

Group Consensus Evaluation: A Procedure for Gathering Qualitative Data

Constance A. Mellon
*University of Tennessee
at Chattanooga*

Abstract. Group Consensus Evaluation, a course evaluation procedure based on qualitative data-gathering techniques, can provide an effective alternative to traditional course evaluation forms. Using this procedure, students come to agreement, first in small groups, then in the class as a whole, on the things they would like to keep the same in a course, the things they would like to see changed, and how that change might occur. This procedure provides the instructor with a single class view rather than a series of conflicting views while providing a forum for students to separate their individual problems with a course from those that are shared with their fellow students.

The young woman with a doctorate in educational technology was the only instructional developer at a small liberal arts college in the midwest. As such, it was her function not only to facilitate the design of courses, but to evaluate them as well. She had been trained at one of the major university development centers whose professional staff included two evaluators and she had studied carefully the evaluation processes they used. Therefore, for each course she designed, she gathered the typical kind of information using a written form composed of a Likert-type scale and a space for comments. It seemed, however, that the results were always the same. About half of the students thought the instruction was clear, interesting, important, and taught at just the right pace. The other half found it to be slightly confusing, somewhat uninteresting, generally unimportant, and taught a little too fast (or a little too slow). The open-ended comments were, perhaps, even less helpful. Some students thought that too

much material had been covered in the course while others found the material repetitive. Some students declared that the class was too structured while others complained of no structure at all. Some students wanted more lecture while others argued for increased student participation.

Written evaluation data of this kind are relatively easy to analyze; however, interpreting the results to faculty may create a problem. Let us suppose that the new instructional design for a sociology class included a role play exercise. The professor, anxious to find out how students reacted to this new educational experience, asks that evaluation data be collected. Presenting the data to the professor might create a scene similar to the following.

"Dr. Johnson," the developer might begin, "of the 32 students enrolled in your class last semester, 15 mentioned the role play exercise as the thing they liked best about the class."

"Excellent," the professor might reply. "But what about the other 17 students?"

"Well," the developer might explain, "eight of them really thought the role play was a waste of time, three said they would never enroll in another course where they were forced into a role play situation, and the other six didn't mention that component of the course at all."

"All right then," might be the response. "You're the instructional developer. What should I do about that role play component next semester?"

This dialogue is typical of the difficulties instructional developers encounter in the higher education setting. On the one hand, they are trained to believe strongly in the necessity for formative evaluation as an integral component of development. On the other hand, the data they are able to collect are often contradictory, confusing, and insufficient to provide any clear direction to the efforts to improve instruction.

Are there other choices? At the College of Saint Benedict in St. Joseph, Minnesota, an alternative procedure was developed. The procedure, Group Consensus Evaluation, was based on the work of D. Joseph Clark and colleagues at the Biology Learning Resources Center, University of Washington (Clark & Bekey, 1979; Redmond & Clark, 1982). Begun as a small effort in conjunction with several newly developed general education courses, it was met with enthusiasm by the faculty and soon expanded into a college-wide program.

were run for classes which ranged in size from five to ninety students.

Each small group is asked to appoint a recorder to report their findings back to the group as a whole. The evaluator then asks each group to take about five minutes to come to agreement on the things they liked best about the class, the things they liked least, and some specific suggestions for improving what they liked best.

At the end of the time period allotted for small group discussion, the evaluator asks for group reports. Spokespersons are asked to report their group's findings

The Theoretical Base of Group Consensus Evaluation

For purposes of expediency, most instructional evaluation in higher education uses a written format. The theoretical base of the written evaluation method is part of the quantitative tradition in research and evaluation in which numbers are used to predict outcomes. These written evaluations of instruction generally focus on one of two factors: course/component design or teacher effectiveness.

Written evaluation data are collected to assure or improve the quality of new instruction designs. This activity, called formative evaluation, is usually conducted by instructional developers using questionnaires or rating instruments which focus upon a course or a course component being designed and is considered an integral part of the instructional development process. However, as described earlier in this article, the data are frequently ambiguous and provide little clear direction for change.

In higher education, there has been a growing interest in student evaluation of teacher effectiveness. This type of evaluation, like the formative evaluation described above, uses written questionnaires or rating instruments; however, these instruments, while mentioning course content and methods, tend to focus on instructor performance. Two important questions have been raised concerning this type of evaluation: Are the factors being rated valid measures of good instruction? Does the feedback received from these instruments improve subsequent instruction? (Chandler, 1978; Levin, 1979; Rotem & Glassman, 1979.) Both types of evaluation share a common problem. They separate components of instruction from each other, from the environment in which they occur, and from the human interaction that comprises the educational setting.

Unlike the written evaluation method, the theoretical base for Group Consensus Evaluation is qualitative in nature, with the focus on viewing educational experiences from the perspective of those involved: student, teacher, administrator. There is an attempt at understanding why participants in an educational experience react as they do to that experience. This is accomplished by applying research procedures from the areas where the traditional focus has been upon in-depth study of people: the ethnographic techniques of anthropology and the qualitative methods of sociology (Bogdan & Biklen, 1982;

Group Consensus Evaluation is a procedure for obtaining information using interview and small group techniques.

What Is Group Consensus Evaluation?

Group Consensus Evaluation is a procedure for obtaining from students much the same sort of information that is solicited on the standard evaluation form. However, the information is obtained in the classroom using interview and small group techniques. The procedure follows a fairly simple and straightforward pattern of activities.

The evaluation is conducted at the end of a regularly scheduled class, taking between twenty to forty minutes to complete. After introducing the evaluator and explaining that he/she will gather information for course improvement, the classroom teacher leaves. The evaluator tells the students that the procedure they will use is designed to facilitate class agreement on the strengths and weaknesses of the course and how the course might be improved. He/She assures the students that only a summary of their discussion will be presented to the instructor and that the things said and who said them will be kept confidential.

The students are then asked to divide into small groups of no more than five or six. This physical division, often in classrooms obviously not designed for small group interaction, is actually one of the more difficult parts of the procedure. Nevertheless, at the College of Saint Benedict, effective evaluations

on each of the three questions; strengths, weaknesses, and suggestions for improvement. The groups who are not reporting are asked to listen carefully to see if they agree or disagree with the group that is reporting. When groups disagree on some issue, it is thoroughly discussed. Sometimes, when the issue has been clarified, the groups will then agree on a shared point of view. If agreement is impossible, the evaluator might try a hand count to see what percentage of the group holds each point of view. This procedure continues until each group has reported.

The evaluator has two main functions during this process. First, he/she must listen carefully and record the findings of each group. Second, where general group discussion is needed, he/she should facilitate clarification and communication by asking necessary questions and providing succinct summaries. He/She also acts as timekeeper for the process, being certain that all groups have a chance to report fully on each of the questions within the time allotted for the evaluation.

It can be seen that both written evaluation and Group Consensus Evaluation solicit the same sort of information about a course. In practice, however, while the factual elements obtained may be similar, the outcome of each method is quite different. This is primarily due to the fact that the theoretical base of Group Consensus Evaluation is different from that of the written evaluation.

Hamilton, MacDonald, King, Jenkins & Parlett, 1977; Spradley & McCurdy, 1972). These procedures involve gathering and analyzing descriptive data in an attempt to see an event or a phenomenon as the participants see it. It is a holistic approach which tries to describe all the bits and pieces, the variables, that comprise an individual's world rather than trying to control for them as quantitative methods do.

Qualitative research differs radically from quantitative research in the data it seeks and in the way that data is judged. The results of qualitative studies cannot be adequately considered within the accepted definition of research "reliability." Analysis of this type of data, as Glaser and Strauss (1967) explain, is "dependent on the skills and sensitivities of the analyst" and "is not designed (as methods of quantitative analysis are) to guarantee that two analysts working independently with the same data will achieve the same results." (page 103). In fact, the question of whether two evaluators working with the same class achieve the same results is of less relevance to the purpose of Group Consensus Evaluation than is the question of whether the results truly reflect the views of the students comprising the class. Since the evaluation data are constantly reviewed with the class during the administration of the procedure, the results tend to be valid for that class. If major, time-consuming revisions appear to be necessary, a professor might decide to have several classes evaluated to be assured that the suggestions for revision are not unique to the needs of only one class.

One useful approach to understanding qualitative methodology is from the perspective of symbolic interactionism. This perspective, based upon the philosophy of George Herbert Mead (1934) and Herbert Blumer (1969), differs from conventional psychological and sociological frameworks in that the human being is viewed as an *acting* rather than a *responding* organism. The assumption is that people act on the basis of the meaning things have for them and that in order to understand the actions of people, it is necessary to interpret objects as they do. "Objects," from this perspective, include not only the person doing the acting, but other people and their actions, the environment, and the physical objects within the environment.

Although qualitative research methods are limited only by the imagination of the researcher, two basic

techniques are consistently associated with this area: participant observation and unstructured interviewing. Group Consensus Evaluation is based upon the latter technique, unstructured interviewing.

The Unstructured Group Interview

The unstructured interview has been described in a variety of books dealing with qualitative methods (Banaka, 1971; Richardson, Dohrenwend & Klein, 1965; Spradley, 1979). Basically, it is a conversation between two people where one can ask questions and the other respond in a comfortable and non-threatening way. The purpose of the unstructured interview is to understand another's point of view. The questions should be aimed at making the person being interviewed comfortable so that he/she will be willing to communicate. Thus, questions should not challenge, confront, or argue with the point of view being expressed. The aim is to understand another's perspective, not to change it.

In the unstructured interview, the person being interviewed should do most of the talking. The interviewer should encourage the respondent to continue talking by nods, smiles, and such verbal encouragement as "yes" and "go on." The interviewer should not interrupt the respondent or cut off the flow of conversation. If questions occur as the respondent talks, the interviewer should remember them or jot them down to ask later.

methodology with the addition of one further element. Interviews are conducted in a group setting. The evaluator who conducts the Group Consensus Evaluation should be familiar and comfortable with group procedures. During the early part of the process, groups must be allowed to function freely in order to arrive at agreement on the points in question. However, the evaluator needs to be aware of groups who require some prodding on time or task and the point at which to end the small group discussion. When the small groups report to the group as a whole, the evaluator should summarize the points as they are noted, then constantly check to be sure that his/her perceptions agree with those of the students who are reporting and the students who are listening. Clarifying questions should be asked as necessary and large group discussion of the points being made should be encouraged.

Applying the procedures of unstructured interviewing to this group situation, the evaluator should remain neutral, neither agreeing nor disagreeing with the opinions being expressed. His/her responsibilities during this data-gathering stage are facilitation and inquiry. Through the use of these procedures, information on the students' perceptions of the class and the basis for these perceptions can be collected.

Clarifying Data Through Group Consensus Evaluation

One of the particular values of this

Students expressed their preference for Group Consensus Evaluation as opposed to the written evaluation method.

Merton, Fiske and Kendall (1956) described a modification of the unstructured interview which they called the focused interview. While still aiming at subjective experience, the respondents are asked about a particular concrete situation in which they were all involved. The intent is to clarify the situation by recording descriptions from a variety of participants.

Group Consensus Evaluation incorporates both the unstructured interview and the focused interview into its

evaluation procedure is the opportunity to clarify perceptions as they are noted. For instance, during a Group Consensus Evaluation with an upper division chemistry class, there was an obvious split of opinion on the value of the lectures the instructor had given. One small group reported that they liked the lectures best because they stressed principles rather than repeating specific material to be memorized. Other students strongly disagreed, claiming that they found much of the lecture material incom-

prehensible. By questioning and discussion, it became clear that the problem was with the entering knowledge of the students rather than the lecture style of the instructor. Those students who liked the lectures had taken their basic chemistry classes from the same instructor. Those who disliked the lectures had taken the introduction from other instructors. The students concluded that there was a real need for consistency in the chemistry sequence and that departmental revision would be necessary to accomplish this. For a short term solution, they suggested that study guides and skill sheets used by the instructor in his basic chemistry class be made available to the students who had taken classes from the two other chemistry instructors and that a weekly review session be held for those who needed it.

In another instance, an introductory math class was being evaluated. The students reported two opposing perceptions of the instructor: one group found her to be encouraging, accepting, and helpful; the other group reported inflexibility and discomfort in raising questions. As these two points of view were discussed, one student noted the fact that the positive perceptions were from the men in the class and the negative ones from the women. Several women then were able to state their feelings that the instructor showed preferential treatment to male students. Once this view had been expressed, many of the men in the class supported it and were able to describe incidents to illustrate it.

evaluation results are reported to the instructor.

Reporting Group Consensus Evaluation Data

Group Consensus Evaluation data is reported orally to the faculty member, preferably within the familiar surroundings of his/her own office, with the positive data reported first. This builds good rapport with the faculty member, even those with whom the evaluator has never worked before. It also helps to prepare an environment of trust and acceptance within which the less favorable aspects of the evaluation data can be discussed in a non-threatening manner. When reporting those things the students liked least about the course, the students' suggestions for improvement should be linked with each point. Where class opinion is divided, this should be mentioned.

The data feedback sessions with faculty were found to be a natural and effective bridge to instructional development activities. In this situation, the necessity for the developer to provide both evaluation and development services was appreciated by both developer and faculty member. Faculty were willing, often anxious, to discuss the points raised in the evaluation and to explain their goals and rationale for the segments of the course that were and were not seen as effective by the students. The willingness of the developer to provide the expertise and assistance necessary to make the suggested improvements in the

uated as well as to improve the course in the future. Evaluations conducted at mid-semester provide enough succinct information to allow for immediate adjustments to the class being conducted. Thus students no longer have to be asked to contribute data for course improvements they will never see. Improvements are immediate and are tailored to the individual needs of the class.

Second, both students and faculty like this method of evaluation. Students expressed their preference for Group Consensus Evaluation as opposed to the written evaluation method. They explained that, while they were asked to fill out innumerable written evaluation forms, they never saw any results. With this procedure, they felt that someone was finally listening to their ideas and acting upon them.

Faculty were also enthusiastic about Group Consensus Evaluation. The data were clearly defined and presented a single class view rather than a compilation of many opposing views. Suggestions for improvement were reasonable, possible, and could frequently be quickly implemented in the class being taught. In addition, the fact that the faculty member had requested the evaluation and then acted upon it improved the communication between the instructor and the students.

At the College of Saint Benedict, faculty requests for Group Consensus Evaluation increased from five during the first semester it was offered to thirty the following semester. Faculty who had requested evaluation for one class they were teaching often followed the first evaluation with requests for their other classes to be evaluated. In addition, many of the faculty requested written summaries of the feedback sessions to include in their documentation for promotion and tenure.

Third, Group Consensus Evaluation allows a clear distinction to be made between the problems of individual students and problems with the course. By allowing time for class discussion of each issue, students can see which problems experienced in the class are shared by the group and which problems are the student's own. While the evaluator using this procedure never challenges the opinions students present, other students discussing how their views differ from the one being expressed may do so. For instance, a small group of students in a religion class under evaluation had no positive comments on the course at

Unlike the written evaluation method, the theoretical base for Group Consensus Evaluation is qualitative in nature.

These two examples were selected to demonstrate the type of data which can be gathered using the unstructured interview procedures of Group Consensus Evaluation. As can be seen, discrepancies in the data can be followed up immediately, and in many instances, the reasons for the discrepancies become clear. This allows for a fuller understanding of the problems with a specific course and how they might be handled, thus providing a broader base of supporting information when the

course was welcomed by faculty under these circumstances where it might have been ignored or resented under other circumstances.

The Advantages of Group Consensus Evaluation

In addition to the points discussed above, there are several other advantages of Group Consensus Evaluation which should be considered. First, due to the nature of the procedure, it can be used to improve the class being eval-

all. They complained that the work was too hard, the assignments too long, and the tests unfair. There was immediate disagreement by the majority of the students who defended the instructor and her teaching methods. Several students pointed out quite strongly to the dissenting group that the assignments and work load were more than fair compared to other college courses and that the problem was that this group of students was just not willing to work.

vided. The Group Consensus evaluation procedure is useful in any setting where course evaluation is used to guide instructional development activities.

Group Consensus Evaluation is a cost-effective way to gather data.

The dissenting students finally admitted that their stand had been somewhat biased and that the workload and testing procedure was indeed fair in the college setting.

Finally, Group Consensus Evaluation is a cost-effective way to gather data. It requires less than an hour of class time, yet yields results which are, in essence, already analyzed. And the analysis is conducted by experts on the meaning of the data: the students whose opinions the data represent. The data reported are clear, present a single view, and can be acted upon without a great deal of debate about meaning. In addition, Group Consensus Evaluation prepares the way for any instructional improvement services available to the faculty at that particular institution.

In Summary

Group Consensus Evaluation, a course evaluation procedure based on qualitative data-gathering techniques, can provide an effective alternative to the written evaluation. By involving the students in a discussion aimed at class consensus, it provides a forum for separating the problems of individual students from the problems with the course and for clarifying ambiguous pieces of data. A single class view is presented for the instructor's consideration rather than a series of conflicting views. Moreover, through the one-on-one feedback sessions, a natural setting for involving faculty members in instructional development activities is pro-

References

- Banaka, W. H. (1971). *Training in depth interviewing*. N.Y.: Harper & Row.
- Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Englewood Cliffs, N.J.: Prentice Hall.
- Bogdan, R. C. & Biklen, S. K. (1982). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn & Bacon.
- Chandler, T. A. (1978). The questionable status of student evaluations of teaching. *Teaching of Psychology, 5*, 150-152.
- Clark, D. J., & Bekey, J. (1979). Use of small groups in instructional evaluation. *POD Quarterly, 1*, 87-95.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Hamilton, D., MacDonald, B., King, C., Jenkins, D., & Parlett, M. (1977). *Beyond the numbers game: A reader in educational evaluation*. Berkeley, CA: McCutchan.
- Levin, B. (1979). Teacher evaluation—A Review of research. *Educational Leadership, 37*, 240-245.
- Mead, G. H. (1934). *Mind, self and society*. Chicago: University of Chicago Press, 1934
- Merton, R., Fiske, M., & Kendall, P. (1956). *The focused interview: A manual of problems and procedures*. Glencoe, IL: The Free Press.
- Redmond, M., & Clark, D. J. (1982). Small group instructional diagnosis: A practical approach to improving teaching. *AAHE Bulletin, 1982*, 35.
- Richardson, S. A., Dohrenwend, B. S., & Klein, D. (1965). *Interviewing: It's forms and functions*. N.Y.: Basic Books.
- Rotem, A., & Glassman, N. S. (1979). On the effectiveness of students' evaluative feedback to university professors. *Review of Educational Research, 49*, 497-511.
- Spradley, J. P. (1979). *The ethnographic interview*. New York: Holt, Rinehart and Winston.
- Spradley, J. P., & McCurdy, D. (1972). *The cultural experience: Ethnography in complex society*. Chicago: Science Research Associates.