

Using Sequential Two-Part Focus Groups As A Supplemental Instrument For Student Course Evaluations

Robert E. Pritchard, Rowan University, USA
Gregory C. Potter, Rowan University, USA
Michael S. Saccucci, Western Connecticut State University, USA

ABSTRACT

This paper describes the use of a sequential two-part focus group program designed to be a supplemental instrument for eliciting student suggestions to improve the teaching/learning process. The sequential two-part focus group program provides students with two opportunities to provide feedback and have that feedback both noted and mirrored back to them. Many of the student comments and suggestions elicited during the focus group sessions were positive and constructive. Moreover, use of the sequential two-part focus group program resulted in statistically significant improvement in student evaluations in all but one of the seven major sections of the Educational Testing Service's SIR II, the primary instrument utilized to obtain the student course evaluations. The "Overall Evaluation" of the course, wherein students are asked to, "Rate the quality of instruction in this course as it contributed to your learning," improved most significantly.

INTRODUCTION

*A*lthough much of the research on student evaluations of teaching (SETs) is critical of their validity and usefulness, a significant number of studies suggest that the evaluation of teaching is nonetheless appropriate and can be made more relevant and useful when multiple measuring instruments are utilized. One such measuring instrument is a focus group.

This paper describes the use of a sequential two-part focus program (Program). The Program was designed to afford opportunities for business students to provide course feedback and have that feedback both noted and mirrored back to them twice during the semester. The Program's purpose was twofold:

1. To elicit student feedback to supplement the Educational Testing Service's Student Instructional Report II (SIR II) – specifically to obtain information pertaining to changes the instructor could make to improve the teaching/learning process.
2. To measure any changes in the students' SIR II responses that may have resulted from the use of the Program.

The Program's effectiveness was measured in two ways. First, the usefulness of the student inputs was assessed. Second, after using the Program for two semesters, the instructor's SIR II evaluations for those two semesters were compared with the evaluations obtained during the two semesters just prior to the start of the Program (baseline semesters). The Program's effectiveness based on these two measures is discussed below in the section, "Research Design and Results."

LITERATURE REVIEW

Collecting feedback from students on individual teachers is a widespread practice in higher education, although one checkered with some criticism of the validity and usefulness of the instruments. Richardson's (2005) review of the literature on SETs suggests that student feedback is a major source of evidence for assessing teaching quality and can be used in trying to improve that quality, in addition to providing information of value to future students. Richardson notes that although many teachers and students believe student feedback is important, many teachers and institutions do not take student feedback sufficiently seriously (p.410). Simpson and Siguaw (2000) highlight research supporting the claim that SETs are "reasonably valid multidimensional measures," although 50 to 80 percent of the total variance of evaluations can be linked to personality variables. (p. 200.)

Mace (1997) acknowledges the long history of SETs and their widespread use today. However, he remarks that the validity of SETs may be questionable and raises the issue of how teaching effectiveness is actually defined. Such a definition depends on how "effective teaching" is operationalized (for example, an outcome of learning measured by student knowledge in a pre- and post-test situation; or, operationalized as the production of a "satisfied customer"). Also noting research critical of SETs, Read, Rama, and Raghunandan (2001) comment that certain "extraneous factors" such as gender, age, race, and national origin may influence ratings.

Mace's research underscores a concern that student evaluations, in the final analysis, are a matter of values (for example, does the instructor value student opinion?). Mace also remarks that evaluations give ratings across a set of general teaching qualities that may lack specificity or clarity (p. 263). More fundamentally, Timpson and Desley (1997) note that a key assumption of student ratings, namely that there is a direct relationship between teaching and learning, is actually a complex matter, including, for example, considerations of student learning style and intellectual commitment. Relating to the matter of complexity, Sheehan and DuPrey (1999), speaking to the multidimensionality of teaching, note a lack of agreement on the factors involved in effective teaching, as revealed in a survey of a large number of close-ended rating scales. Similarly, Salmon, Smith, Lee, and Miller (2005) point to the "mixed evidence" that a positive relationship between student ratings and student learning can be found.

In addition to this research challenging the validity and usefulness of SETs, other observers note that the use of formal instruments, often commercially produced, may encourage, as Richardson remarks, "uncritical acceptance of traditional conceptions of teaching...." (p. 391.) He adds that such instruments also may neglect considerations of critical thinking and self-expression. Ngware and Ndirangu (2005) add that student opinion surveys have been found deficient in critical areas relating to teaching effectiveness.

Some research suggests, however, that SETs may have an appropriate place in assessment. Ahmadi and Cotton (1998) highlight the importance of matching specific evaluative components with the course being evaluated, the classroom environment, and the instructor to ensure appropriate, useful, and accurate evaluation. Simpson and Siguaw (2000), while acknowledging there are no definitive conclusions about the instruments used in SETs, suggest that these measures nonetheless are reasonably valid. These researchers note that contextual variables such as grading leniency, class size, and workload do have an influence, along with such student factors as motivation and grade expectation.

From this perspective, the use of focus groups is one approach toward a diversification of evaluative techniques. Lunt and Livingstone (1996) trace the history of focus group research in the social sciences, noting a resurgence of interest relating to the use of qualitative approaches. These researchers note the evidence for focus groups as a method to generate discussion and also probe the negotiated meanings of that discussion. This leads, according to Lunt and Livingstone, toward an understanding of the "dilemmatic" nature of everyday arguments. (p.90.)

Calderon, Baker, and Wolf (2000) report on the use focus groups for conducting qualitative research in health care. They note that focus groups are useful for gathering perceptions, beliefs, and values, and are particularly suited for a context of cultural diversity. Meltzer, Maughan, and Fry (1995) used focus groups in a survey of undergraduate library usage as a way to address broad customer needs. They point to the use of focus groups as a way to encourage free and open sharing of ideas. Similarly, Hamilton, et al (2002) report that focus

groups used to augment student feedback not provided by a commercial SET provide a rapid means to implement student-suggested changes to the course and, where feasible, the building of trust.

Somewhat pragmatically, Mohr (2002) comments on the need for beginning teachers to plan, present, and assess productive assignments. She stresses the need to seek input from students about the “value and sensibility of assignments and then integrate their suggestions with one’s instructional expertise to refine the assignments.” (p. 470.) Parallel with this observation, Obenchain, Abernathy, and Wiest (2001) caution that what students perceive as effective teaching may not correspond with what the institution defines as effective teaching. Accordingly, seeking a student voice in evaluation may be critical for seeking what Cranton and Carusetta (2004) call “authentic teaching” ... teaching that leads to a critical approach toward the academic “community collective” and allows for a differentiation of one’s thoughts and values from those of the community. (p. 7.) Significantly, these researchers employed focus groups, in a three-year study among 22 educators, basing their study on grounded theory research and reliance on the investigator as the primary instrument of data collection.

Relating to Cranton and Carusetta’s “community collective,” Timpson and Desley (1997) remark that student feedback that is sensitive to the social diversity of the student population may contribute to a more “nurturing environment” and to the development of more credible assessment instruments and mechanisms. Focus groups, which by their nature engage a social dialectic, may help to sensitize the teaching evaluation process to consideration of what Salmon et al (2005) identify as variables lying outside a professor’s control.

Finally, Tiberius (2001) suggests the usefulness, ease of administration, and superiority of focus groups over written methods for student feedback. He mentions the helpfulness of focus groups to reconcile contradictory feedback received from student ratings, as well as for generating evaluative items. Noting that interpretation of focus group results can be time-consuming, he nevertheless provides seven conditions where the use of these groups is of high value. One of those conditions recognizes the possibility that focus groups can provide information to assist a teacher in enhancing what is currently effective, not just a corrective mechanism for “poor teaching.”

METHODOLOGY

The Program was conducted by one instructor teaching multiple sections of a required junior-year business course. The same text book and teaching methods were employed throughout. The Program consisted of two parts. The first part consisted of a focus group conducted in each class section about 10 weeks into the course. During this approximate 15-minute session, students were told that they are viewed as being important and, consequently, their ideas for improving the teaching/learning process in the course are also important. The students from each section were told that their suggestions would be noted and reviewed in their entirety with students in the other sections at the end of the semester. The goal of the end-of-semester review was to obtain additional student feedback and consensus. Finally, the students were told that their suggestions probably could not be implemented until the following year.

Based on this short introduction, student feedback was elicited about various aspects of the course such as the tests, scheduling of classes, etc. Student feedback was noted in a nonjudgmental way and then mirrored back to the students to 1) assure the feedback was recorded accurately and 2) elicit additional feedback. The instructor’s goal was to make sure the students knew that they were being heard and that the instructor valued their feedback as demonstrated by his taking the time to note the feedback and mirror it back to assure its accuracy. At the end of the focus group session, students were thanked for their inputs.

The second part of the Program consisted of a focus group conducted in each class section close to the end of the semester, just prior to the administration of the SIR II evaluations. Students were reminded of the earlier focus groups and then the combined list of the students’ prior suggestions was presented to the students for discussion. Each student comment/suggestion was clearly stated and presented for discussion. The instructor tried to reach a consensus among the students as to the usefulness of each, sometimes asking students to vote on them. Once a consensus was reached or a vote taken, the results were recorded and fed back to the students. This process lasted about 20 minutes in each class section. The students were thanked for their participation and then the SIR II forms were administered.

RESEARCH DESIGN AND RESULTS

The Program’s effectiveness as measured by the usefulness of the students’ suggestions was positive but mixed. A number of students suggested ways to make the course easier. For example, covering the material over two semesters rather than in one semester, skipping certain topics, as well as providing students with sample final exams and lists of items that would be included on multiple choice questions. The latter would provide the students with a better idea of what to study (and what not to study), thereby making the course easier. These suggestions were of limited value.

Other student suggestions were more helpful in terms of future planning for the course. These included using real-life company information (rather than hypothetical company information) in examples, including more in-class examples of some specific types of problems that were particularly difficult for students, including more real-world examples, and having more frequent tests (five instead of four). These and other student suggestions were useful for future course planning.

Beyond the student suggestions, the focus groups reinforced the notion discussed in the Literature Review – that students want to be heard. Students whose participation during many classes had been limited spoke up enthusiastically and wanted to talk about their ideas; they very clearly wanted to be heard. The process of recording student suggestions, repeating them back to students to assure accuracy, and repeating them again during the second focus group session let students know that they had been heard.

The second measure of the Program’s effectiveness was the improvement in the instructor’s SIR II evaluations in the two semesters following implementation of the program versus the two preceding semesters. The Program was initiated spring 2003 and continued fall 2003. The analysis presented below includes SIR II data from the two preceding semesters (spring and fall 2002, $n_{2002} = 79$). The data from these two semesters represents the baseline used for comparison. The SIR II data from the spring and fall 2003 semesters ($n_{2003} = 84$) is compared with the baseline data in Table 1.

Figure 1: Comparison of Mean Student Evaluation Scores on Seven Sections of SRI II Form

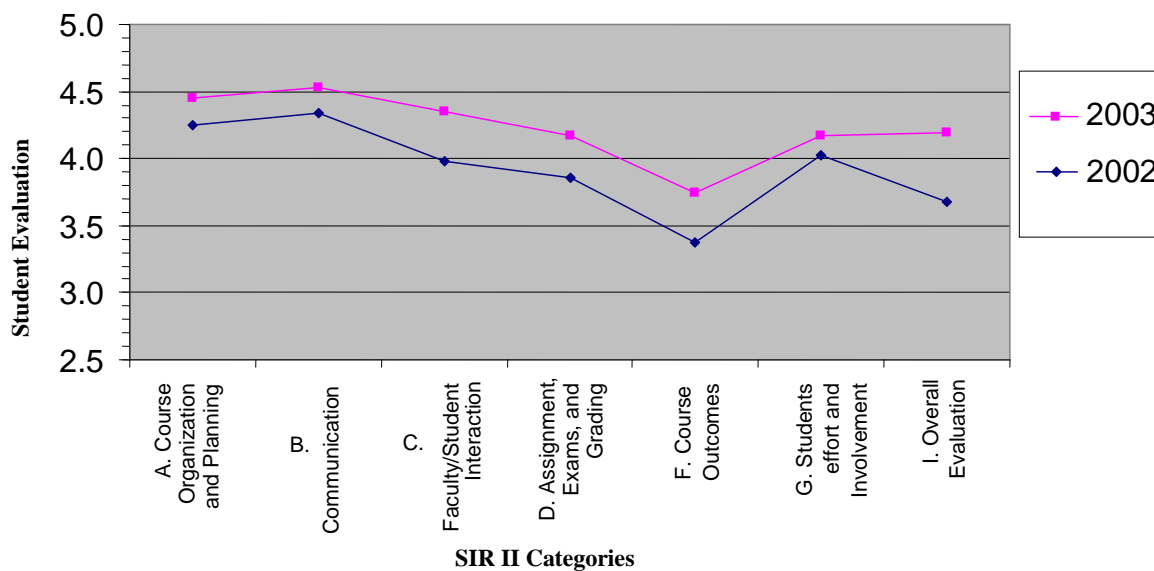


Table 1: SIR II Data

Description	2002 Average	2003 Average	Difference
A. Course Organization and Planning			
1. The instructor's explanation of course requirements	4.25	4.36	0.11
2. The instructor's preparation for each class period	4.42	4.57	0.15
3. The instructor's command of the subject matter	4.41	4.62	0.21
4. The instructor's use of class time	4.28	4.52	0.24
5. The instructor's way of summarizing or emphasizing important points in class	3.91	4.19	0.28
Overall	4.25	4.45	0.20
B. Communication			
6. The instructor's ability to make clear and understandable presentations	4.15	4.40	0.25
7. The instructor's command of spoken English	4.73	4.89	0.16
8. The instructor's use of examples or illustrations to clarify course material	4.18	4.40	0.22
9. The instructor's use of challenging questions or problems	4.16	4.39	0.23
10. The instructor's enthusiasm for the course material	4.47	4.59	0.12
Overall	4.34	4.53	0.19
C. Faculty/Student Interaction			
11. The instructor's helpfulness and responsiveness to students	4.00	4.34	0.34
12. The instructor's respect for students	4.04	4.43	0.39
13. The instructor's concern for student progress	3.71	4.09	0.38
14. The availability for extra help for this class	4.09	4.22	0.13
15. The instructor's willingness to listen to student questions and opinions	4.08	4.69	0.61
Overall	3.98	4.35	0.37
D. Assignment, Exams, and Grading			
16. The information given to the students about how they would be graded	4.27	4.44	0.17
17. The clarity of exam questions	3.63	4.02	0.39
18. The exam's coverage of important aspects of the course	4.00	4.21	0.21
19. The instructor's comments on assignments and exams	3.41	4.15	0.74
20. The overall quality of the textbook	3.95	4.01	0.06
21. The helpfulness of assignments in understanding course material	3.91	4.18	0.27
Overall	3.86	4.17	0.31
F. Course Outcomes			
29. My learning increased in this course	3.54	4.01	0.47
30. I made progress toward increasing course objectives	3.27	3.75	0.48
31. My interest in the subject has increased	3.25	3.60	0.35
32. This course helped me to think independently about the subject matter	3.46	3.77	0.31
33. This course actively involved me in what I was learning	3.38	3.57	0.19
Overall	3.38	3.74	0.36
G. Students effort and Involvement			
34. I studied and put effort into this course	4.06	4.30	0.24
35. I was prepared for each class	3.56	3.68	0.12
36. I was challenged by this course	4.44	4.52	0.08
Overall	4.02	4.17	0.15
I. Overall Evaluation			
40. Rate the quality of this course as it contributed to your learning	3.68	4.19	0.51
Number of Students	79	84	

The weighted average mean student rating for each of the SIR II questions as well as the seven major SIR sections and the differences are shown in Table 1. The weightings were based on the number of students in each of the two 2002 sections and the two 2003 sections. Higher evaluation ratings were obtained for all of the questions shown on Table 1 as well as for all of the seven major sections of the SIR II. Figure 1 displays the average mean student rating by SIR II category.

The data for the two years 2002 (baseline) and 2003 (using focus groups) were compared using 95% confidence intervals. The confidence intervals were calculated using a standard error obtained from the individual semester scores for 2002. The confidence intervals for the differences in the seven sections of the SIR II are shown in Table 2. All but one of the confidence intervals for the seven major SIR sections shows a statistically significant improvement from 2002 to 2003. The single exception was student effort and involvement.

Table 2: 95% Confidence Intervals for Differences in Seven SIR II Evaluation Sections

Description	Lower 95% Confidence Limit	Difference	Upper 95% Confidence Limit
A. Course Organization and Planning	0.04	0.20	0.34
B. Communication	0.04	0.19	0.34
C. Faculty/Student Interaction	0.22	0.37	0.52
D. Assignment, Exams, and Grading	0.17	0.31	0.45
F. Course Outcomes	0.21	0.36	0.51
G. Students effort and Involvement	-0.04	0.15	0.34
I. Overall Evaluation	0.17	0.51	0.85

CONCLUSION

Many of the students’ suggestions and comments obtained during the focus group sessions were useful for future course planning. Clearly the focus groups are valuable supplemental instruments for obtaining student course feedback. They provided student feedback that is not possible to obtain using the Educational Testing Service’s Student Instructional Report II. Furthermore, using the supplemental two-part focus group sessions allowed the instructor to obtain a consensus of student suggestions obtained earlier in the semester. This increased the value of the student feedback by identifying those particular suggestions that were most important to the students.

As measured by the positive changes in student evaluations, the Program had a salutary effect on the teaching-learning experience. Particularly noteworthy is the response to Question 40, “Overall Evaluation,” which increased from 3.68 for the baseline year 2002, to 4.19 for 2003, the year the focus groups were used. The national average scores for “Overall Evaluation” reported by ETS for both 2002 and 2003 was 3.96 while the average score for the college of business (COB) where this Program was used was 3.88. Thus, the instructor’s “Overall Evaluation” in the baseline year (2002) was lower than both the national average and the COB average. By contrast, the “Overall Evaluation” received during 2003 when the Program was used was considerably higher than both the national and COB averages.

It is also interesting to note that the greatest increase in the student evaluation scores were in their responses to the questions in Section C, “Faculty/Student Interaction,” (instructor’s helpfulness and responsiveness to students, respect for students, concern for student progress, availability for extra help, and willingness to listen to students) and Section F, “Course Outcomes,” (student learning increased during course, made progress toward achieving course goals, interest in subject increased, course helped me to think independently, course actively involved me in what I was learning). The significant improvement in student course ratings appears to be largely the result of more positive feeling on the part of the students about the course and the instructor. Since, as noted above, the instructor used the same text, teaching methods, etc., during both 2002 and 2003, the changes in the students’ feeling likely resulted from the use of the Program.

Consistent with research suggesting the usefulness of SETs as part of a multifocus approach toward faculty teaching assessment, this study suggests that focus groups evidently contribute to higher levels of student satisfaction when viewed in the context of supplementing a standardized commercial evaluative instrument (the SIR II in this instance). Research on SETs, in general, clearly shows that the routine collection of standardized commercial evaluative instruments contributes little toward helping faculty improve their individual teaching (Richardson 2005). This study supports the research (Richardson 2005) indicating that student feedback (such as

that obtained through the use of focus groups) can, however, be helpful to individual teachers for the improvement of their individual teaching.

NOTES

¹ SIR II questions relating to “Supplementary Instructional Methods” (Section E number 22 through 28) were not included in the study since they were not applicable. In addition, “Course difficulty, Work Load, and Pace” (Section H questions 37 through 39) were not included since mean scores not are provided by ETS for these items.

REFERENCES

1. Ahmadi, R. & Cotton, S. (1998). Assessing students' ratings of faculty. *Assessment Update*, 10(5), 5-7.
2. Calderon, J., Baker, R., & Wolf, K. (2000). Focus groups: a qualitative method complementing quantitative research for studying culturally diverse groups. *Education for Health*, 13(1), 91-95.
3. Cranton, P. & Carusetta, E. (2004). Perspectives on authenticity in teaching. *Adult Education Quarterly*, 55(1), 5-22.
4. Hamilton, D., Pritchard, R., Welsh, C., Potter, G. & Saccucci, M. (2002). Effects of using in-class focus groups on student course evaluations. *Journal of Education for Business*, 77(16), 329-333.
5. Livingston, S. (1996). Rethinking the focus group in media and communications research. *Journal of Communication*, 46(2), 79-99.
6. Mace, J. (1997). Valuing the individual student: Using single-subject design as a tool for evaluating classroom teaching performance. *Journal of Social Work Education*, 33(2), 261-273.
7. Meltzer, E., Maughan, P., & Fry, T. (1995). Undergraduate in focus: Can student input lead to new directions in planning undergraduate library services? *Library Trends*, 44(2), 400-423.
8. Mohr, K. (2002). Planning for productive college-level work: Using the course assignment framework. *Community College Journal of Research and Practice*, 26, 469-477.
9. Obenchain, Abernathy, T., & Wiest, L. (2001). The reliability of students' ratings of faculty teaching effectiveness. *College Teaching*, 49(3), 100-4.
10. Moses, W. & Ndirangu, W. (2005). An improvement in instructional quality: Can evaluation of teaching effectiveness make a difference? *Quality Assurance in Education*, 13(3), 183-201.
11. Mohr, K. (2002). Planning for productive college-level work: Using the course assignment framework. *Community College Journal of Research and Practice*, 26, 469-477.
12. Meltzer, E. & Maughan, P. (1995). Undergraduate in focus: Can student input lead to new directions in planning undergraduate library services? *Library Trends*, 44(2), 400-23.
13. Ngware, M. & Ndirangu, M. (2005). An improvement in instructional quality: Can evaluation of teaching effectiveness make a difference? *Quality Assurance in Education*, 13(3), 183-201.
14. Obenchain, K, Abernathy, T. & Wiest, L. (2001). The reliability of students' ratings of faculty teaching effectiveness. *College Teaching*, 49(3), 100-104.
15. Read, W., Rama, D. & Raghunandan, K. (2001). The relationship between student evaluations of teaching and faculty evaluations. *Journal of Education for Business*, 76(4), 189-192.
16. Richardson, J. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387-415.
17. Salmon, C., Smith, S., Lee, B., & Miller, V. (2005). Exploring assorted beliefs, doubts, and legends regarding quantitative teaching evaluations. *Journalism & Mass Communication Education*, 60(3), 258-271.
18. Sheehan, E. & DuPrey, T. (1999). Student evaluations of university teaching. *Journal of Instructional Psychology*, 23(3), 188-93.
19. Simpson, P. & Siguaw, J. (2000). Student evaluations of teaching: an Exploratory study of the faculty response. *Journal of Marketing Education*, 22(3), 199-213.
20. Tiberius, R. (2001). Making sense and making use of feedback from focus groups. *New Directions for Teaching and Learning*, 87, 63-75.
21. Timpson, W. & Desley, A. (1997). Rethinking student evaluations and the improvement of teaching: Instruments for change at the University of Queensland. *Studies in Higher Education*, 22(1), 55-6.

NOTES