majority had difficulty in giving and taking, ideas from the feedback. This is interesting in that it indicates the different kinds of feedback that may affect people; in different ways. The tendency to converge strongly has elsewhere been shown by Schofer and Skutsch [8] to be due to emphasis in the visible consensus and on the need to create consensus. Satisfaction with the process was also negatively correlated with the conformity variable: (Satisfaction and agreement with feedback were strongly cross-correlated.) Clearly, the people who were strongly conforming were not happy with the Delphi at all. The question of what is cause and what is effect, however, remains to be answered. Yet one might speculate that, especially in a value-oriented Delphi, the group pressure from some forms of feedback can be overly strong, forcing participants to take positions which they find uncomfortable. While compromise may be uncomfortable in any situation, the real danger here is that participants may leave the process without really compromising their feelings at all. That is, perhaps the anonymity of the Delphi itself may have encouraged participants to capitulate, but only on paper. They may later hold to their original views, and, if the results of the Goals Delphi are used to develop programs to meet their needs, participants might ultimately be quite dissatisfied with the results.

A cautionary note is relevant at this point. Another study by Skutsch [9] has shown that the form of the feedback itself influences consensus development. Despite the fact that participants in this experiment were encouraged to report their verbal rationale for their positions, the rapidity with which the process was carried out tended to discourage such responses. As a result, histograms of value weights formed the bulk of the feedback. It is just this kind of limited, "hard" feedback which tends to force what might be an irrational consensus, one which might be only temporary.

Opinion Stability as a Method of Consensus Measurement

In most Delphis, consensus is assumed to have been achieved when a certain percentage of the votes fall within a prescribed range-for example, when the interquartile range is no larger than two units on a ten-unit scale. Measures of this sort do not take full advantage of the information available in the distributions. For example, a bimodal distribution may occur which will not be registered as a consensus, but indicates an important and apparently insoluble cleft of opinion. Less dramatically, the distribution may flatten out and not reach any strongly peaked shape at all. The results of the Delphi are no less important for this, however. Indeed, considering that there is a strong natural tendency in the Delphi for opinion to centralize, resistance in the form of unconsensual distributions should be viewed with special interest.

A measure which takes into account such variations from the norm is one that measures not consensus as such, but stability of the respondents' vote distribution curve over successive rounds of the Delphi. Because the interest lies in the opinion of the group rather than in that of individuals, this method is preferable to one that would measure the amount of change in each individual's vote between rounds.

To compare the distributions of opinion between rounds, the histograms may be subtracted columnwise and the absolute value of the result taken. In Table 2 this approach is applied to the weight histograms reported for Goal 5 in the first three rounds of the Delphi. Columnwise subtraction between the first and second, and the second and third, rounds gives the results shown in Table 2. The absolute values of the differences between histograms are aggregated to form total units of change; but since any one participant's change of opinion is reflected in the histogram differences by two units of change, net personchanges must be computed by dividing total units of change by two. Finally, the percentage change is determined by dividing net changes by the number of participants. Clearly, in the example shown in Table 2, the distribution of value weights for Goal 5 became more stable¹ between rounds one and three.

The question 'of what represents a reasonable cut-off point at which the response may be said to be unchanged, and therefore finally in its stable position,' poses some problems however. Since there is no underlying statistical theory in what has so far been proposed, no true statistical level may be set, as might, for example, be possible with a statistical change in variance test.¹

Empirical examination of the responses in the Delphi, however, showed that at any point in time a certain amount of oscillatory movement and change within the group is inevitable. This might be conceptualized as a sort of underlying error function, a type of internal system noise. What is needed is a "confidence" measure which allows the distinction to be drawn between this kind of movement and strong group movements that represent real changing opinion. Such an estimate has tentatively been made from studies of observed probability of movement.

Leaving aside objective 3 (for reasons made apparent earlier), the propensity of the individual to alter his score as a function of distance from the center point was measured. This was done by calculating the proportion of respondents at each scale distance from the mode that moved toward the mode between rounds. The results, displayed in Fig. 6, show a strong tendency for increased amounts of movement with distance from the center point. They also show that a percentage change is to be expected among respondents who are already dead on the mode itself. The amount of movement at the mode (about 13%) has therefore been taken to represent the base of oscillatory movement to be expected, and this is supported by the fact that the amount of change at the centroid does not alter appreciably between rounds.

¹ Conventional variance tests were found to be unsuited to the case of change in histogram shape in this context. Most rely on independent samples; none is strong enough to pick up small changes in shape, and none robust enough to deal with non-normal distributions.

**Table 2: Example of Stability Measurement Computations
for Goal 5**

<u>Rating</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>
Absolute difference in number selecting rating, rounds 1-2 (a)	0	0	0	2	2	2	4	0	2
Total units of change (b)	12								
Net person-changes (c)	6								
Number of participants	21								
Percent change (d)	28.6%								
<u>Rating</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>
Absolute difference in number selecting rating, rounds 2-3 (a)	0	0	1	0	1	1	0	0	1
Total units of change (b)	4								
Net person-changes (c)	2								
Number of participants	21								
Percent change (d)	9.5%								

- (a) These numbers are the absolute differences in the histograms for the two successive rounds.
- (b) These numbers are the sums of the absolute differences in the histograms.
- (c) Net changes are total units of change divided by 2.
- (d) Percent change is net change divided by the number of participants.

Using the 15% change level to represent a 'state of equilibrium, any two distributions that show marginal changes of less than 15% may be said to have reached 'stability; any successive distributions with more than 15% change should be included in later rounds of the Delphi, since they have not come to the equilibrium position.

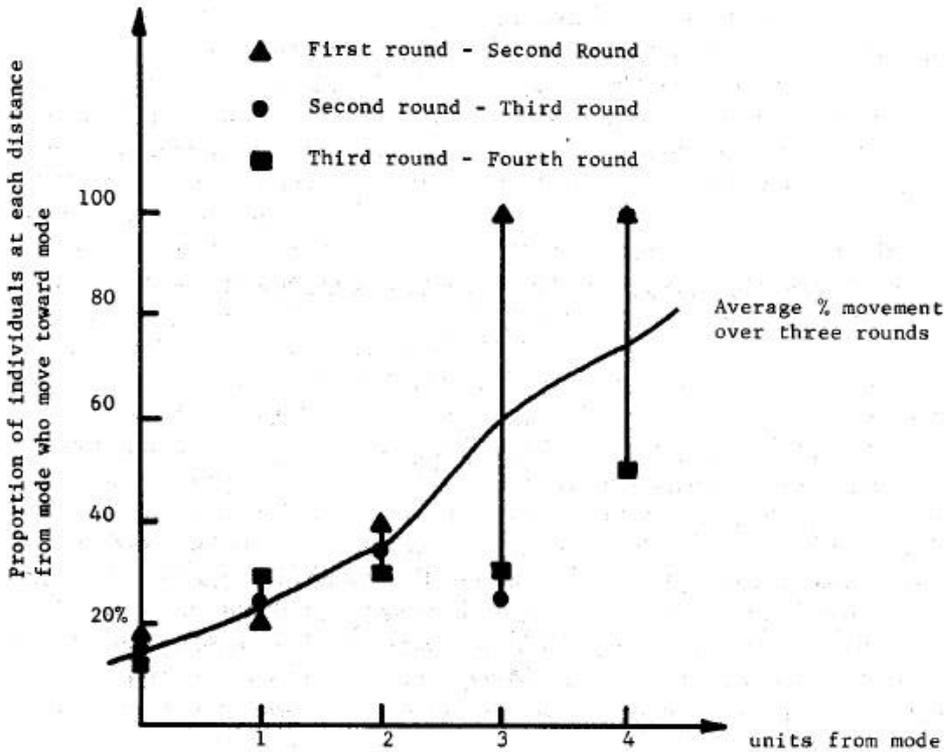


Fig. 6. Proportion of respondents shifting their positions, as a function of distance from the mode.

The results for all nine goals included in this experimental Delphi using this analysis are shown in Table 3. From these data, there can be no doubt as to the general tendency toward stabilization. Only one goal, 7, had not reached a stable position by the end of the third round, although 3, 8, and 9 were all only just stable.

In general, this method seems to have a number of advantages. Firstly, it allows the use of more of the information contained in the distributions. There are applications in which, at the end of the Delphi process, the entire distribution may be used, as for example in linear-weighting evaluation models where goal-weight distributions are treated stochastically, such as that by Goodman [10]. In addition, this stability measure is relatively simple to calculate, and has much greater power and validity than parametric tests of variance.

Table 3: Results of Stability Analysis

<u>Goal</u>	<u>Amount change Rounds 1-2</u>	<u>Amount change Rounds 2-3</u>
1	13% stable	10%
2	17%	10% stable
3	38%	14% just stable
4	10% stable	10%
5	26%	10% stable
6	2% stable	2%
7	24%	33% increasingly unstable
8	22%	14% just stable
9	24%	14% just stable

Perhaps most important, one of the original objectives of Delphi was the identification of areas of difference as well as areas of agreement within the participating group. Use of this stability measure to develop a stopping criterion preserves any well-defined disagreements which may exist. To the organizer of a Goals Delphi, this information can be especially useful.

Delphi Service Program

In order to make several iterations possible in the space of a very short time period, a computer time-sharing terminal was used to process the results of this Delphi experiment. Unlike the systems described by Price (see Chapter VII, B), the program used in this Delphi was an accounting device only; verbal feedback was compiled and read to participants by the organizers.

In this application, histograms produced by the computer terminal were copied by hand onto an overhead projector transparency to provide immediate feedback to participants, who themselves determined their positions in the distributions relative to the group. It is anticipated that, for future experiments, computer-generated histograms will be produced in multiple copies, one of which will be provided to each participant.

This type of computer support, oriented toward the use of a single terminal for all participants, may be especially desirable for Goals Delphi applications, where, because of the lay nature of the respondents, it seems especially desirable to keep all of those involved in a single room, and to maintain a relatively high rate of progress throughout-the survey.

Closure

The potential applicability of the Delphi method to goal formulation and priority determination in public systems is very great. Yet, because the detailed characteristics of the design of the process can have important effects on the nature of the outcomes, it will be important to tailor the Goals Delphi to the problems at hand. The structuring of internal characteristics which are appropriate to a Goals Delphi should be based on a rather complete understanding of the linkages between form and function in the Delphi environment. While considerable experience must be gained before Delphi can be offered as a routine goal-formulation process, this discussion has suggested some structural and process features relevant to this important application of the Delphi method.

Appendix I: Hypothetical Decision Scenario

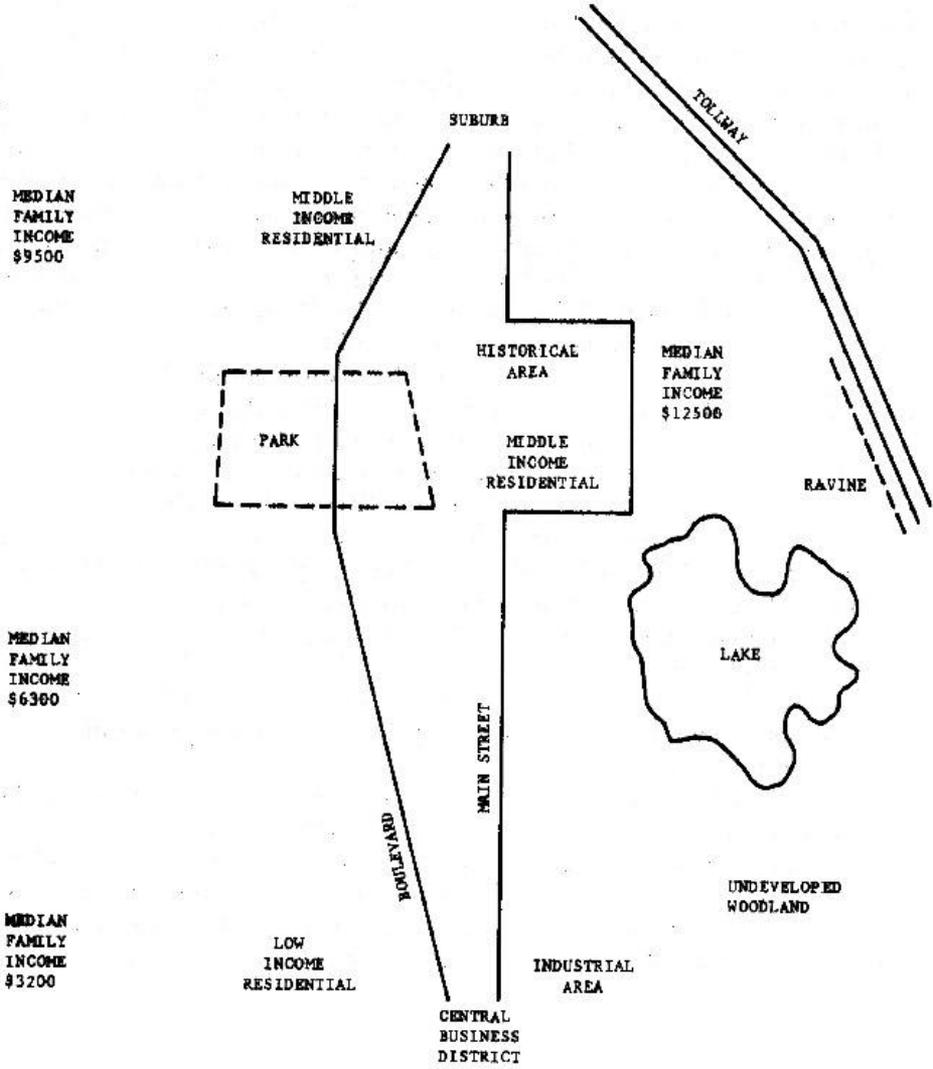
The following transportation-facility-location problem is offered as an appropriate context for developing local-scale transportation planning objectives. Within this context, there is a need to establish an objective set, and to evaluate *quantitatively* several alternative plans in the context of the objectives.

A two-mile transportation link is proposed in an urban area. It is to run from the Central Business District (CBD) to new, developing suburbs to the north. This area is presently served by a four-lane boulevard with an average daily traffic (ADT) of forty thousand vehicles, and by a four-lane street with an ADT of twenty thousand vehicles. This street, however, circles an historical area by means of four 90° turns, and traffic must travel this section at twenty-five mph. The southeast corner of the historical area comes within five hundred yards of the edge of a lake, and the main street presently is only one block from the lake at this point. A tollway also passes the suburb and proceeds in a southeasterly direction. The tollway passes within one mile of the CBD, with its alignment located in a ravine. The elevation of the ravine is such that to build a connector to the tollway from the CBD would require a great deal of earthwork, and even with this the grade would be about 3%.

The alignment of the boulevard is such that it begins in the CBD, proceeding northwesterly through a low-income area to a large park, where it turns and continues in a northeasterly direction through a middle-income residential area to the suburbs.

The four-lane street heads due north from the CBD, passes through an industrial park, and then makes four sharp turns around the historical area and proceeds directly into the suburbs.

Citizen opposition can be expected in four areas. Public opinion has long been against any changes in the historical area. A citizen group can be expected to form opposing an alignment through the park. One can also be expected to form opposing removal of houses in the middle-income areas north and east of the park. Problems can also be expected if the alignment goes through the low-income area, requiring relocation of some households.



Appendix III: Pre-Delphi Self-Rating Form

Code Number _____

PRE-DELPHI SURVEY

- | 1. As a transportation planner, in a class of transportation planners, my skills in planning would put me about here, relative to the others. | <table border="0"> <thead> <tr> <th colspan="5">Very Highly Skilled</th> <th colspan="2">No Skill at all</th> </tr> </thead> <tbody> <tr> <td style="border-top: 1px solid black;">1</td> <td style="border-top: 1px solid black;">2</td> <td style="border-top: 1px solid black;">3</td> <td style="border-top: 1px solid black;">4</td> <td style="border-top: 1px solid black;">5</td> <td style="border-top: 1px solid black;">6</td> <td style="border-top: 1px solid black;">7</td> </tr> </tbody> </table> | Very Highly Skilled | | | | | No Skill at all | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|-------------------------|---|---|---------------------------------------|---|---------------------------------------|--|---|---|---|---|---|---|---|
| Very Highly Skilled | | | | | No Skill at all | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | | | | | | |
| 2. I think my ideas are, in essence, in agreement with the rest of the class. | <table border="0"> <thead> <tr> <th colspan="5">Yes, absolutely</th> <th colspan="2">No, not at all</th> </tr> </thead> <tbody> <tr> <td style="border-top: 1px solid black;">1</td> <td style="border-top: 1px solid black;">2</td> <td style="border-top: 1px solid black;">3</td> <td style="border-top: 1px solid black;">4</td> <td style="border-top: 1px solid black;">5</td> <td style="border-top: 1px solid black;">6</td> <td style="border-top: 1px solid black;">7</td> </tr> </tbody> </table> | Yes, absolutely | | | | | No, not at all | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Yes, absolutely | | | | | No, not at all | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | | | | | | |
| 3. I know most of the people in the class very well. | <table border="0"> <thead> <tr> <th colspan="5">Yes, pretty much</th> <th colspan="2">No, none at all</th> </tr> </thead> <tbody> <tr> <td style="border-top: 1px solid black;">1</td> <td style="border-top: 1px solid black;">2</td> <td style="border-top: 1px solid black;">3</td> <td style="border-top: 1px solid black;">4</td> <td style="border-top: 1px solid black;">5</td> <td style="border-top: 1px solid black;">6</td> <td style="border-top: 1px solid black;">7</td> </tr> </tbody> </table> | Yes, pretty much | | | | | No, none at all | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Yes, pretty much | | | | | No, none at all | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | | | | | | |
| 4. I have some definite ideas about what the goals in transportation planning are and should be. | <table border="0"> <thead> <tr> <th colspan="5">Yes, lots</th> <th colspan="2">No, none</th> </tr> </thead> <tbody> <tr> <td style="border-top: 1px solid black;">1</td> <td style="border-top: 1px solid black;">2</td> <td style="border-top: 1px solid black;">3</td> <td style="border-top: 1px solid black;">4</td> <td style="border-top: 1px solid black;">5</td> <td style="border-top: 1px solid black;">6</td> <td style="border-top: 1px solid black;">7</td> </tr> </tbody> </table> | Yes, lots | | | | | No, none | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Yes, lots | | | | | No, none | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | | | | | | |
| 5. I have been in transportation for longer than most of the other people here. | <table border="0"> <thead> <tr> <th colspan="5">Yes</th> <th colspan="2">No</th> </tr> </thead> <tbody> <tr> <td style="border-top: 1px solid black;">1</td> <td style="border-top: 1px solid black;">2</td> <td style="border-top: 1px solid black;">3</td> <td style="border-top: 1px solid black;">4</td> <td style="border-top: 1px solid black;">5</td> <td style="border-top: 1px solid black;">6</td> <td style="border-top: 1px solid black;">7</td> </tr> </tbody> </table> | Yes | | | | | No | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Yes | | | | | No | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | | | | | | |
| 6. I have a lot of experience in planning outside of school. | <table border="0"> <thead> <tr> <th colspan="5">Yes</th> <th colspan="2">No</th> </tr> </thead> <tbody> <tr> <td style="border-top: 1px solid black;">1</td> <td style="border-top: 1px solid black;">2</td> <td style="border-top: 1px solid black;">3</td> <td style="border-top: 1px solid black;">4</td> <td style="border-top: 1px solid black;">5</td> <td style="border-top: 1px solid black;">6</td> <td style="border-top: 1px solid black;">7</td> </tr> </tbody> </table> | Yes | | | | | No | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Yes | | | | | No | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | | | | | | |
| 7. I am anticipating that the Delphi is going to be a good thing for goal setting. | <table border="0"> <thead> <tr> <th colspan="5">Yes, I think it will be</th> <th colspan="2">No, I think it may be a waste of time</th> </tr> </thead> <tbody> <tr> <td style="border-top: 1px solid black;">1</td> <td style="border-top: 1px solid black;">2</td> <td style="border-top: 1px solid black;">3</td> <td style="border-top: 1px solid black;">4</td> <td style="border-top: 1px solid black;">5</td> <td style="border-top: 1px solid black;">6</td> <td style="border-top: 1px solid black;">7</td> </tr> </tbody> </table> | Yes, I think it will be | | | | | No, I think it may be a waste of time | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Yes, I think it will be | | | | | No, I think it may be a waste of time | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | | | | | | | | | |

Appendix IV: Post-Delphi Evaluation Form

Code Number _____

POST-DELPHI SURVEY

- | | | |
|--|------------------------|---|
| 1. I feel satisfied with the results in general. | _____
1 2 3 4 5 6 7 | I'm not really happy with the results at all. |
| 2. I learned ideas from the feedback. | _____
1 2 3 4 5 6 7 | I didn't learn a thing from the feedback. |
| 3. In general, I agreed with the ideas in the feedback. | _____
1 2 3 4 5 6 7 | I disagreed with everything in the feedback. |
| 4. I could express my ideas OK this way | _____
1 2 3 4 5 6 7 | I couldn't really write what I wanted to say. |
| 5. I feel as if I really wanted to talk to people. | _____
1 2 3 4 5 6 7 | I didn't feel the need to talk at all. |
| 6. I have a feeling people didn't understand or think about my reasons. | _____
1 2 3 4 5 6 7 | I think people understood my reasons pretty well. |
| 7. I think the Delphi could be operational in goal setting more generally. | _____
1 2 3 4 5 6 7 | I don't think it could be operational, really. |
| 8. I think it went too fast. | _____
1 2 3 4 5 6 7 | I think it went too slowly. |

References

1. T. H. Ostrom and H. S. Upshaw, "Psychological Perspective and Attitude Change," in A. G. Greenwald *et al.*, *Psychological Foundations of Attitudes*, Academic Press, New York, 1968.
2. M. Wachs and J. L. Schofer, "Abstract Values and Concrete Highways," *Traffic Quarterly*, 23, No. 1 (1969).
3. G. S. Rutherford, J. L. Schofer, M. Wachs, and M. Skutsch, "Goal Formulation for Socio-Technical Systems," *Journal of the Urban Planning and Development Division, Proceedings of the American Society of Civil Engineers*, forthcoming.
4. N. C. Dalkey, B. Brown, and S. Cochran, "The Delphi Method IV.:Effect of Percentile Feedback and Feed-in of Relevant Facts," Rand Corporation, RM-6118-PR, 1970.
5. N. S. Dalkey and D. L. Rourke, "Experimental Assessment of Delphi Procedures with Group Value Judgment," Rand Corporation, R612-ARPA, 1971.
6. W. S. Torgerson, *Theory and Methods of Scaling*, Wiley, New York, 1958.
7. N. C. Dalkey, "The Delphi Method: An Experimental Study of Group Opinion," Rand Corporation, RM-5888-PR, 1969.
8. M. Skutsch and J. L. Schofer, "Goals Delphis for Urban Planning: Concepts in their Designs," *Socio-Economic Planning Sciences*, 7, pp. 305-313, 1973.
9. M. Skutsch, "Goals and Goal Setting: A Delphi Approach," unpublished master's thesis, Department of Industrial Engineering and Management Science, Northwestern University, 1972.
10. K. M. Goodman, "A Computerized Investigation and Sensitivity Analysis of a, Linear Weighting Scheme for the Evaluation of Alternatvies for Complex Urban Systems;" unpublished master's thesis, Department of Civil Engineering, Northwestern University, 1971.