Toward Quality Criteria for Collective Judgments

Stuart L. Hart

Institute for Social Research, University of Michigan

While most organizational and social decision making is done in a group or collective mode, there are few guides or evaluative criteria for judging when a high-quality outcome has been reached. Most past studies of group decision making have been conducted in laboratories using student subjects and factual problems with correct answers as means for judging outcome quality. Such proxies are rough approximations at best of real-world conditions where value differences can be intense and problems have no correct or best answer. Drawing upon the existing literature, evaluative criteria are proposed including process, content, and outcome concerns. An instrument based upon these criteria is then applied retrospectively to six cases of ad hoc collective decision making. Using dimensional analyses, a set of more detailed evaluative factors is derived from actual participant responses. The results provide some insights into the nature of high-quality collective judgments as well as the most effective procedures for their achievement. © 1985 Academic Press, Inc.

Few important organizational decisions are made by individuals. Complexity, social turbulence, and uncertainty have forced most administrative and strategic decision making into a group or "collective" mode. By including a broader range of expert and stakeholder perspectives in decision efforts, the likelihood that critical dimensions of complex problems will be recognized and incorporated into decisions is presumably increased. Wider participation also seems to enhance commitment to decisions among involved parties.

Direct participation has become an American entitlement as major social and environmental legislation in the 1960s and 1970s has provided citizen and stakeholder groups with access to decision forums through formal administrative and legal procedures. As a result, decision making has become less an analytic endeavor and more an exercise in mediating among parties with differing images and values concerning particular problems or issues (i.e., less decision and more judgment). Even when there is an underlying desire, on the part of participants, to reach a mu-

tually acceptable outcome, it is often difficult to know when a high-quality decision or viable consensus has been reached.

While the measurement of group problem-solving effectiveness is not a new methodological concern, most past efforts suffer from one of two major design flaws.[1] First, the vast majority of studies have been con-ducted in laboratories using student subjects as proxies for real group members or participants. The contrived nature of such groups necessarily limits the generalizability of the results, since value conflicts and emo-tional commitment are seldom captured in such experiments (Hoffman, 1979). Second, many studies have used factual problems with correct answers as the means for judging problem-solving effectiveness. This proxy, in effect, substitutes objective *accuracy* for assessments of sub-jective quality in determining group success (e.g., Einhorn, Hogarth, & Klempner, 1977; Hall, 1971). Such a design bears little resemblance to the complex, judgmental problems often faced by real-life decision-making groups, making generalization of the results even more difficult.

Where experiments have utilized complex, judgmental problems (i.e., problems without correct answers), panels of judges have often been employed to assess the quality of decisions reached (e.g., Nutt, 1979). Few such experiments, however, have compared the expert evaluations with the subjective assessments of group participants for convergence. Experimental studies using participant assessments as measures have used a variety of criteria to assess effectiveness; these have included participant satisfaction with, support for, and commitment to the collec-tive outcome. There has been little effort, however, to construct system-atically a set of empirically derived criteria against which to judge out-come quality.

This paper is intended as a step in the process of developing more meaningful measures in this area. It also provides some preliminary thoughts about the nature of high-quality collective judgments (defined as decisions requiring the integration of diverse expert and/or stakeholder perspectives) and the most effective procedures for their achievement. The paper addresses the following questions:

- What procedures or activities are most conducive to the development of mu-tually acceptable collective judgments in real-world settings?
- How does one know when a superior (or even good) collective outcome has been reached?
- What are the criteria for judging the quality of collective decision outcomes?

The article begins by reviewing the literature in the areas of group decision making and consensus. From this review, a set of three inter-

---

[1] The work of Kenneth Hammond and his associates stands as an exception to these general statements (e.g., Balke *et al.*, 1973; Hammond & Adelman, 1976).

related criteria (process, content, and outcome) were used to derive five theoretical constructs for evaluating the quality of collective judgments. These constructs served as the basis for designing a questionnaire used to evaluate the effectiveness of six cases of actual collaborative problem solving. Dimensional analysis was then employed to discover a more refined set of evaluative factors based upon actual respondent ratings to questionnaire items. The article closes with a discussion of the nature of high-quality collective judgments and the most effective procedures for their achievement.

TOWARD QUALITY CRITERIA

What Procedures Nurture Consensus?

Collective judgments involve many perspectives and differing assumptions concerning an issue or problem; stakeholders generally do not share the same initial problem definition. It is therefore inevitable that differences of opinion will arise with regard to appropriate solutions. The traditional orientation toward differences arising within groups has been to attempt to reduce tension through compromise. But as Follett (1924) pointed out

> compromise is suppression . . . so we see again and again that what has been suppressed in the compromises of politics and labor disputes crops up anew to bring more disastrous results. (p. 164)

Indeed, initial disagreement should be regarded as a strength, since it provides groups with different options to work from. The need, therefore, is to find ways to manage conflict constructively without eliminating the differences which promote diversity and creativity. Hall and Watson (1970) found that with a minimum of process training, groups can greatly enhance their performance. They found that trained groups tended to avoid premature convergence and actively promoted divergence during the early stages of deliberation. Opinion differences were viewed as natural resultants of multiperson situations. They concluded that group performance is not only a function of the attitudes of group members but also of their shared perceptions of what constitutes appropriate member behavior.

Another approach is for people with different perspectives to interact in a manner which fosters the envisioning and achievement of "superordinate goals." Eiseman (1978) concisely described the nature of such "integrative solutions":

> When parties can transcend what had appeared to be an irreconcilable conflict. when they perceive that their initial positions have been enhanced by their comprehending in a new way positions they had previously rejected . . . then synergy becomes more than a possibility: it becomes a probability. (p. 137)

Likert (1976) suggested a similar consensus approach to group decision making, stressing the importance of a "win–win" approach to problem solving rather than the traditional "win–lose" paradigm:

> Consensus, in Quaker terms, is 'the sense of the meeting', a willing acceptance of the groups' conclusions. . . . [It] is a cooperative effort to find a sound solution acceptable to everyone rather than a competitive struggle in which an unacceptable solution is forced on the losers. (p. 146)

More recently, Likert and Likert (1978) suggested that when an additional step is added to the problem-solving process, the possibility for a win–win solution is greatly increased. This consists of the listing of conditions that the participants feel a solution must meet for it to be acceptable to them. This step decreases the probability that members will remain emotionally committed to particular solutions. It is similar to the advice given by Fisher and Ury (1981) and Mitroff and Emshoff (1979). The former suggest focusing on problems rather than positions while the latter advocate working backward to the underlying assumptions behind initial positions. Both tactics enhance the likelihood of getting participants to work on a group task rather than arguing personal positions.

Image Expansion and Problem Redefinition

Traditionally, consensus decision has meant that (a) each individual understands the others' points of view and (b) whether or not they prefer the decision, each will support it, because it was arrived at in a fair and open manner. In the context of collective decision making, however, such a definition of consensus seems to unnecessarily constrict the ability of individuals to both *influence* and *be influenced by* the viewpoints or analysis of others. Problem-solving interactions thus may change what has to be thought about and what has to be analyzed. Indeed, it is only through interaction on complex issues that participants confront the conflicts and contradictions implicit in their own initial positions.

Considerable evidence on the effects of discussion on judgments has been generated by Persuasive Arguments Theory (Vinokur & Burnstein, 1974). This theory, like the Social Comparison Theory of Festinger (1954) predicts no change of opinions or attitudes unless there is knowledge of the judgments of others and opportunity to rethink one's initial position. Social Judgment Theory (e.g., Axelrod, 1976; Stewart & Gelberd, 1976) also indicates that information about differences in "judgment policies" of participants should reduce disagreement and improve the quality of group judgment. Thus, research indicates that a critical feature in achieving consensus about an issue is the cognitive remapping of the situation by the individuals involved (Dalton, 1969). Similarly, Kaplan

and Kaplan (1982) observed that if one is not successful at generating possible paths between the problem state and the preferred state, then

> . . . alternative ways of viewing the problem must be sought. Anything that changes how one defines some aspect of the problem has a chance of ending the impasse. (pp. 156–157)

From this perspective, differing initial images of a given situation may be described as *subjective:* changes occur within participants themselves (and their orientations to the issue) rather than in the ''objective'' situation external to them from which the original images arose (Mitchell. 1973). The goal in this regard is to achieve (through interaction), a common perspective on the issue or problem; this requires a mutual ''re-perception'' of the situation by participants (Hill, 1982). The desired approach could thus be described as a ''search into the minds of the conflicting parties, probing for those cognitive and emotional elements that might constitute a mutually satisfactory solution'' (Levi & Benjamin. 1977, p. 421).

Given its orientation toward redefinition and image expansion. the above approach is quite different from mathematical models of decision making which assume access to all necessary information at the outset (e.g., Hobbs & Voelker, 1978; McKenna, 1980). It is also different from theoretical voting procedures for the aggregation of individual judgments which also assume fixed preferences (e.g., Arrow. 1951). Indeed, it is precisely the assumption of fixed and known preferences that artificially constrains the applicability of most quantitative decision-making and voting approaches. In order for the preferences of individuals to be summed together in an aggregated ''welfare function,'' it must be assumed that individual ''utilities'' are not influenced by the preferences of others; thus, all forms of interaction are discounted or assumed away, limiting the feasible solutions to the array of alternatives which exist a priori. It is precisely the interaction of diverse and cooperating individuals, however, which offers the possibility for learning and human growth, thereby extending the feasible ''solution space'' to frontiers not envisioned at the beginning of the process.

Consensus and Optimality

Given the above discussion, consensus becomes less a compromise and more a shared image of both the problem being addressed and the most appropriate configuration of responses to the situation, arrived at through a process of discourse deemed reasonable by all involved. Since the size of the solution ''pie'' is not necessarily fixed, it is possible to first ''enlarge'' the pie before ''slicing'' it. This suggests that it is feasible to think about a decision process which yields consensus results which

more closely approximate an interactively derived optimum in the Pareto sense—that is, a solution configuration that makes everyone "better off" (see Pareto, 1916). Realizing, however, that such ideal solutions are seldom possible, it is more realistic to strive for the criterion of "Pareto Improvement"—the combination of policies that makes as many people better off as is possible, without making anyone worse off. Facilitators of such a process search for that configuration of alternatives which increases the zones of support and acceptance, while simultaneously decreasing the zones of indifference and conflict (see Summer, 1980).

Drucker (1980) recently called for an organizational decision strategy which *optimizes* the performance of the primary task of the organization while, at the same time, *satisficing* constituency groups with the minimum needed to placate or appease. It seems feasible to go beyond this strategy by seeking something closer to Zelany's (1982) concept of *conflict dissolution*. In traditional problem resolution, each party gives up something which reduces the conflict or diffuses it temporarily. But to dissolve a problem, one must find or invent a "prominent alternative" which represents a point where all multiple objectives are optimized. If this ideal cannot be achieved, the objective becomes reduction of the distance between the prominent alternative and the compromise solution. The "optimization" process implicit in the above approach is quite different from that associated with statistical decision theory. Rather than aggregating fixed utilities, individual and group preferences are sequentially discovered, influenced, and reformulated by a process of exploration and evaluation. This results in a more complete image of revealed preferences rather than a collection of espoused strategic positions.

The pattern of the "optimum" response can thus be defined operationally as the point at which participants can no longer supplement, rearrange, or otherwise alter the configuration in a way deemed better by the collective. Press (1978) and Radford (1977) both suggested a similar criterion in determining when an acceptable interactive outcome has been reached for complex, judgmental problems. The "stopping rule" suggested by the former was defined by *stability* of the assessments of all participants (i.e., the process converges when none of the participants change their assessments from one round to the next). The latter defined the criterion of "P-Stability"; an outcome is said to be P-stable if, regardless of which participants act in which sequence in each successive round of consideration, the same policy is adopted as the output of the process. In operation, it would seem that this point is somewhere in between consensus and unanimity since all are in concurrence but none are forced to acquiesce.

Criteria for Judging Outcome Quality

The various concepts regarding decision quality discussed above seem to consist of three interrelated criteria: process, content, and outcome. In order for a proposed recommendation or decision to be considered an improvement over what previously existed, all three components must be successfully blended. Identifying the *process* by which the decision was reached as fair, open, and collaborative (as opposed to manipulative and coercive) is necessary but not sufficient for a high-quality decision. While the idea of "procedural rationality" as the ultimate criterion for decision is alluring (see Nagel, 1980), it is also vulnerable to criticism: though the process may be deemed just, its product may still be both unjust and unrelated to the knowledge base (Rein, 1976).

It is for this reason that the interactive evaluation of the *content* of the participants' points of view is so important. The first step here is to gain an understanding of the preferences and viewpoints of others; the second step is for individuals to gain an appreciation of the implications of their own initial preferences in relation to the other participants, thereby presenting the possibility for expansion or for modification. Indeed, research by Myers and Lamm (1976) concluded that learning in a passive context is not sufficient to produce a change in attitudes; mere awareness of differences among group members is insufficient. It is only through critical self-reflection that participants in a fair process can be more confident in the quality of its product. Thus, while one can never be in a position to know with absolute certainty that a successful "enlightened consensus" has been reached (as opposed to some sort of inauthentic consensus, where parties may agree but their relationships are still not desirable), the combination of procedural rationality and critical content reflection serve as cross-validating mechanisms.

Finally, the ultimate *outcome* of the decision is of importance in understanding whether or not an improvement was produced. In the absence of knowledge about alignment of the group recommendations with the preferences of society at large, the *utilization* of the results by participants and clients can serve as an indicator for this dimension.

FINDINGS FROM A STUDY OF COLLABORATIVE PROBLEM SOLVING

A study was conducted concerning the effectiveness of an approach to making collective judgments (Hart, 1983). This offered the opportunity both to utilize and refine the broad criteria for collective decision quality proposed above.

Methods and Measures

The research focused on six applications of an ad hoc, collaborative approach to complex problems.[2] The approach, which consists of a sequence of eight steps, is divided into three stages. The first stage is oriented toward expanding the range of information available through literature review, networking, and surveys. Stage 2 then narrows to a highly structured workshop with a relatively small (and therefore manageable) number of participants. The results of this effort are, in turn, disseminated to a wider audience for review, comment, and implementation or use (Stage 3).

In the first stage of the process, it is necessary for staff to become familiar with the needs of the client and the nature and scope of the problem. Through networking and surveys, the staff identifies the key issues as well as identifying important perspectives and stakeholders. This enables the staff (in conjunction with the client) to better "bound" the problem for further consideration.

The workshop stage is then used as a tool for focusing attention on a few key questions about the problem identified in the first stage; it is based upon the notion that interaction among diverse yet cooperating individuals can produce creative, integrative solutions that would not be possible by communicating with each participant separately. The workshops are task oriented, ad hoc sessions involving 15–40 people providing key perspectives on the issue at hand. The workshops make extensive use of structured group techniques, particularly nominal group technique (Delbecq, Van de Ven, & Gustafson, 1975) and brainwriting (Geschka, Schaude, & Schlicksupp, 1973), in combination with presentations and summaries in large (plenary) sessions.

The third stage is largely a reporting and dissemination stage; it is geared toward the needs of the client. Depending upon the nature of the particular problem or issue, this may involve the preparation of targeted technical reports solely for the client or it may entail broader publications intended for dissemination to a much wider audience.

In an effort to evaluate the effectiveness of this collaborative problem-solving approach, data were collected from participants and clients in six past applications. All six projects were funded by public agencies and ranged in topic area from a citizens' review of proposed criteria for power plant site selection to identification of participatory management procedures for use within a State Department of Labor.[3] All of the projects had certain characteristics in common, which enabled the use of common data collection instruments. Common features included the following.

[2] For a more detailed description of the approach, the reader is referred to Hart, 1983, and Enk and Hart, 1984.

[3] See Hart, 1983, for details about the six projects.

- All projects were third-party oriented (i.e., all involved a "staff" relationship with a client organization). None of the projects were conducted "in-house."
- All projects involved some form of initial staff networking and research in order to scope the project and prepare for the workshop.
- All projects involved staff in synthesis, report writing, and follow-up.
- All projects made use of interactive workshops as a major component.
- Every workshop made use of small group work in task teams with plenary sessions serving as forums for information sharing and presentation.
- All workshops requested that participants be in residence (i.e., to be housed on site and take meals together).

The evaluation entailed three forms of data collection: mail question-naires to the 170 workshop participants in the six projects (80 were re-turned and used in data analysis), telephone interviews with 2–3 key individuals in each of the six client organizations, and a group interview with the project staff. The purpose of the mail questionnaire was to gain participants' insights, impressions, and criticisms of the approach after the passage of time (1–2 years). Client and staff interviews were used for comparative analysis but are not reported in this paper, given their focus on particular projects.

Based upon the criteria articulated in the previous section, five basic theoretical constructs were conceptualized and used in instrument design.

1. Mix of Activities

The purpose of this construct was to understand what parts of the approach stood out in participants' minds as most useful and important. It was concerned with the actual activities and techniques used during the projects. For the overall approach, this included an assessment of the relative contributions of the various steps—literature reviews, partici-pant profiles, project materials, the workshop itself, and interim and final reports. Within the context of the individual workshops, this entailed an assessment of the relative contributions of the various activities and pro-cedures—large group (plenary) sessions, structured small group work, informal time (e.g., social hours), meals, and unstructured time (e.g., evenings). As noted above, two structured group techniques were exten-sively used in the six projects: nominal group technique (see Delbecq et al., 1975) and brainwriting (see Geschka et al., 1973).

2. The Personal Project Experience

This construct was concerned with the more affective or evaluative aspects of participation in the projects. It attempted to tap how partici-pants felt, in retrospect, about the experience from a personal point of view (e.g., excited, overwhelmed, stifled); this was considered important since, as the proposed criteria suggest, participant satisfaction with the process of problem solving seems to be an important facet to overall "success" of the collective effort.

3. *Information and Learning*

This construct attempted to assess the extent to which involvement in the projects resulted in the exchange and gain of information among participants; it was concerned with the level and type of information gain and not the larger question of whether individual perceptions of the problem or issue were changed or expanded. As the proposed criteria suggest, the exchange of information and understanding of new perspectives is thought to be a prerequisite to effective collective decision making.

4. *Image Expansion*

Unlike Construct 3 which dealt with information gain and exchange per se, the purpose of this construct was to assess the extent to which participation in the projects resulted in an expanded image of the problem or issue among participants. Associated with such expansion was the extent to which project involvement altered or expanded participants' conception of the appropriate responses or solution to the problem. As the proposed criteria suggest, the extent to which a decision-making approach fosters such "image expansion" is thought to be a critical dimension to the success of the effort. A further goal of collective decision making is the achievement of an interactively derived outcome which is acceptable to all and greater than the sum of the initial perspectives of the participants. Questions were designed to assess the extent to which such synergy and integration of differing viewpoints took place.

5. *Implementation/Utility*

This construct addressed the extent and type of use associated with each of the projects. The concern here was not for the "quality" of either the process or the outcome but rather its effect—particularly within the client organization. Included in this construct, however, was the extent to which involvement in the projects fostered personal use of the results as well as the development of personal contacts. Participants were also asked to make specific suggestions for how utilization of the project or its results could have been increased.

In an effort to operationalize the above a priori constructs, a number of scaled questionnaire items were constructed. These were divided into those items related to Construct 1 (mix of activities) and those items designed to tap the other four constructs.[4] The latter are henceforth referred to as the *functional* items, while the former are called the *structural* items.

The questionnaire items were processed through dimensional analyses

[4] Copies of the instrument can be obtained from the author upon request.

for the purposes of data reduction and the construction of a factor structure based upon actual responses. Rather than assuming that the a priori conceptualization (i.e., the five constructs) is the most appropriate, dimensional analysis allows one to explore, based upon actual responses, the adequacy of the researcher's initial categories and measures. Thus, dimensional analysis serves as a way of enhancing knowledge of theoretical frameworks by putting one more "in touch" with the data.

To help ensure validity, two procedures were selected that use distinctively different algorithms (see Kaplan, 1974). The first of the two programs used was the Smallest Space Analysis III program (SSA III), which is a nonmetric factor analysis program (Lingoes, 1972; Lingoes & Guttman, 1967). The second procedure used was a metric, hierarchical cluster analysis program called ICLUST (Kulik, Revelle, & Kulik, 1970). The results of both programs were compared in order to identify and resolve conflicting messages before decisions about the final dimensions were made. Generally, items clustering differently in the two programs were dropped. But in a few cases, conflicting single items were retained on theoretical grounds.

The factors resulting from the two dimensional analyses were then analyzed for interrelationships using single-linkage cluster analysis (Bailey, 1975). The results are presented in the form of a Correlogram.

RESULTS

The structural dimensionalization was based upon responses to the items dealing with specific project and workshop activities (Construct 1). Its results provide insight into which procedures and activities are most conducive to the development of mutually acceptable collective judgments. The functional dimensionalization was based on the items pertaining to the overall experience, information and learning, image expansion, and utilization (the other four constructs). Its results provide insight into the criteria appropriate for judging the quality of collective judgments. Finally, the single-linkage cluster analysis combined the factors resulting from the above two dimensional analyses. Its results, presented in the form of a Correlogram, suggest how evaluational (functional) and procedural (structural) factors interact.

1. The Structural Dimensions

The eight structural dimensions emerging from the analysis are described in Table 1 in roughly sequential order (i.e., as they would occur over the course of a project). Thus, the survey and information dimensions consist of activities which precede the conduct of the workshop while the next four dimensions all reflect various aspects of the workshop itself. The reports dimension describes activities which occur subsequent

TABLE 1
STRUCTURAL DIMENSIONS

Dimension or item	Mean	SD	n	Items included
Survey	3.1	1.1	62	Questionnaire/profile, interview/ profile
Information	3.0	1.0	66	Preworkshop material, literature review/bibliography
Presentation	3.8	0.8	78	Large group presentations, introductory plenary
Group techniques	4.2	0.7	79	Brainwriting, nominal group technique, the workshop
Free time	4.1	0.9	76	Informal time, meals, unstructured time
Plenary	3.5	0.9	75	Large group discussion, field trip/ site visit, final plenary
Reports	3.3	1.1	65	Interim reports, final report
Minority report	3.4	1.0	68	The filing of dissenting statements[a]

[a] Dimension based upon a single item.

to the workshop phase. Finally, the minority reports dimension at the bottom of the table was included as a single item due to its practical and theoretical relevance.

Overall, the group techniques and free time dimensions (associated with the workshop) were rated as significantly more important than any of the other dimensions. Participants apparently felt that the interplay between structured small group tasks and informal conversation during breaks, meals, and evenings was important to the success of the collaborative efforts. This is not surprising considering that the small group sessions afford participants the greatest opportunity to interact and be heard. The importance of the free time activities are also probably related to the opportunity for personal and professional bonds to be developed during such periods of informal interaction.

Activities associated with the large (plenary) group split into two dimensions, with participants rating the items associated with the presentation dimension significantly higher than those within the plenary dimension. This suggests that participants sensed a difference between the various large group activities, that the size of the group was not as critical as the function of the session. Apparently those large group activities associated with participant introductions or presentations were evaluated more positively than those having to do with other functions.

In sharp distinction to the above four workshop dimensions, the survey, information, reports, and minority reports dimensions were rated significantly lower, with greater variability, and by fewer people. This result

suggests that the activities preceding and following the workshops are not viewed by participants as being as important as the workshop itself. The efforts implicit in the survey dimension, for example, are heavily oriented toward staff planning rather than participant use. Similarly, the reports dimension rating probably reflects the extent to which report writing is targeted at the client rather than at participants. Also, since only a few highly motivated or disgruntled people exercised the option of filing a minority report, it is not surprising that there is mixed feeling about its overall importance as an activity. In addition to the above considerations, the low number of people responding to these four dimensions (all with n's in the 60s) suggests that they were not as memorable as some of the other aspects of the projects (especially the workshop).

2. The Functional Dimensions

The nine functional dimensions emerging from the analysis are listed in Table 2. For each dimension, the table includes a description of the items from the questionnaire listed in descending order of loading, the mean and standard deviation, and the number of complete cases available for use in the calculations.

The dimensions add richness to the more aggregated notions contained within the constructs. The first four dimensions (the uppermost portion of Table 2), for example, can be interpreted as representing various aspects of the "personal project experience" and "information and learning" constructs described above. Similarly, the next four dimensions can be interpreted as representing the concerns outlined in the "image expansion" construct. The utilization dimension at the bottom of the table was included as a single item due to its theoretical relevance and directly reflects this construct.

The ratings indicate that participants generally felt quite positively about the decision process. The learning and collaboration dimensions were rated quite highly (3.6 and 3.5), while the abusive and bombarded dimensions received significantly lower (although more variable) ratings. Those dimensions dealing with the problem or issue level, however, were not as clearly interpretable. The ratings on the integration, image expansion, and synergy dimensions suggest, however, that participants felt that the approach generally succeeded at broadening people's perspectives and moving them toward an integrative, consensus response. Indeed, the ratings for these three dimensions were significantly higher than those for the complexity and utilization dimensions. The high standard deviation on the synergy dimension, however, suggests that while some people felt very strongly that the approach led to superior results or recommendations, others were less enthusiastic in their assessment. Finally, participants rated the utilization dimension quite low (2.7). Again, the high

TABLE 2
FUNCTIONAL DIMENSIONS

Dimension or item	Mean	SD	n	Items included
Learning	3.6	0.7	80	Educated, enlightened, informed, challenged, personal information gain, excited, incorporation of contributions, enjoyable
Collaboration	3.5	0.6	79	Cooperative, integration of diverse viewpoints, incorporation of unpopular views, fair
Abusive	2.0	0.9	78	Manipulated, coercive, heavy handed, stifled, pressure to conform
Bombarded	1.8	0.9	80	Confused, intimidated, overwhelmed
Integration	3.3	0.7	80	Effective, sharing of information, shared image, stable solution, efficient, comprehensive, best combination of responses
Image expansion	3.2	0.8	77	Uncertainty, what is not known, number of factors involved, alternative theories, decision-making context, needed next steps
Complexity	2.9	0.9	77	Scope, complexity of the problem or issue
Synergy	3.2	1.2	74	Response greater than the sum of individual contributions[a]
Utilization	2.7	1.0	63	Use of the results or recommendations[a]

[a] Dimension based upon a single item.

standard deviation here suggests that there are some strongly differing images about the extent of use of project results. The low number of people responding to this question (63) would also suggest that many people lacked either the information or motivation to answer the question.

3. Intercorrelation of the Dimensions

To get a sense of how the above 17 dimensions interrelate, a single-linkage cluster analysis was performed (Bailey, 1975). The product of such an analysis—a Correlogram—is derived directly from a matrix of bivariate relationships. In this case, a correlation matrix of Pearson r's using pairwise deletion of missing data to preserve as many cases as possible was used. The results of the analysis are summarized in Fig. 1. A cutoff level of .47 was used in constructing the Correlogram.

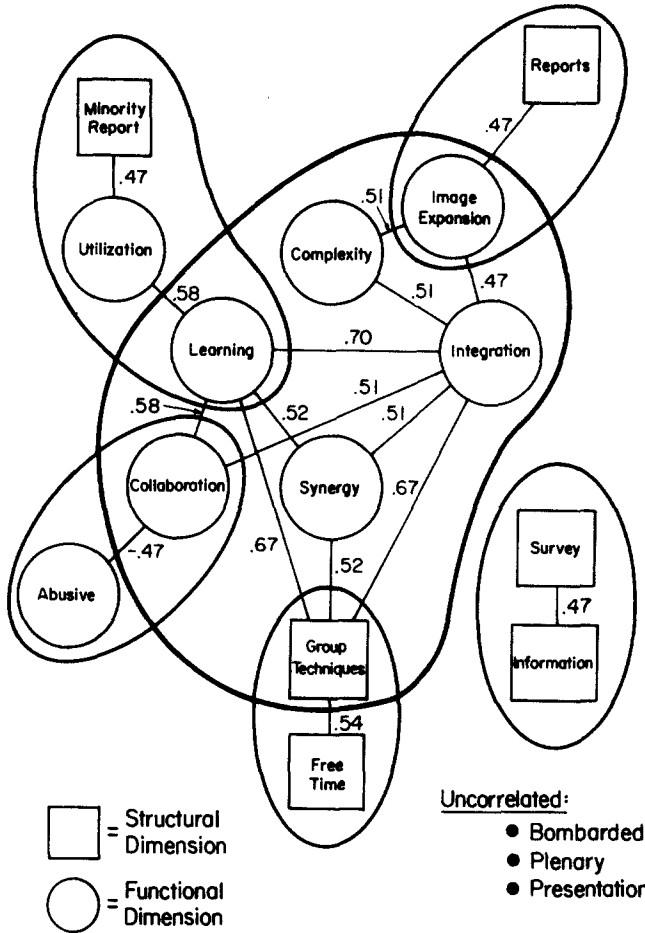The Correlogram can be interpreted at two different levels. First, one

FIG. 1. Correlogram of the dimensions.

large cluster seems quite evident, consisting of those dimensions encircled by the thick line. This large cluster has the "triad" of learning, integration and group techniques as its center (all with correlations of more than .65). Thus, the structured group aspects of the workshops seem to be highly correlated with personal learning and consensus building necessary in order for an integrative solution to emerge. This conclusion is further corroborated by the other dimensions which more loosely interrelate with the above three. These include (a) collaboration; (b) complexity of the situation; (c) diversity of the interests and potential solutions; (d) synergy of response; and (e) image expansion.

The second level of interpretation involves subclusters or secondary "nuclei" associated with the Correlogram. Those emerging from this

analysis tend to complement the interpretation of the large central cluster made above. For example, there is a strong *inverse* relationship between the collaboration and abusive dimensions. Thus, those who tend to rate the central cluster items highly tend to rate the items in the abusive dimension quite low. Also, there is a fairly strong relationship between the group techniques and free time dimensions, confirming our earlier suspicion about the apparent relationship between these two dimensions. The other two subclusters are not as readily interpretable, although one relates the utilization dimension to the central cluster while the other relates reporting to the main cluster. Not unexpectedly, the preworkshop activities associated with the survey and information dimensions are clearly related to each other but do not tend to cluster with any of the other items.

Finally, three of the dimensions (bombarded, plenary, and presentation) are not strongly related with any other dimension or cluster. While this does not necessarily reflect on their perceived importance, it does suggest that they are tapping domains not covered by the other dimensions.

SUMMARY AND DISCUSSION

The review of the literature in the beginning of this paper suggested that there were three broad and interrelated criteria for judging the quality of collective judgments (defined as problems without correct answers, requiring the integration of diverse points of view): (a) the acceptability of the *process* by which the decision was reached, (b) the quality or *content* of the decision, and (c) the *outcome* or effect of the decision.

Recognizing that these criteria were broad, a set of five theoretical constructs was developed to aid in the construction of a survey questionnaire for evaluating the effectiveness of six cases of ad hoc collective problem solving. Dimensional analyses conducted on the large number of structural (activities) and functional (process) items produced a manageable number of factors based upon actual responses which were more refined than the a priori conceptualizations.

Consistent with much of the literature summarized in the beginning section of this paper, the results of the dimensional analyses suggest that a high-quality collective judgment begins with a shared image of the problem or issue among participants; this is achieved through constructive interaction and mutual reperception of the situation. Through learning and expanding participants' image of the problem, it is then possible to create solutions which are greater than the sum of the initial individual positions. All this must be accomplished through a collaborative process deemed fair and noncoercive by all involved.

As the Correlogram (Fig. 1) showed, there was a strong relationship between participant evaluations of learning, integration, and group techniques. Also highly interrelated with the above three dimensions were

participant assessments of collaboration, image expansion, and an appreciation for the complexity of issues. Furthermore, there was a strong inverse relationship between ratings of collaboration and abusiveness, suggesting that those who rated the former highly also tended to feel the least abused.

Group techniques played a particularly important part in the configuration of dimensions presented in the Correlogram. Indeed, it appears that structured group procedures facilitated the personal learning and behavioral norms necessary for a creative, integrative solution to emerge in the ad hoc groups examined in this study. While these results would seem to be quite generalizable to other ad hoc group problem solving situations, they might be quite different for permanent groups where problem solving and behavioral norms have already been established.

The literature on applied group techniques (e.g., Van de Ven & Delbecq, 1974; Gustafson *et al.*, 1973) suggests that freely interacting groups tend to inhibit creative thinking, producing fewer problem dimensions and a smaller number of different kinds of solutions than do groups in which members are constrained from verbal interaction during the generation of ideas (nominal groups). The findings in this study are consistent with this conclusion; ratings for the structured group activities were significantly higher than for any of the other project or workshop activities with the exception of free time (breaks, meals, informal interaction). The importance of free time may derive largely from the opportunity it presents for rest, individual reflection, and informal social interaction in the context of collaborative problem solving. There is no way of knowing, however, whether the lower ratings for the plenary (large) group activities were due to the espoused inhibiting properties of larger groups (e.g., Bales *et al.*, 1951), or due to other reasons such as boredom, fatigue, or reticence.

The higher ratings for plenary sessions emphasizing presentations by other participants rather than discussion or evaluation does, however, suggest that the function of the large group session is an important factor. Apparently participants prefer plenary sessions as forums for organized information transfer and not discussion or critical evaluation. Many participants, for example, commented that the reports of the small group tasks in the plenary session would be more useful if people had adequate time to organize the presentations so that they were meaningful to others. Thus, prepared presentations were among the most preferred uses of the plenary mode.

CONCLUSIONS

The results of this analysis suggest that a theoretically based yet empirically derived set of evaluative factors for collective judgments in ad hoc groups is both possible and meaningful. The use of dimensional anal-

yses facilitated the reduction of a large number of survey items to a smaller and more manageable set of factors based upon actual responses. In addition to developing empirically a set of evaluative dimensions, the results also provided some insights into the most effective procedures for the achievement of high-quality group decisions.

In particular, the data indicated the usefulness of structured small group techniques (e.g., nominal group techniques) where integration is required among diverse perspectives. Such techniques seem to be particularly useful in establishing the behavioral norms necessary for effective ad hoc group functioning. Informal interaction among participants (e.g., meals, breaks, evenings) was shown to be an important complement to the structured sessions. Finally, the results suggested that use of the traditional large group (plenary) format to discuss or evaluate issues should be discouraged; plenary sessions should emphasize organized information transfer.

While this paper has explored possible criteria and proposed quantitative dimensions for assessing the relative effectiveness and quality of ad hoc, collective judgments, it has not examined these issues for established groups, nor has it considered the effect of key independent variables (e.g., situational differences in the decision settings or differences in participant characteristics). Future research might test the stability of the dimensions developed here in other decision-making settings. Another fruitful line of future research would be to examine differences in the patterns of correlation among evaluative criteria for established versus ad hoc groups.

Indeed, the results of this analysis suggest that much can be learned about group decision making in real-life settings. More work is needed in this area to test the validity of existing laboratory research findings and to ground more firmly our understanding of collective decision processes in reality.

REFERENCES

Arrow, K. (1951). *Social choice and individual values*. New Haven: Yale Univ. Press.
Axelrod, R. (Ed.). (1976). *Structure of decision*. Princeton, NJ: Princeton Univ. Press.
Bailey, K. (1975). Cluster analysis. In D. Heise (Ed.), *Sociological Methodology–1975*. San Francisco: Jossey–Bass.
Bales, R. et al. (1951). Channels of communication in small groups. *American Sociological Review*, 16, 461–468.
Balke, W., Hammond, K., & Meyer, G. (1973). An alternate approach to labor management relations. *Administrative Science Quarterly*, 18, 311–327.
Dalton, G. (1969, May). *Influence and organizational change*. Paper presented at the Conference on Organizational Behavioral Models, Kent State University.
Delbecq, A., Van de Ven, A., & Gustafson, D. (1975). *Group techniques for program planning*. Glenview: Scott, Foresman.
Drucker, P. (1980). *Managing in turbulent times*. New York: Harper & Row.
Einhorn, H., Hogarth, R., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin*, 84, 158–172.

Eiseman, J. (1978). Reconciling incompatible positions. *Journal of Applied Behavioral Science*, 14, 133–150.

Enk, G., & Hart, S. (1985). An eight-step approach to strategic problem solving. *Human Systems Management*, 5, in press.

Festinger, L. (1954). A Theory of social comparison processes. *Human Relations*, 7, 117–140.

Fisher, R., & Ury, W. (1981). *Getting to yes: Negotiating agreement without giving in.* Boston: Houghton Mifflin.

Follett, M. (1924). *Creative experience.* New York: Longmans, Green.

Geschka, H., Schaude, G., & Schlicksupp, H. (1973, August). Modern techniques for solving problems. *Chemical Engineering*, 6, 91–97.

Gustafson, D. et al. (1973) A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, delphi groups and nominal groups. *Organizational Behavior and Human Performance*, 9, 280–291.

Hall, J. (1971). Decisions, Decisions, Decisions. *Psychology Today*, 5, 51–88.

Hall, J., & Watson, W. (1970). The effects of a normative intervention on group decision-making performance. *Human Relations*, 23, 299–317.

Hammond, K., & Adelman, L. (1976). Science, values, and human judgment. *Science (Washington, D.C.)*, 194, 389–396.

Hart, S. (1983) *Strategic problem solving in turbulent environments: A description and evaluation.* Ph.D. dissertation, University of Michigan.

Hill, B. (1982). An analysis of conflict resolution techniques. *Journal of Conflict Resolution*, 26, 109–138.

Hobbs, B., & Voelker, A. (1978). *Analytical multiobjective decision making techniques and power plant siting.* Oak Ridge: ORNL.

Hoffman, L. R. (1979). Applying experimental research on group problem solving to organizations. *Journal of Applied Behavioral Science*, 15, 375–391.

Kaplan, S., & Kaplan, R. (1982). *Cognition and environment.* New York: Praeger.

Kaplan, R. (1974). A strategy for dimensional analysis. In D. Carson (Ed.), *Man–environment interactions.* Stroudsburg: Dowden, Hutchison & Ross.

Kulik, J., Revelle, W., & Kulik, C-L. (1970). *Scale construction by hierarchical cluster analysis.* Ann Arbor: University of Michigan.

Levi, A., & Benjamin, A. (1977). Focus and flexibility in a model of conflict resolution. *Journal of Conflict Resolution*, 21, 405–425.

Likert, R. (1976). *New ways of managing conflict.* New York: McGraw–Hill.

Likert, R., & Likert, J. (1978). A method for coping with conflict in problem solving groups. *Group and Organizational Studies*, 3, 427–434.

Lingoes. J. (1972). A general survey of the Guttman–Lingoes non-metric program series. In A. Shepard, K. Romney, & S. Nerlove (Eds.), *Multidimensional Scaling.* New York: Seminar.

Lingoes, J., & Guttman, L. (1967). Nonmetric Factor Analysis: A Rank Reducing Alternative to Linear Factor Analysis. *Multivariate Behavioral Research*, 2, 485–505.

McKenna, C. (1980). *Quantitative methods for public decision making.* New York: McGraw–Hill.

Mitchell, C. (1973). Conflict resolution and controlled communication: Some further comments. *Journal of Peace Research*, 10, 123–132.

Mitroff, I., & Emshoff, J. (1979). On strategic assumption making: A dialectical approach to policy and planning. *Academy of Management Review*, 4, 1–12.

Myers, D., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83, 602–627.

Nagel, S. (1980). The means may be a goal. *Policy Studies Journal*, 9, 567–578.

Nutt, P. (1979). On the quality and acceptance of plans drawn by a consortium. *The Journal of Applied Behavioral Science*, 15, 7–21.

Pareto, V. (1916). *The mind and society* (A. Livingston & A. Bongiorno, Trans.). New York: Harcourt, Brace.

Press, S. (1978). Qualitative Controlled feedback for forming group judgments and making decisions. *Journal of the American Statistical Association, 73*, 526–535.

Radford, K. (1977). *Complex decision problems: An integrated strategy for their resolution.* Reston: Reston Publishing Co.

Rein, M. (1976). *Social science and public policy.* New York: Penguin.

Stewart, T., & Gelberd, L. (1976). Analysis of judgment policy: A new approach to citizen participation in planning. *Journal of the American Institute of Planners, 42*, 33–41.

Summer, C. (1980). *Strategic behavior in business and government.* Boston: Little, Brown.

Van de Ven, A., & Delbecq, A. (1971). Nominal versus interacting group processes for committee decision making effectiveness. *Academy of Management Journal, 14*, 203–212.

Vinokur, A., & Burnstein, E. (1974). Effects of partially shared persuasive arguments on group-induced shifts: A group problem solving approach. *Journal of Personality and Social Psychology, 29*, 305–315.

Zelany, M. (1982). *Multiple criterion decision making.* New York: McGraw–Hill.