

AD-A154 201

AN EMPIRICAL STUDY OF A NEW METHOD FOR FORMING GROUP
JUDGMENTS: QUALITATIVE CONTROLLED FEEDBACK(U) RAND CORP
SANTA MONICA CA S J PRESS ET AL. MAY 79 RAND/P-6333
SBI-AD-F630 632 F/G 5/10

1/1

UNCLASSIFIED

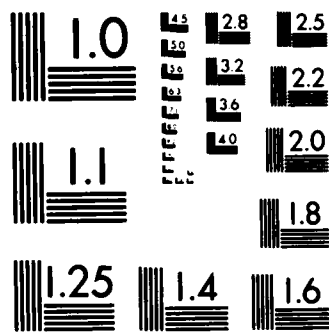
F/G 5/10

NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A154 201

AN EMPIRICAL STUDY OF A NEW METHOD FOR FORMING GROUP JUDGMENTS:
QUALITATIVE CONTROLLED FEEDBACK

S. J. Press, M. W. Ali, Chung-fang Elizabeth Yang

May 1979

DTIC FILE COPY



DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

P-6333

85 5 28 08 M

The Rand Paper Series

Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406

AN EMPIRICAL STUDY OF A NEW METHOD FOR FORMING
GROUP JUDGMENTS: QUALITATIVE CONTROLLED FEEDBACK

By

S.J. Press, M.W. Ali, and Chung-fang Elizabeth Yang

S. James Press and Mirza W. Ali
Department of Statistics
University of California
Riverside, California 92521

Chung-fang Elizabeth Yang
Department of Marketing
University of Southern California
Los Angeles, California 90007



Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

ABSTRACT

The Qualitative Controlled Feedback (QCF) method was developed by Press (1978a) to assist policy makers in forming judgments and making decisions that reflect the careful interactive reasoning and arguments of all of the members of a group or population. Since the QCF method involves controlled feedback, it tends to minimize the effects of face-to-face group interaction pressures. Since the feedback is "qualitative," however, the procedure tends not to artificially induce a consensus on the group. This paper summarizes a feasibility study of the procedure. A sample of 111 faculty and staff members of the University of British Columbia participated in the testing of the method. The participants were asked to make a judgment on the importance of building an aquatic center on campus. A second (control) group of 89 faculty and staff members was surveyed on the same issue, but using the conventional survey method, that is, no feedback. It was observed that Qualitative Controlled Feedback created a good interaction (in the sense of exchanging arguments and reasons) among group members. Changes in judgments occurred as subjects went from one stage to another after having qualitative feedback of information. It was also found that the judgments given by the subjects in the qualitative controlled feedback group were distributed quite differently from those given by the control group. The method suggests a significant new way of collecting and interpreting group judgments.

1. INTRODUCTION

This paper describes an empirical study we carried out to apply the new method of qualitative controlled feedback (QCF) to a realistic problem and situation. The QCF method was developed by Press (1978a) to help policy makers order priorities by assessing reasoned individual judgments after the individuals have benefited from group interactions. The method was extended to the multivariate case of multiple question situations in Press (1978b). Bayesian inferences using QCF are treated in Press (1979). The QCF method differs from many other, already existing, group judgment or decision-making methods in that: (1) it does not require the members involved in the group interactions to reach a consensus of "group decision;" (2) it does not allow the group members to have face-to-face interactions when they are making judgments; and (3) it does not permit the group members to know the identities of fellow group members, or their individual judgments on the issues in question. In fact, the method was originally proposed to meet the criticisms of some of the traditional group judgment forming methods. These methods usually involve a group of individuals discussing an issue in question on a face-to-face basis. The ultimate goal of the discussion is to obtain a unanimous judgment or decision from the group, namely, the "group judgment." In situations involving judgment or decision making, it is generally felt to be desirable to examine the alternatives in light of opposing arguments and reasons. It is also generally understood that in situations involving group judgment or decision-making, the group members will generate a greater number of alternatives, contrasting arguments, and reasons, than would be generated in situations involving individual judgment. Social psychologists (e.g., Collins and

Gnetzknow, 1964), however, have found that group judgments perform better only when the interpersonal environment fosters independent thinking, free expression of ideas and views, and the sense of respect in one member about other members' views. The QCF method is designed to capitalize on this aspect of group interaction.

There are various obstacles in the interpersonal environment standing in the way of free interaction among group members in the traditional group judgment methods. The main sources of the obstacles are status hierarchy, personality styles, leadership styles, pressure of group conformity, and power of dominance. The impact of the obstacles may be that alternative views do not get represented, and also that undue weights are sometimes given to irrelevant factors. Effects of some of these factors (such as group cohesiveness, group norm, and leadership) on the performance of face-to-face decision-making groups, have been examined by Janis (1972). He has found, after analyzing the deliberations of actual world decision-making groups, that these factors may give rise to a phenomenon he called, "groupthink." Groupthink is said to occur in a group if, in spite of their high intellectual ability, the group members refrain from critical thinking, and manifest concurrence (or consensus) seeking behavior with an illusory optimism about the success of their decision.

The QCF method is designed to minimize such obstacles. The Delphi method of forming a group consensus (Dalkey and Helmer, 1951) is often used to forecast future technological events based upon the combined judgments of experts. The QCF method shares with Delphi the use of "controlled feedback." Using controlled feedback, a panel of experts may be asked to give judgments, individually and independently, on questions of interest.

The judgments requested are generally numerical, and are collected by an intermediary. In Delphi, the intermediary then computes a summary of some quantitative measures of the response distribution, such as the mean, inter-quartile range, etc. He then feeds back one or more of the measures (sometimes the entire distribution) of the judgments to each member on the panel individually. We call such a data collection protocol, quantitative controlled feedback. Each member is then asked again to give his judgment independently. Anonymity of the panel members is preserved, and this process of feedback and soliciting judgment is continued until convergence (consensus) is reached. In QCF, no quantitative measures are fed back to panel; instead, panelists are asked to provide reasons for their answers, the reasons are merged into a composite, and the composite is fed back to the panelists. This is called qualitative controlled feedback.

While the Delphi method reduces the effects of some of the undesirable factors mentioned by Janis (1972) it has been criticized as unsuitable for social research applications (see, e.g., Sackman, 1975). A main objection is related to the idea of quantitative feedback. When quantitative measures, such as the mean, are fed back, the panelists are psychologically pressured to shift their answer on the next round towards the given mean. Social psychologists are very familiar with this phenomenon of group conformity (e.g., Ash, 1962). In other words, members of a Delphi panel are artificially "forced" to agree, possibly on the wrong answers, and probably without careful consideration of all arguments, and all alternatives. In social science research where a true answer, or an optimal solution to a problem often does not exist, it is very important to ensure that every member of a group is presented with all points of view, and all arguments

for and against these points, before the panelist makes a judgment. Since QCF involves feeding back only reasons supporting the panelists' individual judgments, panelists are likely to shift their responses on subsequent rounds only on the basis of being persuaded by the logic of the reasons generated by the panel (and not because they know the average value of the group's response, since this will not be known by panel members).

A related issue is whether or not a single "group judgment" needs to be reached at the end of a group judgment session. As mentioned earlier, social research problems seldom have a true answer, or an optimal solution. The resultant unanimous judgments reached by a traditional face-to-face discussion group, or by a Delphi panel, often reflect artifacts, more than anything else. Furthermore, diversity of opinion among group members may be more meaningful to social scientists than (artificially induced) agreement. Additional arguments for the QCF approach and additional references to related literature is given in Press, 1978a.

Whether or not a newly proposed methodological procedure is easy to implement, is useful, and is workable, is a matter of empirical study. Accordingly, it was decided that we should try out the QCF procedure on a real problem in order to study the difficulties that might arise in practical situations and to see what kinds of results might be obtained. This study is detailed in Ali, 1978. The research is summarized, interpreted, and evaluated in the remainder of this paper.

2. BACKGROUND OF THE EMPIRICAL STUDY

2.1 History

In June 1976, the University of British Columbia was starting construction of an indoor aquatic center on the campus. Construction of this

center had been an issue of a good deal of controversy for more than six years within the university community (consisting of more than 25,000 students, faculty and staff). In spite of the involvement of such a large population and a considerable construction cost, little systematic effort was made to see how the community really felt about the importance of having the center built. Furthermore, the university was not committed to construct the center, beyond a small initial preliminary construction stage, which could have made use of the site for many other purposes.

Apparently the university community was divided into two major sub-groups: one in favor of constructing the center, and the other against construction. A third group of neutral and uniformed people also existed. This situation afforded us an opportunity to apply the QCF method to the problem of determining how important the university community felt it was to construct the center at that time.

It is well known that in the Vancouver area, rain occurs very frequently and that forty inches per year is typical. For some reason, however, the existing swimming pool at the University of British Columbia is outdoor, which prevents its usage by most of the university community a large portion of the time.

After considerable discussion over a period of about 6 years, the decision was made to construct the first phase of the center. It was planned that the center be constructed in two phases, and that the center be used for a variety of recreational and academic purposes. The total cost of construction was estimated to be 4.5 million dollars in 1974. It was decided that if the center was to be completed it would have to be financed by a combination of the University, Provincial and Federal Government grants, donations from university faculty, staff and students,

and contributions from the general public. Phase one would begin in June, 1976, and if support and funding did not grow in sufficient amounts, phase two would not occur, and the site would be used for other purposes.

The proposed site was also a matter of debate in the community. Since the construction of the center at the site would destroy some beautiful trees and scenic beauty of the site, and also the ground could be used for more needed academic buildings, objections were raised to construction of the center. Apart from these issues, it was also a question of debate as to whether it was really worthwhile, with regard to the alternative needs of the university community, to have such a large facility, by spending a large amount of money that could be used for more demanding needs, such as academic buildings (or other purposes). At that time, there were still considerable uncertainties and debates about the usefulness of the already started swimming pool construction and, most important of all, about whether or not the construction should proceed beyond the first phase.

2.2 Target Population

We decided to try out the QCF methodology on the swimming pool issue in the Spring of 1976. We decided that because the students were not likely to be available for questioning during the coming summer months, we would confine our attention to the less committed and less transient population of faculty and staff, comprising a total population of 5,019 including 2,194 faculty and 2,825 staff. We designed a questionnaire and pilot tested it on a small group (after which the survey instrument was modified). Simple random samples were selected from each of the two strata of the target population (the original numbers were originally proportional

to the population proportions, but for various logistical reasons, the samples turned out to be of about equal size).

In July, 1976, we began the first stage of our empirical study by mailing a questionnaire to each member of the randomly selected groups from our target population. In a cover letter, we asked the subjects to participate in a research study that would require them to answer the same question three distinct times (we choose three stages so that there would be a sufficiently large number of stages so as to permit us to study effects resulting from feedback, yet still have few enough stages so that non-response would be minimized). Of course in a non-academic exercise the process could be continued until it stabilizes, rather than operating it with a fixed, predesignated, number of stages.

In total, 111 subjects complied with our request by completing all three stages of the study. Subsequently, the second stage of this study took place in October, 1976, and the third stage in January, 1977. At the time we distributed our third-stage questionnaire to the group that had already completed two earlier stages, we also sent out a questionnaire to a control group of 110 subjects (of whom 89 responded) who had not previously been involved in the study. It should be noted that future applications of the QCF procedure could be carried out within hours, instead of months, with the aid of computers and pre-programmed reasons. We were handicapped by having to do this on a very part-time basis.

3. METHOD

3.1 The Questionnaire

At each of the three stages of the study, the main task of the participants was to answer a basic question (although the methodology is

applicable to studying many related questions simultaneously, it was decided that one question would suffice for a initial feasibility study). The question required them to make a judgment, and it read:

"How important (necessary) do you feel it is for the University of British Columbia to complete construction of an indoor aquatic center on the campus that would be available for use by students, faculty and staff and their families, and the general Vancouver community?"

The participants were instructed to answer this question on a 9-point rating scale. Each of the 9 points on the scale was given a numerical label as well as a descriptive label. They were:

00.0	extremely unimportant;
12.5	very unimportant;
25.0	moderately unimportant;
37.5	somewhat unimportant;
50.0	indifferent or neutral;
62.5	somewhat important;
75.0	moderately important;
87.5	very important;
100.0	extremely important.

(a) The First-stage Questionnaire. In addition to the basic question, the first-stage questionnaire included, on the cover page, some background information about the construction of the aquatic center, the methodology surrounding the research, the study objectives, and instructions for completing the questionnaire. In particular, it was emphasized that participants should complete the questionnaire independently (without consulting others).

Following the basic question, the participants were asked to list the reasons for their answering the basic question the way they did. At the end of the questionnaire, the participants were instructed to answer some

The fact that participants showed willingness to consider reasons supporting views opposite to their own stand demonstrates that the QCF method accomplishes what it was designed to do: to create an interactive environment where arguments representing different points of view can be considered objectively.

4.1.4 Non-response

To discuss non-response, it is convenient to return to Table 1 from which it can be seen that there was a sufficiently large proportion of non-response at each stage of the study which we could not safely ignore (17% at the first stage, 20% at the second stage, and 7% at the third stage). We were concerned as to whether or not the non-respondents were likely to cause any serious bias on the final stage distribution of responses. Throughout the study, there were a total of 66 non-respondents out of the 177 originally solicited at the first stage (for an overall response rate of 63%). Thirty of them did not answer our initial request for participation in the study, therefore we had no information at all about them. The remaining 33 of the non-respondents, however, dropped out of the study after they participated in the first stage of the study, and thus, by answering the subsidiary (demographic) questions asked, they had left us with some background information about them. This information provided us with an opportunity to compare the people who completed all three stages with the non-respondents.

To perform the comparison, we carried out a logistic regression analysis. The dependent variable in the analysis was a 0 or 1 variable indicating whether or not a participant had completed the entire study. There were ten independent (explanatory) variables in the regression

regarding the construction of the pool. If the participants had checked a number category between 0 and 37.5, they were classified as "opponents," i.e., those who did not think the construction of the aquatic center was important. If the participants checked number category 50, they were classified as "neutral." The 62.5 and above group was classified as "proponents." The reasons given by the participants were also conveniently arranged as "negative" reasons (those against the building of the center), and "positive" reasons (those supportive of constructing the building). These two sets of classifications provided us with a much clearer picture of the relationship between participants' responses and the reasons they presented. Table 8 gives the total numbers of positive and negative reasons given by people who responded to the issue favorably or unfavorably. Of course the same reasons are given many times by different subjects, while there are only 17 distinct reasons on the first stage, and 26 distinct ones on the second stage.

Not surprisingly, Table 8 clearly indicates that proponents tended to give positive reasons and opponents tended to give negative reasons, especially at the first stage. As the study progressed, more reasons were generated of both the negative and positive variety, and so reasons representing the other point of view (regardless of which one was held) emerged steadily, while the number of reasons supporting the participants' own stand also increased. One interesting result was that participants who gave neutral responses tended to provide negative reasons. This result suggests that such so-called "neutrals" were often not real neutrals at all, but rather were neutrals leaning toward the opponent side and for some reason were reluctant to express their true feelings.

is somewhat similar to what happens in an ideal face-to-face interaction where, when faced with new arguments and reasons, participants often abandon the arguments and reasons with which they entered the discussion and accept new ones, or, perhaps, while still holding to their original arguments they pick some new reasons to strengthen their position. It is of interest to observe and study this phenomenon in our new setting. Our discussion will be based on a set of conditional probabilities, called the transition probabilities for reasons.

Table 6 and 7, respectively, give the averaged transition probabilities of reasons from the first stage to the second stage and from the second to the third stage. We can see from the large diagonal elements of these tables that the probability was very high that a participant who gave (or did not give) a particular reason at one stage, still gave (or did not give) the same reason at the following stage. This result indicates that it is much more likely for a participant to repeat his last stage's reasons at the present stage. Furthermore, we have also found that 33% of the participants gave one reason, or more, at the second stage that they had not given at the first stage. This result provides further support for the position that the QCF method tends to stimulate the participants to reflect upon arguments for or against a position that they had not thought about in their original judgments.

(c) The Relationship Between Participants' Responses and Reasons.

Another important aspect of reason-giving behavior is how participants in different response categories differ in the types of reasons they give. In order to explore this relationship, we first divided our sample into three subgroups based on the participants' answers to the basic question

These results may indicate that the QCF method stimulated participants to think more about the issue. An alternative explanation, which probably accounts for part of the observed increase in reasons given, relates to the design of this study (see below).

At the first stage of the study each subject was asked to write his reasons independently, and thus, he had to undergo a process of introspecting and reasoning, which is not a very simple task for many people. Also, each subject had limited information at that time. These factors might have resulted in his giving very few reasons at the first stage. However, on the second and third stages, the participant was asked to check off reasons he agreed with from a composite list, and then to add any new ones he felt would be appropriate. We expect that whenever he came across a reason he agreed with, and (or) he could have used to support his present judgment, he would have checked that reason. This relatively effortless procedure, on the part of the participants in the second and third stages, might have contributed to the increase of the length of the reason lists also. However, since the standard deviations of the distributions also increased over the stages, we can safely say that participants did exercise care in checking more or fewer reasons, as they saw fit, and therefore have not been totally influenced by the ease of use of the method.

(b) Transition Probabilities of Reasons. From the second stage onward, a subject, while checking reasons from the composite list, sometimes added new reasons to his own previous stage slate of reasons, or dropped reasons from the slate. This phenomenon of adopting new reasons and abandoning old ones is indicative of a process of interaction, and

the third stage, and 10% changed on both stages. From tables 4 and 5 we also see that, starting from a diagonal cell, as we move to the left or to the right, the probabilities decrease, the highest probability occurring in the adjacent left or adjacent right cells. Thus when a subject changes response, it is most likely that he will move to one of the adjacent response categories. It is also worth noticing that changes took place only among categories 25 through 87.5; participants who fell into the two extreme categories at the previous stage, did not change their responses in the subsequent stage (these were the people least persuadable by feedback of reasons). Meanwhile, participants who responded at the neutral point (score of 50 at a previous stage), were most likely to be persuaded, and thus, to change their responses at the subsequent stage. There seems also to be a general trend of changing from a relatively favorable position on construction of the pool, to a relatively less favorable position.

4.1.3 Reasons Given by the Participant

As has been stated before, an important aspect of QCF is that at each stage, subjects not only answer the basic question, but also give reasons supporting their answer. The reasons generated by the participants should prove very useful in a decision making context to gain insight about opinion, and to better understand the attitudes of the group members. It is worthwhile to examine and understand some basic features of reasoning behavior of the participants.

(a) The Distributions of Number of Reasons Given. The data show that the average numbers of reasons given per person increased from 1.66, to 4.60, to 6.15, on successive stages; the standard deviations in the number of reasons given per person increased from 1.0, to 2.5, to 3.

In contrast to the means, the standard deviations increased steadily from stage 1 to stage 3, except for the faculty sample at the second stage (the standard deviation dropped from 26.8 to 25.8). When comparing the third sample standard deviations with the control group's, we found that for the faculty sample, the third stage sample standard deviation was smaller than that for the control group, and that the reverse is true for the staff sample.

4.1.2 Judgment Changes

Changes in individual judgments from stage to stage were not always reflected in changes in the judgment distributions. In other words, while individual judgments change, the distributions of judgments may, or may not, change. Therefore, to study judgment changes, it is necessary to carry out a detailed analysis of the changes at an individual level. First, we estimated the transition probabilities⁴ for changes in response, over stages, for a given individual. Table 4 gives the frequency counts and the estimated transition probabilities of participant's changes from one response at the first stage to another at the second stage. Table 5 gives the changes from the second stage to the third stage. Both tables are for the combined faculty and staff groups.

From the tables, note that the largest transition probabilities occur in the diagonal cells, indicating that a subject is most likely not to change his response. Nonetheless, 33 out of the 111 participants (30%) did change their responses at the second stage, and merely 23 changed their responses (21%) at the third stage. Through some calculations not shown here, we also found that more than 41% changed responses either on

basic differences when compared stage by stage. The bimodality character is prominent in faculty distributions from the very first stage, whereas it is only in the second stage that the staff distribution appeared to become bimodal. The latter underwent greater change from stage to stage, relative to that of faculty

These results may be interpreted to imply that the staff group had a rather unanimously favorable attitude toward the building of the aquatic center, but were gradually dissuaded by the opposing arguments, mainly given by the faculty group. This is a very interesting finding and there are many related social psychological questions that could be asked at this point. Due to the limited scope of this paper, however, we will demur on further discussion on these issues at this time. We conclude here only that the feedback effect seems to have worked very well.

(d) Means and Standard Deviations. Table 2 and 3 give the sample means and standard deviations of the response distributions for the faculty, staff, and control groups at all three stages. Note that in this context, the means are not as meaningful as the modes because most of the distributions found are bimodal. That is, although the means still denote the average response, they do not reflect the point where most of the mass of a distribution lies. In fact, the shifts of the locations and heights of the modes are more reflective of the distribution changes in this context. Nevertheless, we have noticed that the means decreased stage by stage for each of the groups, and for the combined group. The mean for the staff is always larger than that for the faculty. The control group means are near the first stage means (except for faculty).

although this could be due to sampling fluctuation. In figure 1(a), the modal values of the principal modes differ (75 in first stage and 87.5 in control) in the two curves. A chi-squared test for equality of the two distributions (due to small expected cell frequencies, some of the response categories were merged, resulting in a total of six categories) showed them to be different (at the 5% level). That is, the distributions of the first stage and the control group responses differ, signifying the presence of a time effect on the control group distribution.

Now note that since the control group questionnaires were administered simultaneously with the third stage, we may also conclude that the third stage distribution has a time effect. However, the third stage responses were obtained after information feedback; so this distribution may also have a feedback effect. In order to see this we compare the third stage distribution with the control distribution in Figure 1(c). A difference between these distributions would reflect the presence of a feedback effect in the third stage distribution (since feedback is the factor which differs in the two situations).

A chi-squared test of equality of these two distributions³ required rejection of the equality hypothesis at the 5% level of significance. We (marginally) conclude, therefore, that there is an effect of feedback. Thus, although the tests are difficult to carry out, there nonetheless is indication that both time effect and feedback effects were responsible for the observed changes in judgment.

(c) Faculty vs. Staff Subsamples. Figure 2(a), (b), and (c) give the frequency distributions at the three stages for the faculty and staff subsamples separately. A comparison between the two groups reveals some

the two stages. At the time we conducted the third stage research, the first phase of the construction of the aquatic center was well under way. The controversy about this issue had quieted down a little. The entire community's attitude toward the aquatic center issue might have changed because of this fact, or for other reasons that had nothing to do with the study at all. To determine whether the feedback or the time factor generated the observed changes we needed to compare the experimental group with the control group.

(b) Experimental vs. Control Group Responses. There are three things to consider about the control group. First, the control group data collection was carried out simultaneously with the third stage; second, participants in the control group are different from those of the experimental group (in that they have not received any feedback) but they form a random sample of the same population from which the first stage experimental group was drawn; and third, the control group panelists were given the identical questionnaire as the first stage group. Since the set of responses at the first stage, and that of the control group, may be considered to have come from the same population (if all other things were equal), the frequency distributions should conform to one another. If the distributions are different, however, the difference can only be attributed to time (that is, changes in attitude that have taken place in the population over time, because of factors that relate to the issue), since time is the only factor which is different in the two situations.

The dotted lines in Figure 1 are the same control group distribution superimposed on the experimental group distribution at each of the three stages. The control group does not appear to have a regular shape,

Responses to the basic question were conceptualized as being observations from some underlying continuous distribution of importance rating. Accordingly, the data were grouped, and histograms were plotted to estimate the underlying density. Figures 1, 2 give the sample frequency distributions of responses for each stage, for the faculty and staff combined and separately, and for the combined and control groups. Our first observation is that the experimental distributions in each of the three stages appears to be bimodal¹, meaning that there are two sub-groups of thought in the panel. It is helpful to note how the distributions changed from stage to stage. In going from the first stage to the second stage, the subgroup on the left side of the scale (unimportant side) became larger, the apparent mode shifted from 25 to 37.5, resulting in a higher modal ordinate (notice that the judgment group 37.5 had the smallest ordinate in the left subgroup on first stage)². On the other hand, the subgroup on the right side of the scale became smaller, the modal value remained about the same, but the ordinate (in contrast to the modal ordinate on the left side) became smaller. In the third stage, the mode on the left side returned to 25, and the subgroup became more concentrated, whereas, the right side subgroup remained almost unchanged. Note also that the ordinate at 50, the indifferent group, became smaller and smaller at every stage.

We have seen above that the judgment distributions have undergone change from stage to stage. We can think of two major reasons that might have generated these changes. They are: (1) feedback of new information (in terms of reasons) at every stage, and (2) time. The time effect means that a difference, found between the response distributions on two stages, is assumed to be caused, in part, by the difference of execution time of

limited use," "plenty of recreational facilities on campus," "alternative proposals for spending money," and "alternative for physical education department."

The preparation of the composite reason lists relied very much on the ability of the investigators to understand and interpret the reasons given by the participants; to combine and to differentiate; and finally, to paraphrase. This phase of the process is one which could be computerized perhaps by coding all reasons, using computer editors to block out various portions of prose, etc. This notion will be discussed in more detail in Section 4.2.1.

4. RESULTS AND DISCUSSION

In this section, we will first present and discuss the substantive findings of the study. Then, we will discuss some of the technical issues related to the use of the method.

4.1 Substantive Findings

4.1.1 Judgment Distributions

(a) Responses at Each Stage. In this section first we compare the judgment distributions among the three stages of the experimental group to see how the distributions change from stage to stage. These distributions are also of importance in understanding the judgment structure of the group. For example, they reveal such facts as: whether or not there exist subgroups of differing judgments; if there are, how divergent the subgroups are in their judgments; and so on. Second, we will compare the experimental group with the control group to see whether there is any effect of information feedback.

3.3 Preparation of the Composite Reason List

At the end of the first stage, the reasons every participant used to support his judgment were gathered. Similar reasons given by different participants, but given in paraphrased form, were combined, and reasons that consisted of too many words were shortened. On the average, panelists gave a total of 2, 5 and 6 reasons at the end of the first, second and third stages, respectively. After these preliminary editing procedures, reasons were carefully studied. We found that it was possible for us to conveniently classify these reasons into two basic categories, "pro" and "con;" "pro" meaning in favor of having an aquatic center, and "con" meaning against having such a center. Seventeen distinct reasons were placed into the composite at the beginning of the second stage, and nine new distinct reasons were added on the third stage for a composite of twenty-six distinct reasons presented at the beginning of the third stage.

The major reasons frequently cited within the pro category were: "year-round swimming facility (with emphasis for winter)," "need of physical education department for training purposes," "the center's social role as a mixing-up media between university people and the surrounding community," "need for general recreation," "shortage of swimming facilities in Vancouver," and "good publicity for the school." The reasons frequently cited within the con reasons could be identified as: "university is mainly an academic institution," "priority of spending money on academic rather than recreational buildings," "covering the existing pool," "abundance of indoor and outdoor pools in the city," "center's limited expected use to community," "large cost compared to

subsidiary questions. Moreover, the control group questionnaire was administered at the identical time as the third stage questionnaire so that differences between results from these two could be attributed to experimental effects, and not to changes in attitudes of the participants over time.

3.2 The Participants

As described earlier, the population under investigation was limited to Faculty and Staff members of the University. Those two groups of people are quite different with respect to type of job, educational background, and probably, but most importantly, with respect to the attitudes towards construction of the center. A faculty member may be expected to prefer spending money on a project that is of an academic nature to spending money on a project that is more of a recreational nature. A staff member, on the other hand, may not be that biased towards having an academic project. In short, on the average, a staff member may be expected, a priori, to attach greater importance to the center than that of the faculty member.

Table 1 gives a summary of the numbers of participants surveyed and the numbers who responded in each of the two strata and at each of the three survey stages. The last column gives the response rates. The last rows give the corresponding figures for the control group.

In all, 177 randomly selected faculty and staff members of the University were approached, and 111 of them went through the entire process required by the QCF method; these 111 formed the basic sample. The control group consisted of 89 participants who responded to the Control Group Questionnaire sent out to 110 randomly selected faculty and staff members.

subsidiary (background) questions that we thought might reflect their biases in answering the basic question, such as: academic status (faculty or staff); their sex; whether or not they know how to swim; how frequently they and their family would use the center; whether or not they live on campus; whether or not the participants commuted more than 30 minutes to the campus; whether or not they had a swimming pool nearby, or in their residence; whether or not they had already donated to the construction of the center; whether or not they would be willing to donate in the future if they had not yet donated (they were advised these questionnaires would not be used to solicit funds); and how much annual fee they would be willing to pay for the use of the center.

(b) The Second-Stage Questionnaire. All reasons provided by the panelists on the first stage were merged by the authors (this process is detailed in Section 3.3) into a single composite of reasons. This composite was presented to each panelist at the beginning of the second stage. Each panelist was also reminded of his earlier response to the basic question.

After answering the basic question, the panelists were requested to select reasons from the composite list that they had used to support their second stage answer, and in addition, they could give new reasons not contained in the list.

(c) The Third-Stage Questionnaire. The general procedure for conducting the third stage of the survey was the same as that in the second stage.

(d) The Control Group Questionnaire. The control group questionnaire was identical to the first stage questionnaire, including the

equation (namely the ten mentioned in Section 3.1): participant's status (faculty or staff); sex; whether or not his family would use the center; whether or not he had donated money; whether or not he had a swimming pool in or near his residence; whether or not he would ever use the center, or use it occasionally; amount of annual fee he was willing to pay; whether or not he had to drive 30 minutes or more to use the center; and finally, how he responded (in the first stage) to the basic question. Doing inference on the estimated coefficients in the regression showed not only that no coefficients were significantly different from zero, but also, that taken simultaneously, the estimated coefficients had no significance; that is, there were no significant relationships between these variables and the response-non-response variable. We therefore concluded that by ignoring the non-respondents, we had not systematically excluded an important subgroup of the population.

4.2 Technical Issues

4.2.1 Preparation of the Composite Reason List

One difficulty we encountered when we carried out this study was related to the preparation of the composite reason list. At the first stage of the study, the participants were asked to write down in their own words all the reasons for their answer to the basic question. Naturally, many of the participants gave the same reason but in different words. Sometimes, they wrote down similar words, but the words meant very different things. The task of combining the similar reasons, and differentiating distinct reasons, became very difficult, time consuming, and sometimes, quite subjective. Furthermore, there was heavy reliance on the judgments of the intermediary who was responsible for preparing the list. The subjectivity of such a task sometimes lead to other problems.

The feedback process was designed so that in addition to the composite list, each participant was reminded of the reasons he had given on the previous stage, as well as his own response on the previous stage. Since the reminder of earlier given reasons was provided, not in the participant's original words, but in an edited form (perhaps a paraphrase of the original reason), there is the possibility that the participant might not have recognized his own previously given reasons. Subsequently, he may not have checked the same reasons on the following stage, not because he did not think those reasons important any more, but because he thought them to be different reasons from his original ones.

For future studies we propose several solutions to these problems. One alternative is to select a panel of judges to prepare the composite list so that the resultant list can be more objective. Another alternative is to partially computerize the process of preparing the reason list. The computerization will not only permit the analyst to formulate a more objective list, but also, speed up the preparation process so that the method can be executed in a matter of one or more hours, with the participants or their surrogates sitting at computer terminals at various locations.

4.2.2 Stabilization

As stated in Section 1, the feedback process of the QCF method should continue until the participants' responses have stabilized. Stabilization is defined as the state when no participant has changed his response at the present stage from his response at the previous stage. In this feasibility study we did not attempt to carry the process out until it was reasonably stable; rather, we pre-fixed the number of stages at three,

for convenience. We did some related analysis, however, that bears on the stabilization issue.

The first analysis involved conducting two regressions. One used participant's first stage responses as the dependent variable, and their second stage responses as the independent variable. The other regression used their second stage responses as the dependent variable, and their third stage responses as the independent variable. The results of these analyses showed that 83% of the variation on the second stage response can be explained by the first stage response and 92% of the variation on the third stage can be explained by the second stage response. We concluded from these results that there was a tendency toward stabilization as the feedback process went along.

In the second analysis we did to study the stabilization question, we utilized an index developed in Press (1978a). Based on the definition of stabilization, he designed a stopping rule index defined as

$$Q_j = \frac{1}{N} \sum_{i=1}^N (Z_{ij} - Z_{i,j-1})^2, \quad j \geq 2,$$

where Z_{ij} denotes the response of subject i on stage j , and $i=1, \dots, N$. The point at which the feedback process should be stopped is when Q_j is sufficiently small.

We calculated index values for our sample at stages two and three. We found that for the combined group, $Q_2 = 95.9$ and $Q_3 = 62.5$. The index value for the third stage was 35% less than that for the second stage. Moreover, the third stage histograms of responses had not changed much from the second to the third stage. It is unlikely that much more would

have been gained by a fourth or fifth stage. In future studies the Q index can be used as a basis for termination of the feedback process. Computerization of the Q index can also accelerate the process.

5. CONCLUSIONS

Our empirical investigation of the QCF procedure has brought to light some interesting and important aspects of measuring group judgment; most important, it has demonstrated the feasibility of using the QCF method. We showed in this study that even when doing the study on a part-time basis, and with limited support, we were able, in a straightforward way, to carry out three stages of a QCF procedure. The response rates in our survey were, fortunately, very high. The study illustrated that the QCF method did provide an environment of group interaction, and the participants did show signs of actively being involved in the interaction process. Finally, the study demonstrated that the individual judgments derived from the group interactions were different from those derived from conventional survey methods. The major difference lies in the fact that the participants, after being exposed to the totality of reasons for their answers to the basic question by all the participants, tended to polarize into two camps, those favoring and those opposing the issue under study. The fact that the participants did not tend towards consensus supports our original argument that the so-called consensus reached at the end of other group judgment methods is oftentimes artificial, and may be very far away from the true sentiment of that particular group. The result has an important implication for policy makers who are often the ultimate users of such judgmental data. If polarized judgments result from carefully reasoned, interactive thought (such as judgments resulting from use of the QCF

method), the results probably signal problems that need to be handled with special care and consideration. By studying the relative heights of the modes, and their shapes and separations, policy makers who make use of the QCF method are likely to have available better quality, as well as quantity, information to assist them in assessing risks, forming judgments, and making informed decisions.

FOOTNOTES

1. This observation is supported by the fact that at an importance rating of 37.5 where the sample category probability is about .04, the sample standard deviation is readily found to be .018; for category probabilities of greater value, the sample standard deviation rises to only about .04.
2. It should be noted that changes across stages are noted only qualitatively since it is too difficult to carry out statistical tests for such changes. The reason is that because of the feedback, the responses are mutually correlated and tests involving such data are unknown at this time.
3. The chi-squared test of goodness of fit is not strictly applicable in this case. One basic assumption of the test is that the observations be independent. However, in our case the observations (responses) obtained in the third stage are correlated because of the feedback of reasons common to all panelists.
4. Transition Probability, $p(Z_n, Z_{n-1})$ is the probability an individual will change his response from Z_{n-1} at stage $n-1$ to Z_n at stage n . Let z_n be the response of a subject at stage n ($n=1,2,3$). Then the transition probability is simply the conditional probability $p(z_n, z_{n-1}) = P[Z_n = z_n | Z_{n-1} = z_{n-1}]$ of giving response z_n ($z_n=0, 12.5, \dots, 100$) at stage n , given that the participant gave response z_{n-1} at stage $n-1$. Here we assumed that the subject's response on the present stage depends only on his response at the previous stage. (Recall that, during data collection each subject was reminded of his previous stage response.) A sample estimate of $p(z_n, z_{n-1})$ is given by

$$\hat{p}(z_n, z_{n-1}) = \frac{f(z_n, z_{n-1})}{f(z_{n-1})}$$

where $f(z_n, z_{n-1})$ denotes the number of persons who gave response z_{n-1} at stage $n-1$ and response z_n at stage n , and $f(z_{n-1})$ denotes the total number of persons who gave response z_{n-1} at stage $n-1$.

5. In many situations it is likely that a large fraction of the reasons likely to be generated by the panel could be generated beforehand and then tabulated on a checkoff list for use by the panel. Of course the panel would still be free to generate other reasons. For most reasons, however, the procedure would become greatly simplified.

REFERENCES

- Ali, M. (1978), "An Empirical Study of Judgment Making In Groups Using Qualitative Controlled Feedback," a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Mathematics and The Institute of Applied Mathematics and Statistics, The University of British Columbia.
- Asch, S. E. (1962), "Effects of Group Pressure Upon the Modification and Distortion of Judgments," in Cartwright, D. and A. Zander (eds.), Group Dynamics: Research and Theory, New York: Peterson and Company.
- Collins, E. B. and H. Guetzkow (1964), A Social Psychology of Group Processes for Decisionmaking, New York: John Wiley & Sons, Inc.
- Dalkey, N. C. and Helmer, O. (1951), "The Use of Experts for the Estimation of Bombing Requirements--A Project Delphi Experiment," R-1283-PR, The Rand Corporation.
- Janis, I. L. (1972), Victims of Groupthink: A Psychological Study of Foreign Policy Decisions and Fiascoes, Boston: Houghton-Mifflin Co.
- Press, S. J. (1978a), "Qualitative Controlled Feedback for Forming Group Judgments and Making Decisions," Journal of the American Statistical Association, September, 1978.
- ____ (1978b), "Multivariate Group Judgments By Qualitative Controlled Feedback," Technical Report #39, Department of Statistics, University of California, Riverside, June, 1978. This paper was presented at the Fifth International Symposium on Multivariate Analysis, University of Pittsburgh, June 1978. It will also appear in the Symposium Proceedings entitled Multivariate Analysis V, edited by P. R. Krishnaiah.
- ____ (1979), "Bayesian Inference in Group Judgments and Decision Making Using Qualitative Controlled Feedback," Technical Report # 47, Department of Statistics, University of California, Riverside, January, 1979.
- Sackman, H. (1975), "Delphi Critique - Expert Opinion, Forecasting, and Group Processes," Lexington, Massachusetts: Lexington Books, D. C. Heath & Co.

TABLE 1
SUMMARY TABLE OF THE SIZES AND RESPONSE
RATES OF THE SAMPLES USED IN THIS STUDY

<u>Stage</u>	<u>Stratum</u>	<u>Number Solicited</u>	<u>Number Responded</u>	<u>Response Rate (percent)</u>
1	Faculty	96	76	79.17
	Staff	81	71	87.65
	Total	177	147	83.05
2	Faculty	76	58	76.32
	Staff	71	61	85.92
	Total	147	119	80.95
3	Faculty	58	54	93.10
	Staff	61	57	93.44
	Total	119	111	93.28
Control Group	Faculty	55	46	83.64
	Staff	55	43	78.18
	Total	110	89	80.91

TABLE 2

MEANS OF RESPONSE DISTRIBUTIONS AT THE
THREE STAGES AND THE CONTROL GROUP,
STRATIFIED BY STATUS

<u>Groups</u>	<u>First Stage</u>	<u>Second Stage</u>	<u>Third Stage</u>	<u>Control Group</u>
Faculty	50.93	48.38	46.53	60.33
Staff	63.38	59.61	57.40	61.63
Combined	57.32	54.15	52.14	60.96

TABLE 3

STANDARD DEVIATIONS OF RESPONSE DISTRIBUTIONS AT THE
THREE STAGES AND THE CONTROL GROUP,
STRATIFIED BY STATUS

<u>Groups</u>	<u>First Stage</u>	<u>Second Stage</u>	<u>Third Stage</u>	<u>Control Group</u>
Faculty	26.77	25.80	27.58	29.29
Staff	25.63	27.88	29.60	27.54
Combined	26.92	27.47	29.15	28.47

TABLE 4

FREQUENCY COUNTS (Upper Left of Each Cell), AND MLE's OF
TRANSITION PROBABILITIES (Lower Right of Each Cell)
FROM FIRST STAGE TO SECOND STAGE***

		THIRD STAGE											
		0	12.5	25	37.5	50	62.5	75	87.5	100	Total		
FIRST STAGE	0	6 1.0	**								6 1		
	12.5		7 .78	1 .11			1 .11				9 1		
	25		3 .30	6 .60	1 .10						10 1		
	37.5			1 .20	4 .80						5 1		
	50			1 .08	5 .38	5 .38	1 .03	1 .08			13 1		
	62.5			1 .07	3 .20	1 .07	10 .66				15 1		
	75			1 .03	1 .03	2 .06	6 .17	23 .65	2 .06		35 1		
	87.5							1 .06	15 .94		16 1		
	100									2 1.0	2 1		
Total		6	10	11	14	8	18	25	17	2	111		

*Maximum likelihood Estimates

**Blank cells designate zero frequency count and zero transition probability

***For Combined Faculty and Staff

TABLE 5
FREQUENCY COUNTS (Upper Left of Each Cell), AND MLE*s OF
TRANSITION PROBABILITIES (Lower Right of Each Cell)
FROM SECOND STAGE TO THIRD STAGE***

		THIRD STAGE											
		0	12.5	25	37.5	50	62.5	75	87.5	100	Total		
FIRST STAGE	0	6 1.0	**								6 1		
	12.5	1 .10	8 .80	1 .10							10 1		
	25		1 .09	9 .82	1 .09						11 1		
	37.5		2 .14	5 .36	6 .43	1 .07					14 1		
	50		1 .13	1 .13		5 .62	1 .12				8 1		
	62.5						15 .83	3 .17			18 1		
	75							23 .92	2 .09		25 1		
	87.5					1 .06		2 .12	14 .82		17 1		
										2 1.0	2 1		
Total		7	12	16	7	7	16	28	16	2	111		

*Maximum likelihood Estimates

**Blank cells designate zero frequency count and zero transition probability

***For Combined Faculty and Staff

TABLE 6
 AVERAGED* TRANSITION PROBABILITIES OF REASON
 FROM THE FIRST STAGE TO THE SECOND STAGE

<u>AT FIRST STAGE</u>	<u>AT SECOND STAGE</u>	
	<u>The Same Reason is Not Given</u>	<u>The Same Reason is Given</u>
A particular reason is not given	.79	.17
A particular reason is given	.21	.83

*Averaged over 17 reasons.

TABLE 7
 AVERAGED* TRANSITION PROBABILITIES OF REASON
 FROM THE SECOND STAGE TO THE THIRD STAGE

<u>AT SECOND STAGE</u>	<u>AT THIRD STAGE</u>	
	<u>The Same Reason is Not Given</u>	<u>The Same Reason is Given</u>
A particular reason is not given	.87	.17
A particular reason is given	.13	.83

*Averaged over 26 reasons.

FIGURE 1
EXPERIMENTAL VS. CONTROL GROUPS

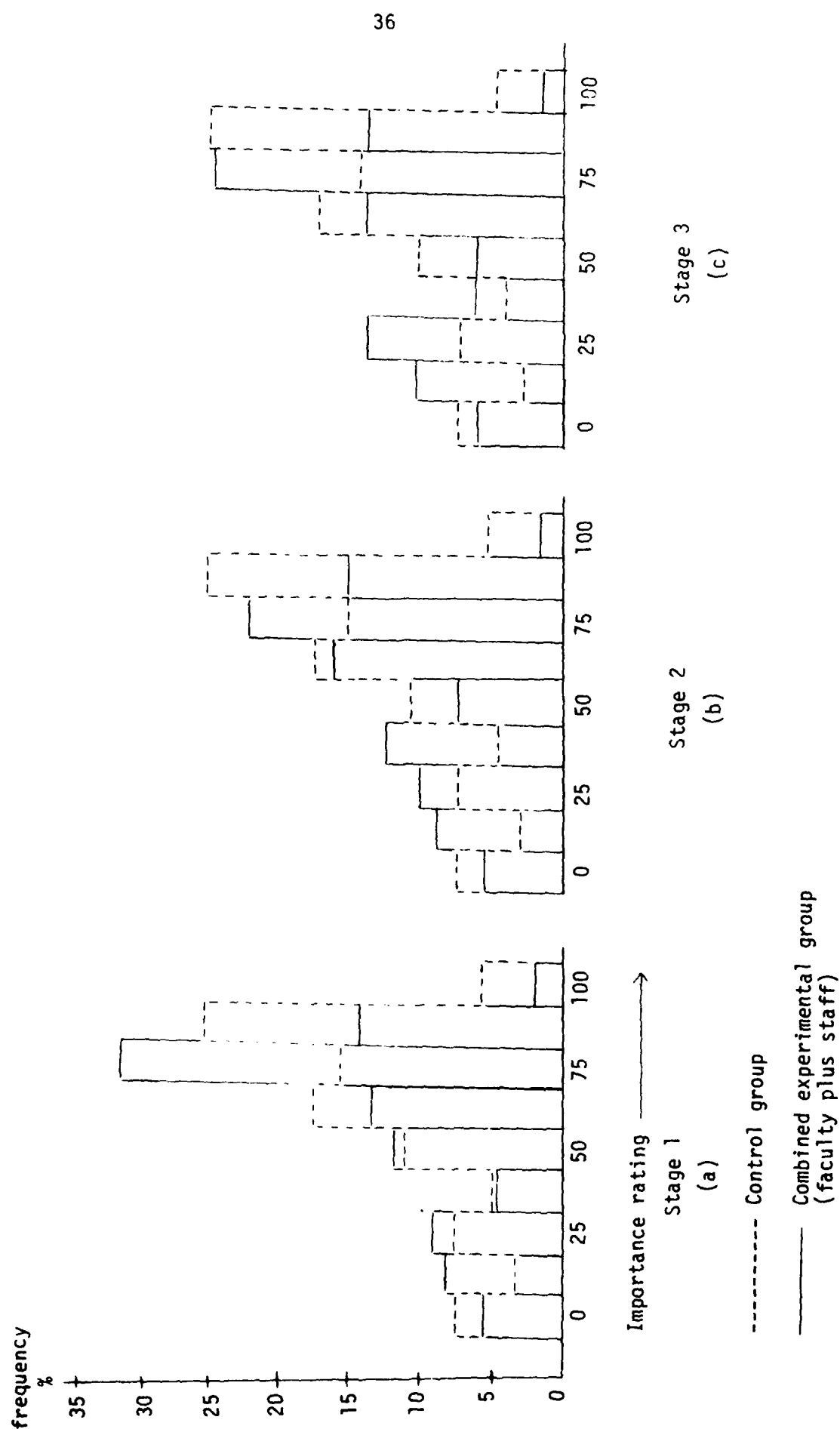
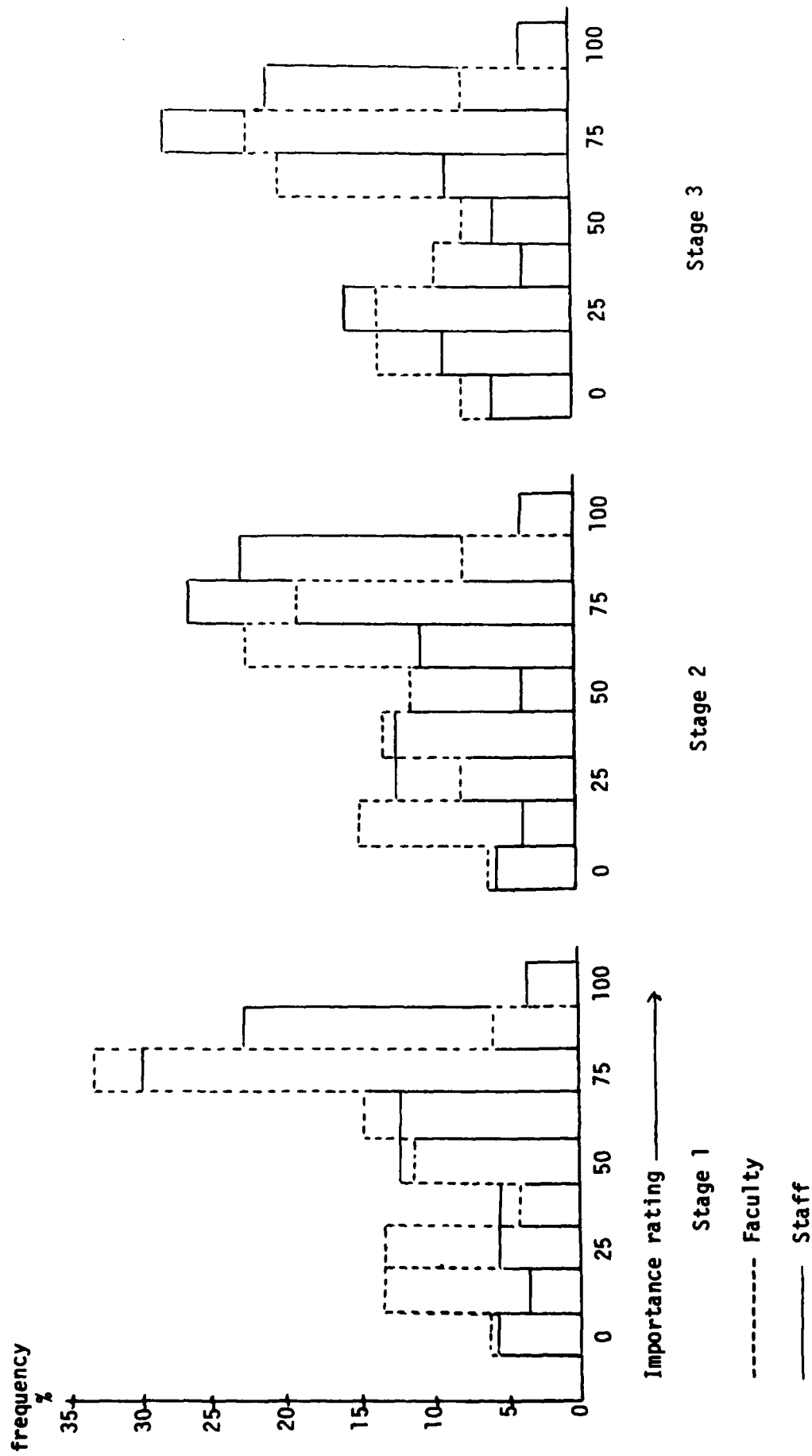


TABLE 8
CROSS CLASSIFICATION OF NUMBERS OF REASONS GIVEN, BY TYPE OF RESPONSE,
AND BY TYPE OF REASONS (FOR THE COMBINED FACULTY AND STAFF GROUP)

Response Category	Stage					
	First		Second		Third	
	Type of Reason					
	Positive	Negative	Positive	Negative	Positive	Negative
Proponents	105 (1.5)*	15 (.2)	231 (3.7)	50 (.8)	335 (5.4)	60 (1)
Neutral	1 (.07)	14 (1.1)	6 (.8)	20 (2.5)	9 (1.3)	19 (2.7)
Opponents	2 (.07)	47 (1.6)	13 (.3)	191 (4.7)	20 (.5)	239 (5.7)

*Numbers in parentheses are the average numbers of reasons of a given type given, per person, in a particular response category.

FIGURE 2
FACULTY VS. STAFF GROUPS



RAND/P-6333

AN EMPIRICAL STUDY OF A NEW METHOD FOR FORMING GROUP JUDGMENTS:
QUALITATIVE CONTROLLED FEEDBACK

S. J. Press

M. W. Ali

Chung-fang Elizabeth Yang

END

FILMED

6-85

DTIC