
Utilizing Instructional Consultations to Enhance the Teaching Performance of Engineering Faculty

CYNTHIA J. FINELLI

*College of Engineering and Center for Research
on Learning and Teaching
University of Michigan*

MOLLY OTT

*School of Education
University of Michigan*

AMY C. GOTTFRIED

*Department of Chemistry
University of Michigan*

CHAD HERSHOCK

*Center for Research on Learning and Teaching
University of Michigan*

CHRISTOPHER O'NEAL

*Center for Research on Learning and Teaching
University of Michigan*

MATTHEW KAPLAN

*Center for Research on Learning and Teaching
University of Michigan*

Keywords: improving teaching, instructional consultation, small group instructional diagnosis (SGID)

I. INTRODUCTION

For many engineering faculty, the desire to engage intellectually with and to make a difference in the lives of students provides strong incentive to be a good teacher (Cross, 1993). But as Felder notes (2004, p. 41), "College teaching may be the only skilled vocation for which systematic training is neither required nor provided." To offer on-the-job-training and continued employee development, a growing number of engineering colleges and universities are establishing teaching centers that offer a range of services, including instructional consultations, classroom observations and interventions, programs on teaching and learning, faculty learning communities, mentoring opportunities, and in-print or electronic resource libraries (Brinko, 1997; Lewis and Povlacs, 2001; National Academy of Engineering, 2007). Instructional consultation, whereby a consultant works with an instructor to assess and enhance teaching, is one of the primary activities provided by such centers, and for good reason. Faculty development scholars assert that instructional consultation is "the most promising way of fundamentally changing postsecondary teaching" (Brinko, 1997, p. vii) and the "best way to instill lasting commitment to and change in a faculty member's teaching" (Lewis and Povlacs, 2001, p. iii).

Of course, various factors influence the impact of consultations, including the kind of data that is used to guide them. *Student ratings of teaching* are one kind of data that instructional consultants often use, and several studies report that they have a positive influence on teaching, and when combined with a consultation; this kind of data is up to four times more effective than student ratings alone (Cohen, 1980; Levinson-Rose and Menges, 1981; L'Hommedieu, Menges, and Brinko, 1990). *Student feedback collected during a Small Group Instructional Diagnosis (SGID)* is another kind of data often used by instructional consultants. Clark and Redmond (1982) and Diamond (2004) show that consultations informed by SGIDs result in significant teaching improvements from both the student and faculty perspectives. A *videotaped class session* is a third kind of data that may be used during consultation. This has the potential to enhance an instructor's teaching performance because it encourages self-reflection (Taylor-Way and Brinko, 1989), but research that systematically assesses the impact of video-based consultations is lacking.

Although instructional consultation has consistently been shown to have a positive impact on teaching (Brinko, 1993; Cashin, 1995; Cohen, 1980; Levinson-Rose and Menges, 1981; Marsh, 1984; Menges and Brinko, 1986; Wilson, 1986), there is

ABSTRACT

Although many kinds of data can be used to guide instructional consultations, research comparing the efficacy of such data is scant, especially in engineering. In this study, multiple modes of assessment were used to evaluate the impact of consultations informed by different kinds of data. This study illuminates two key aspects of instructional consultations: (1) their efficacy varies depending on the kind of data used to guide them, with student feedback from a Small Group Instructional Diagnosis (SGID) having the largest positive impact, and (2) the instructional consultant plays a key role in helping both interpret the available data and identify strategies for improvement. These findings suggest three implications for practice: (1) whenever possible, SGID-based consultations should be offered systematically and proactively for engineering faculty, (2) data for other kinds of consultations should be tailored to the needs of the individual instructor, and (3) instructional consultants should be available to collaborate with faculty to enhance their teaching, thereby building an engineering culture that actively supports teaching and learning.

little research that rigorously compares the efficacy of instructional consultations based on the kind of data used to inform them. Knapper and Piccinin (1999, p. 5) contend that instructional consultation is “based upon the reasonable psychological principle that performance of any task can benefit from knowledgeable feedback, especially when coupled with workable strategies for change and ongoing monitoring of the effect of such change.” They further assert that the discipline of faculty development could benefit from data-driven support of consultation practices.

This study provides such research-based evidence, and it is designed to address the question: *What is the impact of consultations informed by different kinds of data on the teaching performance of engineering faculty?* Specifically, trained instructional consultants used the following kinds of data collected at midterm to guide consultations:

- student ratings of teaching data,
- student feedback collected during a SGID, or
- a videotaped class session.

Three modes of assessment were used to evaluate the impact of the consultations:

- gains in student ratings,
- faculty perceptions of the consultation, and
- reported changes in teaching practice.

II. METHODS

A. The Sample

This project spanned two academic terms, and all engineering faculty who were teaching full-term, undergraduate lecture courses in all engineering departments at the University of Michigan were invited to participate in the study. Faculty members were informed about the methodology of the study¹ through presentations by the research team at department meetings and through group and targeted e-mail messages. Forty-nine engineering faculty participated, representing 12 percent of the eligible faculty population in the College of Engineering, and these instructors agreed *a priori* to participate fully in whichever intervention group they were randomly assigned. Of these, 28 faculty members participated during the first term and 29 participated in the second term. Eight participated both terms, so the sample included 57 instructor units.

The sample includes only 55 separate courses because students evaluated pairs of faculty who team-taught as a single entity for two courses. The 55 courses represent 12 percent of the full-term, engineering undergraduate lecture courses taught during the time period of the study. Although the individual teaching styles of the faculty ranged from straight lecture to extensive use of active and cooperative learning, all courses are considered *lecture courses* in that they took place in a typical lecture setting.

B. Student Ratings of Teaching Survey

Student ratings of teaching derived from multiple-choice surveys are the most common tool used to collect data about teaching effectiveness, and they are one of the most thoroughly studied areas in higher education, with literally thousands of books and articles written on the topic. As a result, a solid literature base supports the

overall validity and reliability of student ratings of teaching (e.g., Aleamoni, 1999; Cashin, 1995; Cohen, 1980; Cross, 1993; Harrison, Douglas and Burdsal, 2004; Marsh, 1984; Marsh and Bailey, 1993; Marsh and Roche, 1993; Menges and Brinko, 1986; Wilson, 1986), despite the controversies that persist about specific issues related to ratings, such as the relationship of grades to ratings. Perhaps most important for this study, ratings have been shown to accurately identify instructors' relative strengths and weaknesses, especially when used at the aggregate level (Aleamoni, 1999; Marsh and Bailey, 1993).

This study employed gains in student ratings as one of three modes of assessment to evaluate the impact of consultations. The survey used to collect student ratings of teaching was drawn from a professionally designed bank of items that comprises the University of Michigan student ratings system. The ratings system has been in use for more than 30 years, and it currently serves over 12,000 classes and has more than 400,000 student responses annually. The 17 items on this study's survey focus on both the research-based traits of effective teaching from *Seven Principles of Good Practice in Undergraduate Education* (Chickering and Gamson, 1987; Sorcinelli, 1991) and classroom behaviors linked to college teaching effectiveness (Murray, 1985). While a few global items were included, namely overall course and teacher ratings, most items purposefully identified areas where instructors could enhance their teaching through specific interventions (e.g., use of class time, level of preparation, enthusiasm, and individual rapport). Although the survey could have solicited open-ended student comments, for the purposes of this project, it did not. A sample student ratings of teaching survey is contained in Appendix A.

Student surveys were collected at midterm and again at the end of the term in all participating classes in order to assess the impact of the instructional consultations. As is common in studies that measure the effect of pedagogical changes on student learning, a normalized gain score (Hake, 1998) was used in this study. The score is defined as the ratio of the actual gain (average end-of-term rating for an individual course minus average midterm rating) to the maximum possible gain (5.0 minus average midterm rating), and it accounts for both individual differences in midterm ratings and potential improvement. For the 55 individual courses in this project, the normalized gain score was computed for each of the 17 items on the student ratings of teaching survey. Then, for each item, the *average normalized gain score* for each intervention group was calculated.

C. Intervention Groups

Faculty members were randomly assigned to a control group that received no instructional consultation (*Group 1*) or to one of several intervention groups that received a consultation informed by different kinds of data (*Groups 2–4*). Table 1 outlines aspects of the project for each intervention group.

Faculty in *Group 2* had an instructional consultation informed solely by a summary report of midterm student ratings of teaching collected three to four days prior to the consultation. Together with an instructional consultant, these instructors drew only on the student ratings data to identify strategies to enhance teaching.

Consultations for faculty in *Group 3* were guided by a Small Group Instructional Diagnosis (SGID) (e.g., Black, 1998; Clark and Redmond, 1982; Lenze, 1997). The instructional consultant met with the instructor a few days before a designated class session to discuss course goals, and on the day of the class, the consultant

¹All aspects of the project were approved for human subject research by the University of Michigan's Behavioral Sciences Institutional Review Board prior to the start of the project.

	<i>Group 1a</i> No consult, no ratings	<i>Group 1b</i> No consult, ratings only	<i>Group 2</i> Consult on ratings	<i>Group 3a</i> SGID, no ratings	<i>Group 3b</i> SGID, with ratings	<i>Group 4a</i> Videotape, no ratings	<i>Group 4b</i> Videotape, with ratings
Middle of term	Summary report on student ratings [†]	✓	✓		✓		✓
	Student feedback from SGID			✓	✓		
	Self reflection on a videotaped class					✓	✓
	Instructional consultation*			✓	✓	✓	✓

[†] Faculty in *Group 2* received a consultation informed by their midterm student ratings data, while faculty in *Group 1b* did not receive the ratings and faculty in *Groups 3b* and *4b* received them only after their assigned consultation was complete.

* Instructional consultation varied by intervention group, but it always involved the presence of a trained consultant working together with the faculty member to interpret the data and develop strategies for improving teaching.

Table 1. Aspects of the project for each intervention group.

observed the first portion of the class session. Then, during the final 20-minutes of the class session, the instructor left the classroom and the instructional consultant collected feedback from students in a two-step process. First, students worked in small groups to fill out a worksheet with two items: (1) "List the major strengths in this course; what is helping you learn in the course? Please explain briefly or give an example for each strength," and (2) "List changes that could be made in the course to assist you in learning. Please explain how suggested changes could be made." One student in each small group recorded responses about which there was general agreement. As a second step in the feedback process, the instructional consultant facilitated a whole-class discussion, noting both the consensus responses and any disagreements. As needed, the consultant probed students in the class about issues that were unclear. Finally, the consultant prepared a report summarizing the student comments and then met with the faculty member to discuss and interpret the data and to develop an approach for addressing the issues. This form of feedback is regularly used for purposes of teaching enhancement, at both the University of Michigan and elsewhere, and it has been shown to be an effective catalyst for change in college classrooms (e.g., Diamond, 2004).

Faculty in *Group 4* received an instructional consultation that was informed by a videotaped class session. As with *Group 3*, the instructor and consultant met in the middle of the term shortly before a given class session to discuss course goals. Then the instructional consultant videotaped the entire class session, capturing both faculty and student activities. The subsequent consultation included reviewing the videotape, reflecting on teaching practices, and establishing a plan to enhance teaching. The protocol used for the consultation is based on work by Taylor-Way and Brinko (1989).

All instructors received an individualized report comparing midterm and end-of-term student ratings when the term was over, but some also received a summary report of the ratings at midterm. Accordingly, *Groups 1, 3, and 4* were further subdivided into faculty who did not receive any student ratings until the term ended (*Groups 1a, 3a, and 4a*), and faculty who were given a summary report at the middle of the term (*Groups 1b, 3b, and 4b*). For *Groups 3b* and *4b*, the summary report was given at the end of their consultation, and the instructional consultant did not refer to the ratings

(there was no consultation for faculty in *Group 1b*). All instructors in *Group 2* received the summary report on midterm student ratings as the basis for the consultation. For this project, the midterm student ratings data was also used to assess changes in teaching performance; however, a more objective measure of teaching performance would have provided a stronger assessment.

Trained instructional consultants at the University of Michigan's Center for Research on Learning and Teaching (CRLT) who are affiliated with CRLT North² conducted all consultations. They hold Ph.D.'s in science, technology, or engineering, and they had all participated in extensive training that included reviewing handbooks for instructional consultation (including the popular references *Practically Speaking* [Brinko and Menges, 1997] and *Face to Face* [Lewis and Povlacs, 2001]), discussing case studies, and engaging in role-plays of consultations. Instructional consultants were randomly assigned to faculty, and each conducted approximately the same number of consultations in each intervention group. All consultations followed a consistent protocol that included a structured format for the pre-class meeting, guiding questions to focus the consultation, and follow-up strategies for developing a specific plan of action. A sample consultation protocol from *Group 3b* is included in Appendix B. The instructional consultants met regularly as a group to standardize their practice during the course of the study.

D. Faculty Feedback

In addition to gains in student ratings, faculty perceptions of the consultation and reported changes in teaching practice were used to assess the impact of consultations. As a way to provide this feedback,

²Established in 1962, CRLT is the oldest teaching center in the nation (Professional and Organizational Development in Higher Education Network, 1995). Its 13 instructional consultants conducted more than 2,000 individual consultations for instructors (including faculty and teaching assistants) and more than 290 SGID-based consultations during the 2007–2008 academic year (Center for Research on Learning and Teaching, 2008). CRLT North is the engineering branch of CRLT that was established in 2004 especially to support teaching and learning initiatives for engineering instructors (Center for Research on Learning and Teaching North, 2008). The office is funded by the College of Engineering, and it employs a full-time director, a half-time assistant director, and a half-time program assistant.

		Sample [*] N = 49	Group 1a No consult, no ratings N = 7	Group 1b No consult, ratings only N = 11	Group 2** Consult on ratings N = 11	Group 3a SGID, no ratings N = 7	Group 3b SGID, with ratings N = 5	Group 4a Videotape, no ratings N = 7	Group 4b Videotape, with ratings N = 9
Gender	Male	39 (80%)	6 (86%)	9 (82%)	8 (73%)	6 (86%)	5 (100%)	6 (86%)	6 (66%)
	Female	10 (20%)	1 (14%)	2 (18%)	3 (27%)	1 (14%)	0 (0%)	1 (14%)	3 (33%)
Rank	Full professor	17 (35%)	0 (0%)	7 (64%)	4 (36%)	3 (42%)	0 (0%)	3 (42%)	3 (33%)
	Associate professor	11 (22%)	4 (57%)	0 (0%)	2 (18%)	2 (29%)	1 (20%)	0 (0%)	2 (22%)
	Assistant professor	14 (29%)	3 (43%)	1 (9%)	3 (27%)	2 (29%)	3 (60%)	2 (29%)	3 (33%)
	Non-tenure track	7 (14%)	0 (0%)	3 (27%)	2 (18%)	0 (0%)	1 (20%)	2 (29%)	(11%)
Had prior consultation	23 (47%)	4 (57%)	4 (36%)	4 (36%)	4 (57%)	4 (80%)	2 (29%)	7 (78%)	
Years teaching experience	13.4 ± 9.8	9.6 ± 7.0	21.2 ± 11.1	11.6 ± 7.3	13.3 ± 10.4	7.3 ± 6.2	10.1 ± 7.8	14.8 ± 11.3	
Years at institution	9.5 ± 9.3	7.3 ± 7.1	15.6 ± 12.7	7.6 ± 5.2	10.1 ± 8.1	3.0 ± 1.2	8.2 ± 7.3	10.7 ± 12.3	
<i>Behaviors to Improve Teaching. Reported time spent engaging in the following behavior on a scale of 1 (never), 2 (occasionally), and 3 (often):</i>									
	Talk to others about teaching		2.3	2.1 ⁺⁺	2.5 ⁺	2.3	1.8	2.3	2.3
	Solicit feedback from students about how a class is going		2.6	2.3 ⁺⁺	2.6 ⁺	2.6	2.8	2.4	2.4
	Spend time revising the content of courses		3.0	2.9 ⁺⁺	2.7 ⁺	2.4	2.5	2.6	2.9
	Spend time revising the methods used to teach courses		2.4	2.3 ⁺⁺	2.5 ⁺	2.1	2.5	2.3	2.7
	Attend programs on effective teaching		1.6	1.8 ⁺⁺	1.7 ⁺	1.4	2.3	2.1	2.1
	Read journals or books on effective teaching		1.7	2.0 ⁺⁺	2.0 ⁺	1.4	1.8	1.7	1.6
<p>[*] Eight individuals participated both times, so the total number of unique participants (49) is eight less than the sum of individuals in the groups (57).</p> <p>^{**} Two courses in <i>Group 2</i> were team-taught, but students rated the team of faculty members as a single entity, so the number of individual faculty members represented in the groups of Table 2 is two more than the number of courses of Table 3.</p> <p>⁺⁺ Data is not available from two individuals in this category, so <i>N</i> is two fewer than listed at the top of the column.</p> <p>⁺ Data is not available from one individual in this category, so <i>N</i> is one fewer than listed at the top of the column.</p>									

Table 2. Faculty characteristics for the entire project and for each intervention group. For years teaching, the average and standard deviation are reported.

instructors completed a short online survey after the final student ratings of teaching were collected for their course. The survey included items for faculty to indicate how often they engaged in several activities related to teaching enhancement and whether they had participated in a teaching consultation of any sort prior to their involvement in the project.

Further, faculty who received a consultation during the project rated both the consultation and the instructional consultant, described the pros and cons of the consultation, and discussed any changes they implemented as a result. Instructors in the two intervention groups that received a midterm report on student ratings at the conclusion of a consultation but who did not have the assistance of an instructional consultant in interpreting the data (*Groups 3b* and *4b*) also indicated whether they had used the report to reflect on their teaching. If they had used the report, faculty in these groups rated how useful it was for improving their teaching. There were three open-ended items at the end of the survey. Faculty members who had received a consultation were asked to identify the most useful and the least helpful aspects of the consultation, and all faculty were

invited to describe anything else they wanted the research team to know.

III. RESULTS

A. Descriptives

Faculty characteristics for the entire project and for each intervention group are summarized in Table 2. Of the 49 faculty who participated, 10 (20 percent) were women. This proportion is slightly higher than the fraction of engineering faculty at the University of Michigan who were women (13 percent) during the time period of the study. All faculty ranks were represented, and the sample includes individuals with a broad range of teaching experience. Twenty-three instructors (47 percent) had received a teaching consultation sometime in the past.

Course characteristics for the entire project and for each intervention group are summarized in Table 3. The courses spanned all undergraduate course levels from introductory (100-level) to senior

	Sample N = 55	Group 1a No consult, no ratings N = 7	Group 1b No consult, ratings only N = 11	Group 2 Consult on ratings N = 9	Group 3a SGID, no ratings N = 7	Group 3b SGID, with ratings N = 5	Group 4a Videotape, no ratings N = 7	Group 4b Videotape, with ratings N = 9	
Course level	100-level	7 (13%)	0	2	2	0	0	2	1
	200-level	8 (14%)	1	2	0	1	1	2	1
	300-level	17 (31%)	3	4	4	2	1	2	1
	400-level	23 (42%)	3	3	3	4	3	1	6
Credits	2-credit	1 (2%)	1	0	0	0	0	0	0
	3-credit	11 (20%)	1	4	3	1	0	1	1
	4-credit	43 (78%)	5	7	6	6	5	6	8
Course size	62 ± 46	43 ± 34	55 ± 32	80 ± 60	38 ± 17	45 ± 24	81 ± 57	81 ± 62	

Table 3. Course characteristics for the entire project and for each intervention group. For course size, the average and standard deviation are reported.

(400-level), represented a range of credits from two to four units, and encompassed a wide range of course sizes.

The assignment of faculty to intervention groups was random, and several descriptive measures indicate that the groups were comparable in composition. There were approximately the same number of participants in each intervention group, and the groups were similar in terms of faculty demographics (gender and rank), teaching experience (years at institution and years of teaching experience), level of engagement in behaviors related to enhancement of teaching, and prior participation in a teaching consultation. In addition, the groups were roughly equivalent with respect to course composition (course level, credits, and course size).

A total of 2,579 midterm student ratings of teaching surveys and 2,296 end-of-term surveys were completed in the 55 participating courses. Table C1 in Appendix C shows the 95 percent confidence interval of the average ratings for each of the 17 items, by intervention group. The groups were similar in terms of initial teaching ability—an Analysis of Variance (ANOVA) test of the average student ratings at midterm indicates that no significant differences existed according to group assignment for any of the 17 items ($F_{6,48} < 2.01; p \geq 0.05$).

B. Gains in Student Ratings

Table C2 in Appendix C shows the 95 percent confidence interval of the average normalized gain score for each item on the student ratings of teaching survey, by intervention group. Every group had positive gain on some of the 17 items; however, only four of the intervention groups had one or more items in which gains were statistically significant ($p < 0.05$), as shown in Figure 1.

The groups that had a consultation informed solely by student ratings of teaching (Group 2) or by a videotape of their class (Groups 4a and 4b) had at most one item with a statistically significant positive gain ($p < 0.05$). However, the groups that had a consultation guided by student feedback from a SGID (Groups 3a and 3b) demonstrated improvement on six and three items, respectively. Notably, student ratings of “enthusiasm,” “use of techniques to foster class participation,” and “level of instructor respect towards students” improved more for faculty who had participated in

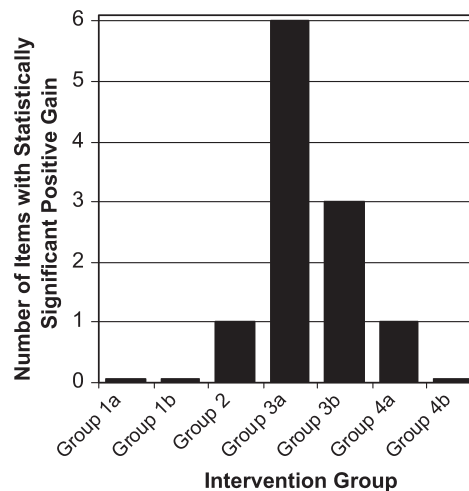


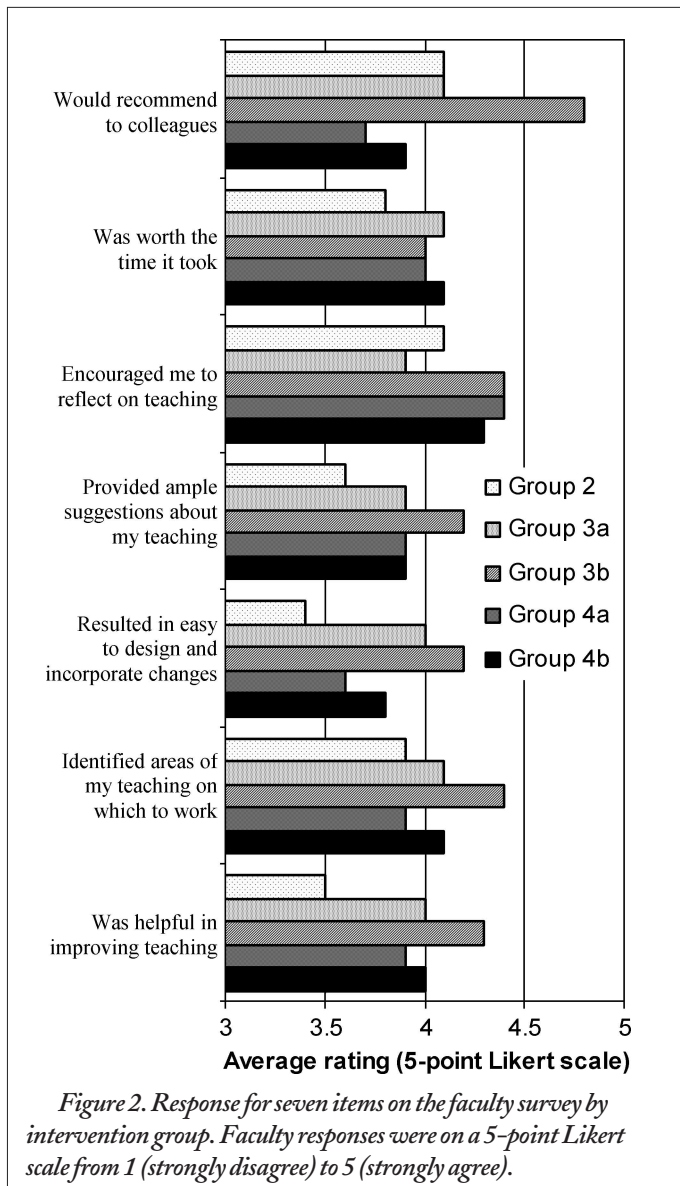
Figure 1. Improvement in student ratings by intervention group.

consultations informed only by student feedback collected during a SGID (Group 3a) than for faculty in all other intervention groups.

Further analysis of the results provides some insight into the value of the instructional consultant. The two intervention groups that received a consultation but did not receive student ratings of teaching (Groups 3a and 4a) had statistically significant positive gains ($p < 0.05$) for six items and one item, respectively. For these two groups, a trained instructional consultant assisted in interpreting the data available to them. On the other hand, the comparable groups that did receive student ratings of teaching data (after the consultation) but lacked the assistance of an instructional consultant to interpret the data (Groups 3b and 4b) had fewer items exhibiting a significant positive gain (i.e., one and zero items, respectively).

C. Faculty Perceptions of the Consultation

Figure 2 shows the average response for seven items of the faculty survey for each of the five intervention groups that had a consultation.



In general, regardless of grouping, instructors agreed that the consultation was helpful in improving teaching, that it was easy to identify areas on which to work, and that it was easy to design and incorporate changes in teaching based on the consultation. Across intervention groups, faculty also agreed that the instructional consultant provided ample suggestions about teaching and encouraged the instructor to reflect on his/her teaching. Finally, across intervention groups, faculty agreed that it was worth the time it took to complete the consultation and that they would recommend the consultation to colleagues in their department.

Despite a number of suggestive differences in faculty perceptions between intervention groups, none was statistically significant (Kruskal-Wallis H Test). This alone is an interesting finding, as it suggests that faculty found consultations informed by the various kinds of data used in this study similarly valuable. Notably, consultations informed by student feedback from a SGID coupled with the receipt of a summary report on student ratings (*Group 3b*) were rated most highly, even though the rating was not significantly higher than for the other groups.

About 80 percent of faculty in the two groups that received a summary report of midterm student ratings of teaching independent

of the consultation (4/5 or 80 percent in *Group 3b*; 7/9 or 78 percent in *Group 4b*) reported using the data to reflect on teaching. However, they consistently rated the “helpfulness of the report in improving teaching” lower than the “helpfulness of the consultation” (3.7/5.0 versus 4.3/5.0 for *Group 3b* and 3.4/5.0 versus 4.0/5.0 for *Group 4b*). Although the instructional consultants had training to interpret student ratings data and relate it to strategies for improving teaching, the research design required consultants *not* to address the student ratings data with faculty in these groups. Without such guidance, instructors may have found the data overwhelming or contradictory, making it difficult for them to identify “workable strategies for change” as advocated by Knapper and Piccinin (1999). This hypothesis is supported by one member of *Group 4b* who noted, “In my opinion, the ratings did not match the positive feedback that I received from the consultation.”

The faculty survey also had open-ended items for faculty to identify the most useful aspects of the project. Among groups receiving student feedback from a SGID, a consistent theme that emerged was the importance of the instructional consultant’s classroom visit for interpreting students’ comments and feedback. One instructor listed “the process of bringing students to consensus” as the most useful aspect, and another stated, “getting feedback from students through a neutral party (mediator) [was useful]. Students are not always forthcoming when asked directly by the instructor.” Many instructors who viewed a videotape of the class session noted its merit during the consultation. For example, one noted, “being able to see what students were doing while I was lecturing was helpful.” Other highly valued aspects of the project were the consultation itself and the suggestions provided by the instructional consultant. One instructor particularly appreciated discovering “strategies to increase student participation. This is something I feel I should improve on, and the consultant gave specific and helpful suggestions.”

When asked to describe the least helpful aspects of the consultation, only half of the faculty responded. One participant replied, “Believe it or not, everything was helpful so I really cannot think of something that could be considered as LEAST helpful.” Faculty responses generally involved study design issues, such as suggestions for changing the student ratings of teaching form, rather than comments about the consultation or its perceived efficacy. However, two main themes did emerge regarding the least helpful aspects of the project. First, instructors noted that significant in-class time was spent administering the student surveys.³ For example, a participant who received a SGID-based consultation stated, “three in-class interruptions was more than I was ready for. I lost a total of about one 50-minute lecture.” Second, a number of instructors commented that the short time horizon between the consultation and the end of the term limited their ability to modify their teaching methods or to assess the effect of making changes. One participant in a video-based consultation noted:

This was a useful exercise. I don’t think there was much time to make a substantial change in teaching style for the current course (that is why I selected *agree* rather than

³Most in-class time for this project was dedicated to collecting student ratings of teaching at midterm and end-of-term for comparative purposes. However, a typical consultation scenario would not require as much assessment as this research project did.

strongly agree), since most of it was done around mid-semester or a little later. I hope to use what I learned in the future, and I am sure it should improve my overall teaching style. So my advice is that if one wants to gather more accurate data, this should continue for a year rather than a semester, that will help to build better teaching practice and may be better to observe its overall impact.

Another instructor whose consultation was informed by student ratings data believed that the short time frame of the study caused student ratings to decrease, when students failed to perceive immediate improvements after having provided feedback at midterm. That instructor commented:

What I did not expect was the impact of the assessment on student expectations. The midterm assessment was okay, because the students hoped that change could come of the course. Unfortunately, many of the changes that needed to be made (e.g., finishing up the textbook that is now only halfway done, not having a robust system for grading) will easily take another year. That is not fast enough for this term's students. Consequently, the study's survey at the end just seemed to upset them more. I would guess it's because it appeared as if I didn't listen to their comments in the first place ... I was disappointed/discouraged at the way the term played out for this class.

The second common theme in the open-ended comments involved the format of the consultation for the video-based groups. For this study, the consultation was designed to be learner-focused, with the learner being the faculty member. In other words, self-reflection, rather than feedback from the students or the instructional consultant, was the primary technique used. Faculty reaction to this approach was mixed. Many comments were consistent with this one:

The fact that the focus appeared to be on me evaluating my own teaching seemed of limited value. I would have preferred a more informed critique in which the consultant had viewed the videotape before and could point out particular issues for discussion.

Other instructors felt that the consultant offered valuable input and found "seeing the tape, and having a chance to discuss it with an objective observer" to be the most helpful part of the project. Although the protocol for viewing the videotaped class session differed slightly for the two groups (for *Group 4a* faculty took the lead in reviewing the videotape, whereas with *Group 4b* the consultant identified several key segments before the consultation), there was a structured consistency within each group. Thus, the discrepancy among faculty opinions is more likely to be a function of instructor preference than of consultant style.

Finally, several instructors responded to the invitation to list anything else they wanted the team to know about their experience with the research project. Most thanked the team for allowing them to participate in the project, and some suggested making the opportunity more widely available to faculty. One person who participated in a SGID-consultation recommended "continuing the consultations and making them standard practice."

D. Changes in Teaching Practice

Faculty in every intervention group that received a consultation responded to the open-ended item, "Please describe any specific changes you made in your teaching based on the consultation," and 27 reported changes they had made. The most commonly cited change (noted by 10 faculty [37 percent]) was introducing more active learning into the course. Explaining concepts more clearly was listed by five instructors (18.5 percent), as was introducing more examples in class. Other common responses included: managing class time differently (three responses or 11.1 percent), giving feedback to students more promptly (two or 7.4 percent), calling on students by name more often (two or 7.4 percent), and changing the pace of the class (two or 7.4 percent). The proportion of faculty who reported making changes was greatest in the group that received a SGID-based consultation without midterm student ratings (changes were reported by 36 percent, 46 percent, 86 percent, 60 percent, 71 percent and 44 percent of the faculty in *Groups 1b, 2, 3a, 3b, 4a, and 4b*, respectively).

In addition, instructors who had a consultation cited more substantial changes than those who did not have a consultation. For example, this response from one member of *Group 4a*, "[I] prepared and used questions and in-class exercises that are more thought-provoking," and this one from a member of *Group 3a* "[I] invited more questions in class, increased the number of examples presented in class, kept the students better informed about schedules, and suggested changes to the teaching assistant" are typical of the types of changes noted by instructors who had a consultation. On the other hand, faculty who did not have a consultation provided less detailed descriptions of the teaching changes they had incorporated. This comment is typical of those from faculty who did not have a consultation, "I tried to modify my teaching to allow more comments from the class."

V. DISCUSSION

Faculty who received a consultation guided by student feedback collected during a SGID saw the greatest gains in student ratings, rated most aspects of the consultation as high or higher than faculty who did not receive a SGID, and reported more detailed changes in teaching than those who did not have a consultation. Although none of these measures is truly indicative of better student learning, each suggests some effect on teaching; and taken together, they provide evidence of the impact of consultations on teaching. There are a number of possible explanations for this. For example, Weimer (1990) posits a five-step assessment/feedback model for changing teaching performance: develop instructional awareness, gather information, select alternatives, implement a plan, and assess effectiveness. This model suggests that successful changes in teaching performance require faculty to gather information about their practice and completely consider alternatives for improvement. Of the kinds of data in this study, only data from a SGID involved careful collection, by the instructional consultant, of student feedback that was then filtered through the prism of a consultant's expertise. Perhaps this approach best exemplifies Weimer's model.

Another explanation is that the SGID process engages not only faculty, but students too; students offered concrete suggestions and reinforced positive behaviors, and the instructional consultant engaged the faculty member during the consultation to underscore student comments. Furthermore, because students define the issues significant

to them rather than responding to a set of pre-determined categories, and because the instructional consultant can probe the class to understand the issues more fully, SGIDs allow the instructor to target solutions to students' expressed needs. The SGID also explores differences in opinion, so if there is a disagreement in the class, the instructional consultant will probably understand it more fully than would be the case from seeing a bi-modal distribution on a ratings item or from written feedback with no opportunity for further exploration of student opinions. There is a third possible explanation for the success of this approach: because students were positively and deeply engaged in the improvement process, they may have responded more generously when rating teaching at the end of the term.

Importantly, consultations informed by other kinds of data were also effective for some individual instructors. While faculty who viewed videotaped classes in this study did not have *uniform* gains in student ratings when analyzed in the aggregate, some individual faculty did show major improvement. Half (8/16) of the faculty in *Groups 4a* and *4b* improved overall (in particular, they had positive gains from midterm to end-of-term on 12 or more of the 17 student ratings items), while the other half worsened overall (i.e., they had negative gains on 10 or more individual items). In fact, one instructor had a positive gain on every single one of the 17 items, while one faculty member had a negative gain on every single item. Faculty perceptions of the video-based consultations also varied—some instructors valued the self-reflection while others disliked that aspect of the consultation. The fact that this intervention was more effective for some specific faculty than for others suggests that video-based consults should only be used when they meet the needs of the individual instructor. Instructors who have previously engaged in reflective practices about teaching, whether through formal processes like a consultation or through informal processes undertaken individually, may be the best candidates for this type of consultation. Further, as Taylor-Way and Brinko (1989) caution, proficiency in analyzing videotaped material is difficult to develop for both instructional consultants and clients and it requires a great deal of concentration, coaching, time, and effort. Thus, this intervention should be used judiciously.

Consultations informed solely by student ratings data resulted in limited improvement in student ratings. Only 11 percent (1/9) of the faculty in *Group 2* improved overall while 44 percent (4/9) worsened overall in terms of student ratings. Additionally, faculty in this group rated most aspects of the intervention lower than other faculty did, and fewer of them reported making changes in their teaching as a result of the intervention. It is worth reiterating that although the student survey included only multiple-choice items for this project, it could have solicited open-ended student responses, possibly resulting in better improvements in teaching than exhibited in the present study. However, doing so would not provide the same benefit as having the instructional consultant in the classroom with the students to clarify and interpret their comments and explore differences in opinion, a critical aspect of the SGID-based consultations that was noted by faculty in those intervention groups.

This study also provides an important understanding of the actual consultation experience. Potential factors that contribute to the success of a consultation are numerous, and they might include: "the number of meetings (and hence, relationship) between the consultant and client; the presence or absence of feedback; the model of consultation espoused by the consultant; [instructor] expectations; personality traits; synergy between consultant and client;

and institutional climate" (Brinko, 1990, p. 75). Whether the consultation is conducted by professional specialists, trained faculty, or graduate students might also affect the consultation (Piccinin, Cristi, and McCoy, 1999). These variables are typically not controlled for or are disregarded in studies of the efficacy of instructional consultation. However, standard consultation protocols developed for this study and consistent consultant training allows others to design consultations that may replicate these successful findings.

Another important factor that may have an impact on the consultation is the academic background of the faculty member and the instructional consultant, especially since successful teaching practices are likely to vary widely by discipline. For example, Weston and McAlpine (1999) recommend discipline-specific consultation approaches, and studies by both Cashin (1990) and Cranton and Smith (1990) show that discipline can be significantly related to how students rate instruction. The engineering focus provided by this research (including the disciplinary expertise of the instructors as well as the instructional consultants' backgrounds in science, technology, engineering, and mathematics) lays the foundation for comparing practices of teaching and faculty development across disciplines.

There are, of course, limitations to the present research. As noted by Hicks (1999), one of the major concerns noted in studies of instructional consultation is the generally small number of faculty who seek this form of assistance. As with other studies of instructional consultation, the relatively small sample size in this research (49 unique faculty teaching 55 distinct courses) does pose a challenge to internal validity and statistical power (L'Hommedieu, Menges, and Brinko, 1990). However, because this sample represents 12 percent of the eligible population, engineering faculty teaching undergraduate lecture courses, sample size may be less of an issue. In addition, like most published studies in which faculty self-present at university teaching centers (e.g., Cashin and Perrin, 1983; Piccinin, Cristi and McCoy, 1999; Piccinin and Moore, 2002), generalizing from the research sample to the entire population of faculty may be problematic. While participants in this study volunteered without knowing the kind of data they would receive (thus addressing some of the self-selection limitation because they could not specify the exact type of feedback they desired), they did know that the study was sponsored by the University of Michigan's Center for Research on Learning and Teaching and was supported by the College of Engineering administration. This may have contributed to some self-selection bias. In fact, for this study, nearly one-half of the participants had previously received a teaching consultation, so it is likely that they were pre-disposed to seeking feedback, and a random sample would have been more representative of the general faculty core.

Another limitation is that, although faculty motivation is a key factor in affecting change (Blackburn and Lawrence, 1995), the design of this study did not necessarily maximize motivation. Brinko (1993, p. 582) reported that "feedback is more effective when the recipient is able to select the way in which it is conveyed," and the instructional consultants who undertook this study agreed with this principle. However, creating a controlled environment for this study required random assignment of faculty to intervention groups, and as a result, the consultation received may not have been aligned with each instructor's individual needs and teaching style. Rarely in practice would an instructional consultant randomly choose a feedback process for a faculty member; rather, the instructional consultant and

instructor would work in tandem to find the best fit for the individual and the course. The gains in teaching performance realized in this research might have been even greater had the consultations been more carefully chosen.

The short duration of this study is also a limitation. Even though the goal of the SGID is to provide instructors with student feedback with the purpose of immediately implementing strategies to improve instruction for the specific course (Diamond, 2004), student ratings of teaching may not reflect this immediacy. Previous research has indicated that consultations informed by some kinds of data may have a lagged effect such that significant improvement in student ratings of teaching may not be evident until one to three years after the consultation (Marsh and Roche, 1993; Piccinin, 1999). The period for this study, from the middle to the end of the academic term, was just six to eight weeks, so understanding the long-term impact is beyond the scope of this study. For this reason, the authors have begun to collect longitudinal data on student ratings of teaching for the faculty involved in this project, with the intention of analyzing the long-term impact of teaching consultations on performance.

VI. CONCLUSION

The present study makes a unique contribution to the literature by filling gaps in existing research and providing a systematic comparison of the impact of consultations informed by different kinds of data (i.e., student ratings data, student feedback collected during a SGID, or a videotaped class session). The study also provides valuable insight into the consultation experience itself, it employs standard consultation protocols and trained instructional consultants, and it focuses exclusively on faculty in engineering. Equally important, it uses multiple modes of assessment (gains in student ratings, faculty perceptions of the consultation, and reported changes in teaching practice) to evaluate the impact of the consultations. Accordingly, this study's findings should be used to inform the consulting work of the growing number of engineering teaching centers (National Academy of Engineering, 2007).

Two principal findings in the data provide insight into the question: *What is the impact of consultations informed by different kinds of data on the teaching performance of engineering faculty?* First, the efficacy of consultations does vary depending on the kind of data used to guide them. In general, faculty in the intervention groups that received student feedback from a SGID had the most positive impact. Consultations guided by a videotaped class session had varying success, and the use of student ratings data alone resulted in the least positive effect. Second, an instructional consultant plays a key role in assisting the faculty member to both interpret the available data and identify strategies for improving teaching. Drawing on their experience and professional judgment, instructional consultants had the ability to quickly direct faculty attention to specific teaching practices and avoid overwhelming the instructor with too much information. By way of contrast, instructors who received feedback data without the assistance of an instructional consultant could not benefit from a trained, neutral third party who invited them to reflect on their teaching or assisted in interpreting the data. These faculty members may not have known where to start, never identifying a manageable number of specific strategies on which to focus.

These findings suggest three implications for practice. First, because SGID-based consultations resulted in the greatest overall impact in a variety of engineering settings for instructors of different experience, SGIDs should be offered systematically and proactively for engineering faculty whenever possible. Second, data for other kinds of consultations should be tailored to the individual needs of the instructor. For example, faculty who are able to view their teaching objectively, who can analyze behaviors from a neutral perspective, or who use a range of teaching techniques over the course of a single class session might especially benefit from video-based consultations. Likewise, student ratings data might be particularly useful for faculty with specific, concrete issues (related to class mechanics, for instance). Ideally, data from multiple, complementary sources should be used to inform the consultation. And third, instructional consultants should be available to collaborate with individual instructors to enhance their teaching. Such consultants may be instrumental in helping faculty with the difficult but essential move from data to strategies for improving their teaching. Although this project does not negate the value of informal consultations with colleagues, it does underscore the importance of having a teaching center where engineering faculty can request the assistance of trained instructional consultants, as well as participate in activities of the center. This will lay the foundation for building an engineering culture that actively supports teaching and learning and normalizing the practice of employing instructional consultants in engineering teaching centers.

V. ACKNOWLEDGMENTS

The authors would like to thank several individuals at the University of Michigan for their valuable assistance: Mary Piontek, Evaluation Researcher at the Center for Research and Learning and Teaching (CRLT), provided helpful guidance and insight for the research design and analysis; Vilma Mesa, Assistant Professor at the School of Education, and Erping Zhu, Coordinator of Instructional Technology at CRLT, conducted several consultations; and Virginia Hamori-Ota, Ryan Hudson, and Marie Kendall Brown offered useful editorial assistance. The authors are also grateful for financial support from the University of Michigan's College of Engineering and from a mini-grant awarded through the 2005 NSF Rigorous Research in Engineering Education: Creating a Community of Practice workshop.

REFERENCES

- Aleamoni, L.M. 1999. Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education* 13 (2): 153–66.
- Black, B. 1998. Using the SGID method for a variety of purposes. In *To improve the academy: Resources for faculty, instructional, and organizational development*, ed. M. Kaplan, 245–262. Stillwater, OK: New Forums Press.
- Blackburn, R., and J. Lawrence. 1995. *Faculty at work: Motivation, expectation, satisfaction*. Baltimore, MD: The Johns Hopkins University Press.
- Brinko, K.T. 1990. Instructional consultation with feedback in higher education. *The Journal of Higher Education* 61 (1): 65–83.
- Brinko, K.T. 1993. The practice of giving feedback to improve teaching: What is effective? *The Journal of Higher Education* 64 (5): 574–93.

- Brinko, K.T., and R.J. Menges, ed. 1997. *Practically speaking—A sourcebook for instructional consultants in higher education*. Stillwater, OK: New Forums Press.
- Cashin, W.E. 1990. Students do rate different academic fields differently. *New Directions for Teaching and Learning* 43: 113–22.
- Cashin, W.E. 1995. Student ratings of teaching: The research revisited. *Idea Paper No. 32*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Cashin, W.E., and B.M. Perrin. 1983. Do college teachers who voluntarily have courses evaluated receive higher student ratings? *Journal of Educational Psychology* 75 (4): 595–602.
- Center for Research on Learning and Teaching. 2008. Summary of CRLT Services for 2007–2008. <http://www.crlt.umich.edu/aboutcrlt/servicessummary0708.php> (last accessed September 2008).
- Center for Research on Learning and Teaching North. 2008. *Overview of CRLT North*. <http://www.engin.umich.edu/teaching/crltnorth/> (last accessed September 2008).
- Chickering, A.W., and Z.F. Gamson. 1987. Seven principles for good practice in undergraduate education. *AAHE Bulletin* 39 (8): 3–7.
- Clark, D., and M. Redmond. 1982. *Small group instructional diagnosis: Final report*. ERIC Document Reproduction Service No. ED217954.
- Cohen, P. 1980. Effectiveness of student-rating feedback for improving college instruction: A meta-analysis. *Research in Higher Education* 13: 321–41.
- Cranton, P. and R.A. Smith. 1990. Reconsidering the unit of analysis: A model of student ratings of instruction. *Journal of Educational Psychology* 82 (2): 207–12.
- Cross, K.P. 1993. On college teaching. *Journal of Engineering Education* 82 (1): 9–14.
- Diamond, M.R. 2004. The usefulness of structured mid-term feedback as a catalyst for change in higher education classes. *Active Learning in Higher Education* 5 (3): 217–31.
- Felder, R.M. 2004. Teaching engineering at a research university: Problems and possibilities. *Educación Química* 15 (1): 40–42.
- Hake, R.R. 1998. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Association of Physics Teachers* 66 (1): 64–74.
- Harrison, P.D., D.K. Douglas, and C.A. Burdsal. 2004. The relative merits of different types of overall evaluations of teaching effectiveness. *Research in Higher Education* 45 (3): 311–23.
- Hicks, O. 1999. A conceptual framework for instructional consultation. *New Directions for Teaching and Learning* 79: 9–18.
- Knapper, C. and S. Piccinin. 1999. Consulting about teaching: An overview. *New Directions for Teaching and Learning* 79: 3–8.
- Lenze, L.F. 1997. Small Group Instructional Diagnosis (SGID). In *Practically speaking—A sourcebook for instructional consultants in higher education*, eds. K.T. Brinko and R.J. Menges, 143–146. Stillwater, OK: New Forums Press.
- Levinson-Rose, J., and R.J. Menges. 1981. Improving college teaching: A critical review of research. *Review of Educational Research* 51 (3): 403–34.
- Lewis, K.G., and J.T. Powlacs, eds. 2001. *Face to face—A sourcebook of instructional consultation techniques for faculty/instructional developers*. Stillwater, OK: New Forums Press.
- L'Hommedieu, R., R.J. Menges, and K.T. Brinko. 1990. Methodological explanations for the modest effects of feedback. *Journal of Educational Psychology* 82: 232–41.
- Marsh, H.W. 1984. Students' evaluations of teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology* 76 (5): 707–54.
- Marsh, H.W., and M. Bailey. 1993. Multidimensional students' evaluations of teaching effectiveness: A profile analysis. *The Journal of Higher Education* 64 (1): 1–18.
- Marsh, H.W., and L. Roche. 1993. The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal* 30 (1): 217–51.
- Murray, H.G. 1985. Classroom teaching behaviors related to college teaching effectiveness. *New Directions for Teaching and Learning* 23: 21–34.
- Menges, R.J., and K.T. Brinko. 1986. Effects of student evaluation feedback: A meta-analysis of higher education research. ERIC 270 408. Evanston, IL: Northwestern University.
- National Academy of Engineering 2007. *Engineering education centers. Roundtable members*. <http://www.nae.edu/nae/caseecomnew.nsf/weblinks/NFOY-5RUQPF?OpenDocument> (last accessed September, 2008).
- Piccinin, S. 1999. How individual consultation affects teaching. *New Directions for Teaching and Learning* 79: 71–83.
- Piccinin, S., C. Cristi, and M. McCoy. 1999. The impact of individual consultation on student ratings of teaching. *The International Journal for Academic Development* 4 (2): 75–88.
- Piccinin, S., and J. Moore. 2002. The impact of individual consultation on the teaching of younger versus older faculty. *The International Journal for Academic Development* 7 (2): 123–34.
- Professional and Organizational Development in Higher Education Network. 1995. *Celebrate Us!* [Brochure].
- Sorcinelli, M. 1991. Research findings on seven principles. *New Directions for Teaching and Learning* 47: 13–25.
- Taylor-Way, D., and K.T. Brinko. 1989. Using video recall for improving professional competency in instructional consultation. *To Improve the Academy* 8: 141–56.
- Weimer, M. 1990. *Improving college teaching*. San Francisco, CA: Jossey-Bass.
- Weston, C., and L. McAlpine. 1999. Toward an integrated approach to instructional consultation. *New Directions for Teaching and Learning* 79: 85–95.
- Wilson, R.C. 1986. Improving faculty teaching: Effective use of student evaluations and consultants. *Journal of Higher Education* 57: 196–211.

AUTHORS' BIOGRAPHIES

Cynthia J. Finelli is director for the Center for Research on Learning and Teaching (CRLT) North and associate research scientist of Engineering Education at University of Michigan (U-M). She holds a Ph.D. in electrical engineering from U-M. Prior to joining U-M in April 2003, Dr. Finelli spent ten years at Kettering University where she was the Richard L. Terrell Professor of Excellence in Teaching, founding director of the Center for Excellence in Teaching and Learning, and associate professor of electrical engineering. Her current research interests include evaluating methods to improve teaching, exploring ethical decision-making in engineering, and assessing the effect of the first year experience on under-represented student retention. She is chair of the Educational Research and Methods Division (ERM) of ASEE.

Address: CRLT North, 4117 Engineering Research Building I, 2200 Bonisteel Blvd, Ann Arbor, MI 48109-2099; telephone: (+1) 734.764.0244; fax: (+1) 734.647.3600; e-mail: cfinelli@umich.edu.

Molly Ott is a Ph.D. student in the Center for the Study of Higher and Postsecondary Education at U-M. She has studied a

number of topics related to faculty work life, including institutional commitment, allocation of effort, and motivation in regards to teaching, research, and service activities. Her research interests also include applying quantitative methods to various issues in higher education and the sociology of higher education.

Address: 2117 School of Education Building, 610 East University, Ann Arbor, MI 48109-1259; e-mail: mollyott@umich.edu.

Amy C. Gottfried is a lecturer in the Department of Chemistry at U-M. She holds a Ph.D. in chemistry from University of North Carolina at Chapel Hill. She has led the development and implementation of a studio-style general chemistry course, and she also teaches organic chemistry. Her research interests include understanding the manner in which students learn chemical concepts and preparing instructors at all levels of chemistry education.

Address: A868 Chemistry Building, 930 North University, Ann Arbor, MI 48109; telephone: (+1) 734.647.9540; fax: (+1) 734.615.3690; e-mail: acgottfr@umich.edu.

Chad Hershock is an assistant director at CRLT, and he is responsible for campus-wide teaching orientations, instructional consultations, workshops on teaching and learning, and preparing future faculty programs. He holds a Ph.D. in biology from U-M and has over 10 years of U-M teaching and course development experience in Ecology and Evolutionary Biology and Environmental Education. His current research interests include the role of teaching assistants on student retention in the sciences and engineering and

the impact of professional development activities on the instructional practice of teaching assistants.

Address: CRLT, 1071 Palmer Commons, 100 Washtenaw Avenue, Ann Arbor, MI 48109-2218; telephone: (+1) 734.615.8060; fax: (+1) 734.647.3600; e-mail: chersh@umich.edu.

Matthew L. Kaplan is managing director of CRLT where he focuses on external and university-wide initiatives. He holds a Ph.D. in comparative literature from University of North Carolina. He has served on the executive board of the Professional and Organizational Development Network in Higher Education (POD) and has edited two volumes of the POD journal, *To Improve the Academy*. He has written on teaching evaluation, multiculturalism, and the use of interactive theater for faculty development.

Address: CRLT, 1071 Palmer Commons, 100 Washtenaw Avenue, Ann Arbor, MI 48109-2218; telephone: (+1) 734.936.0644; fax: (+1) 734.647.3600; e-mail: mlkaplan@umich.edu.

Christopher M. O'Neal is an assistant director at CRLT, and he holds a Ph.D. in ecology from U-M. His current research interests include the impact of early evaluation on teacher performance, the value of interactive theatre for building multicultural competencies in educators, and the impact of teaching assistants on student retention in the sciences and engineering.

Address: CRLT, 1071 Palmer Commons, 100 Washtenaw Avenue, Ann Arbor, MI 48109-2218; telephone: (+1) 734.763.4418; fax: (+1) 734.647.3600; e-mail: coneal@umich.edu.

APPENDIX A. SAMPLE STUDENT RATINGS OF TEACHING SURVEY

This version of the midterm student ratings of teaching survey was administered in *Groups 1a, 3a, and 4a* at midterm. Similar surveys were administered in all participating courses at midterm and the end of the term.



The University of Michigan
Office of Evaluations & Examinations
Teaching Questionnaire



Evaluating Methods for Improving Teaching in the College of Engineering **Center for Research on Learning and Teaching North and College of Engineering**

The Center for Research on Learning and Teaching North (CRLT) and the College of Engineering are conducting a research study to collect data pertaining to methods of improving teaching in the College of Engineering. One aspect of this study includes collecting student ratings of teaching at two points in time: at midterm and at the end of the term. This is the end-of-term student rating of teaching survey.

Your participation is completely voluntary. You may excuse yourself from participating in this survey, in part or as a whole, at any time during the administration of the survey. Any survey questions may be skipped if you do not feel comfortable completing them.

For the purposes of research, no individual's comments will be identified by name, gender, age, ethnicity/race or other combination of characteristics that would reveal his/her identity in this survey. All information collected will remain confidential.

The data collected through this survey will be given to your course instructor in summary form only; your responses will be completely anonymous.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Overall, this was an excellent course.	5	4	3	2	1
Overall, the instructor was an excellent teacher.	5	4	3	2	1
I learned a great deal in this course.	5	4	3	2	1
I had a strong desire to take this course.	5	4	3	2	1
The instructor taught in a manner that served my needs as a student.	5	4	3	2	1
The instructor gave clear explanations.	5	4	3	2	1
The instructor was enthusiastic.	5	4	3	2	1
The instructor acknowledged all questions insofar as possible.	5	4	3	2	1
The instructor treated students with respect.	5	4	3	2	1
The instructor was willing to meet and help students outside of class.	5	4	3	2	1
The instructor kept students informed of their progress.	5	4	3	2	1
The instructor used class time well.	5	4	3	2	1
The instructor seemed well prepared for each class.	5	4	3	2	1
Work requirements and grading system were clear from the beginning.	5	4	3	2	1
The amount of work required was appropriate for the credit received.	5	4	3	2	1
The instructor set high standards for students.	5	4	3	2	1
The instructor used techniques that fostered class participation.	5	4	3	2	1

APPENDIX B. CONSULTATION PROTOCOL FOR GROUP 3B.

Note: At the University of Michigan, the term Midterm Student Feedback (MSF) is used in place of the term Small Group Instructional Diagnosis (SGID).

Pre-class meeting (20-30 minutes). Conducted one to five days prior to the designated class session.

- 1) Complete the *pre-class meeting form* (or take similar notes).
- 2) Describe the Midterm Student Feedback (MSF) process to the instructor (be clear that the instructor will need to leave the room approximately 25 minutes before the scheduled end of the class) and show the instructor the MSF form you will ask students to complete. Also, explain the way in which the instructor should introduce you to the class.
- 3) Confirm the time and location of the MSF class session and schedule a follow-up consultation with the instructor.

Class session

Observation

- 1) Arrive at least five minutes before class is scheduled to begin and sit at the back of the class.
- 2) The instructor should introduce the consultant and explain that the consultant will work with the students during the last 25 minutes of class.
- 3) During the class session, quietly observe and record data (timeline, questions, special events, class map showing interaction, and other evidence that may be useful during consultation).

Midterm Student Feedback session

- 1) Introduce the MSF to the students. Establish the confidentiality of the process, indicate that the instructor invited the consultant into the class, and mention that CRLT has a long history of conducting MSFs.
- 2) Describe the process: students will work in small groups first, then the consultant will collect information from the entire class.
- 3) Distribute the MSF forms, ask students to form groups of 4-8 students each, and allow them approximately ten minutes to respond to the questions. Instruct students to ensure that everybody contributes. When about three minutes remain, give the students a "heads-up."
- 4) On the overhead, write your e-mail, indicating that students should feel free to contact you if there are issues that they prefer to refer privately.
- 5) Call on group spokespersons to indicate one strength per group, and probe for details as to why it is a strength (e.g., Student: "she has a good pace," Consultant: "and that is a strength because...?"). Keep wording as close to students' responses as possible and monitor the overall level of agreement among the rest of the class. Move onto the next group, discussing another strength. Continue until all strengths have been listed.
- 6) Repeat for areas for improvement.
- 7) Thank the students, collect the MSF forms, and gather your transparencies.

MSF report for instructor

- 1) If it is necessary, transcribe your notes so they are legible (the notes will be collected for the research project).
- 2) Prepare an acknowledgement letter without MSF details using CRLT North letterhead, and use it as a cover sheet for the MSF report.
- 3) Create a one or two page typed report that includes a summary of the process (date of class session, name and number of course, number of students involved, etc), as well as a list of strengths mentioned by students and a list of suggestions for improvement. If it is appropriate, include verbatim students' comments. Refrain from including evaluative comments from your observations, as the purpose of the MSF is to focus on students' feedback.
- 4) Make three copies of the report—one for the instructor, one for your own records, and one for the research team.

Follow up visit (30-45 minutes). Conducted as soon after the designated class session as possible.

- 1) Thank the instructor for allowing you to visit the class. Give the instructor a copy of the MSF report, and reiterate that the report is completely confidential.
- 2) Go through list of strengths (one item at a time) and discuss the instructor's impressions. Once all the points have been discussed, ask, "Is there anything surprising in the list?" "Is there anything that you are glad to hear?"
- 3) Go through the list of student suggestions for improvement, making sure the instructor understands each item and has time to reflect on it. Then, encourage the instructor to think about strategies to address each item (e.g., do nothing, make some changes this term, or make some changes in the future). If it is possible, let the *instructor* develop the strategies.
- 4) Use your observation notes only to give additional data on what is being discussed. If you observe something that was not mentioned in the feedback but that you consider deserving of *immediate* attention, offer the data you have and ask the instructor to reflect on the information.
- 5) Finish the consultation by recapping and guiding the instructor about what to do the next time class meets (e.g., thank the students for feedback, discuss what the instructor will do differently or why it is not realistic to do something differently). Remind the instructor that ignoring the session is a bad idea.
- 6) Give the report of the midterm student ratings of teaching surveys to the instructor (try not to discuss the report).

Consultation report for the research project

- 1) Prepare a legible set of notes describing the entire process.
- 2) Include the pre-class meeting form.
- 3) Include your class observation notes, the original student MSF sheets, and the transparencies you created during the class session.
- 4) Include a copy of the MSF report.
- 5) Highlight any interesting comments made by the instructor during the follow-up consultation. Also, list any items from your personal observation that you discuss during the follow-up consultation, and describe your overall impressions of the consultation.

APPENDIX C. STUDENT RATINGS OF TEACHING FOR EACH INTERVENTION GROUP

	<i>Group 1a</i> No consult, no ratings N = 7	<i>Group 1b</i> No consult, ratings only N = 11	<i>Group 2</i> Consult on ratings N = 9	<i>Group 3a</i> SGID, no ratings N = 7	<i>Group 3b</i> SGID, with ratings N = 5	<i>Group 4a</i> Videotape, no ratings N = 7	<i>Group 4b</i> Videotape, with ratings N = 9
1. Overall this was an excellent course.	(3.28, 4.35) (3.42, 4.25)	(3.64, 4.32) (3.63, 4.28)	(3.32, 4.05) (3.49, 4.06)	(3.64, 4.25) (3.60, 4.11)	(3.69, 4.62) (3.31, 4.36)	(3.39, 4.40) (3.46, 4.25)	(3.23, 4.17) (3.49, 4.06)
2. Overall, the instructor was an excellent teacher.	(3.64, 4.78) (3.69, 4.62)	(3.76, 4.54) (3.76, 4.54)	(3.50, 4.15) (3.50, 4.31)	(3.71, 4.61) (3.62, 4.32)	(3.43, 4.88) (3.53, 4.66)	(3.55, 4.73) (3.64, 4.64)	(3.13, 4.29) (3.35, 4.20)
3. I learned a great deal in this course.	(3.29, 4.47) (3.60, 4.40)	(3.91, 4.41) (3.77, 4.29)	(3.65, 4.18) (3.48, 4.09)	(3.74, 4.39) (3.59, 4.19)	(3.94, 4.60) (3.80, 4.21)	(3.74, 4.44) (3.68, 4.34)	(3.72, 4.14) (3.68, 3.90)
4. I had a strong desire to take this course.	(3.15, 4.20) (3.09, 4.08)	(3.30, 4.13) (3.35, 4.10)	(3.15, 4.17) (3.35, 4.14)	(3.32, 4.14) (3.35, 4.10)	(3.24, 4.48) (2.92, 4.26)	(2.80, 4.13) (2.73, 4.03)	(3.20, 4.21) (3.17, 4.22)
5. The instructor taught in a way that served students' needs.	(3.63, 4.21) (3.57, 4.27)	(3.39, 4.27) (3.46, 4.30)	(3.01, 3.91) (3.17, 4.10)	(3.69, 4.27) (3.55, 4.20)	(3.32, 4.58) (3.40, 4.35)	(3.47, 4.37) (3.64, 4.37)	(3.06, 4.14) (3.37, 4.00)
6. The instructor gave clear explanations.	(3.76, 4.41) (3.76, 4.35)	(3.44, 4.42) (3.49, 4.37)	(3.29, 4.06) (3.34, 4.19)	(3.65, 4.45) (3.60, 4.26)	(3.37, 4.48) (3.61, 4.36)	(3.73, 4.51) (3.69, 4.42)	(3.27, 4.25) (3.40, 4.08)
7. The instructor was enthusiastic.	(3.93, 4.84) (3.94, 4.85)	(4.16, 4.68) (4.04, 4.59)	(4.08, 4.68) (3.85, 4.76)	(3.72, 4.88) (3.58, 4.56)	(4.13, 5.07) (3.91, 4.89)	(4.06, 4.76) (3.97, 4.67)	(3.96, 4.71) (4.10, 4.60)
8. The instructor acknowledged all questions insofar as possible.	(4.17, 4.74) (4.15, 4.63)	(4.15, 4.66) (4.07, 4.60)	(4.12, 4.34) (4.11, 4.49)	(4.16, 4.57) (4.11, 4.52)	(4.12, 4.67) (4.08, 4.64)	(4.17, 4.62) (4.26, 4.71)	(3.96, 4.49) (4.09, 4.41)
9. The instructor treated students with respect.	(4.20, 4.85) (4.20, 4.73)	(4.16, 4.76) (4.22, 4.72)	(4.26, 4.56) (4.26, 4.62)	(4.31, 4.76) (4.16, 4.55)	(4.21, 4.81) (4.34, 4.81)	(4.22, 4.69) (4.28, 4.70)	(4.03, 4.51) (4.21, 4.51)
10. The instructor was willing to meet and help students outside of class.	(3.72, 4.71) (3.77, 4.71)	(4.05, 4.54) (3.98, 4.46)	(3.81, 4.26) (3.81, 4.15)	(3.81, 4.53) (3.73, 4.35)	(3.97, 4.34) (3.87, 4.44)	(3.72, 4.57) (3.98, 4.42)	(3.77, 4.36) (3.73, 4.02)
11. The instructor kept students informed of their progress.	(3.32, 4.14) (3.41, 3.98)	(3.68, 4.23) (3.45, 4.05)	(2.86, 3.83) (3.10, 3.64)	(3.54, 4.04) (3.10, 3.70)	(2.93, 4.17) (2.82, 4.11)	(3.18, 4.25) (3.07, 3.73)	(3.11, 4.20) (3.15, 3.58)
12. The instructor used class time well.	(3.46, 4.35) (3.56, 4.53)	(3.75, 4.41) (3.79, 4.36)	(3.49, 4.05) (3.52, 4.15)	(3.84, 4.59) (4.08, 4.38)	(3.37, 4.60) (3.47, 4.41)	(3.80, 4.50) (4.03, 4.49)	(3.27, 4.25) (3.61, 4.10)
13. The instructor seemed well prepared for each class.	(3.96, 4.49) (4.10, 4.60)	(4.24, 4.60) (4.25, 4.62)	(3.95, 4.38) (4.14, 4.60)	(4.31, 4.82) (4.11, 4.66)	(3.56, 4.76) (3.85, 4.86)	(4.12, 4.65) (4.26, 4.77)	(3.72, 4.48) (3.95, 4.42)
14. Work requirements and grading system were clear from the beginning.	(3.61, 4.43) (3.84, 4.23)	(3.89, 4.46) (4.03, 4.41)	(3.62, 4.21) (3.74, 4.19)	(3.97, 4.32) (3.90, 4.22)	(3.48, 4.30) (3.77, 4.35)	(3.85, 4.34) (3.94, 4.28)	(3.12, 4.36) (3.92, 4.24)
15. The amount of work required was appropriate for the credit received.	(3.33, 4.22) (3.50, 4.36)	(3.79, 4.27) (3.67, 4.32)	(3.45, 3.97) (3.36, 4.16)	(3.84, 4.16) (3.80, 4.13)	(3.42, 4.29) (3.13, 4.41)	(3.70, 4.35) (3.73, 4.16)	(2.97, 4.27) (3.74, 4.06)
16. The instructor set high standards for students.	(3.64, 4.36) (3.71, 4.32)	(3.81, 4.39) (3.84, 4.17)	(3.91, 4.47) (3.77, 4.25)	(4.00, 4.41) (3.90, 4.07)	(3.99, 4.57) (3.73, 4.41)	(3.78, 4.52) (3.65, 4.50)	(3.88, 4.23) (3.87, 4.17)
17. The instructor used techniques that fostered class participation.	(3.09, 4.15) (3.20, 4.15)	(3.02, 3.92) (2.98, 3.81)	(3.22, 3.90) (2.97, 3.87)	(3.16, 4.30) (2.75, 3.90)	(3.43, 4.66) (3.21, 4.65)	(3.40, 4.36) (3.06, 4.30)	(3.27, 4.15) (3.30, 4.16)

Table C1. Midterm and end-of-term student ratings for each item by intervention group. The first row listed for each item is the 95 percent confidence interval for the average midterm rating, and the second row is the 95 percent confidence interval for the average end-of-term rating. Student responses were on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

	<i>Group 1a</i> No consult, no ratings <i>N</i> = 7	<i>Group 1b</i> No consult, ratings only <i>N</i> = 11	<i>Group 2</i> Consult on ratings <i>N</i> = 9	<i>Group 3a</i> SGID, no ratings <i>N</i> = 7	<i>Group 3b</i> SGID, with ratings <i>N</i> = 5	<i>Group 4a</i> Videotape, no ratings <i>N</i> = 7	<i>Group 4b</i> Videotape, with ratings <i>N</i> = 9
1. Overall this was an excellent course.	0.02 (-0.11, 0.16)	0.01 (-0.14, 0.16)	-0.06 (-0.16, 0.03)	0.08 (0.00, 0.17)	0.27 (0.05, 0.49)	0.07 (-0.12, 0.26)	-0.03 (-0.23, 0.17)
2. Overall, the instructor was an excellent teacher.	0.17 (-0.13, 0.48)	-0.10 (-0.42, 0.22)	-0.14 (-0.37, 0.08)	0.22 (0.03, 0.40)	-0.16 (-1.58, 1.27)	0.03 (-0.20, 0.25)	0.03 (-0.14, 0.20)
3. I learned a great deal in this course.	-0.08 (-0.30, 0.14)	0.12 (-0.01, 0.25)	0.09 (-0.08, 0.25)	0.16 (0.07, 0.26)	0.25 (-0.12, 0.62)	0.08 (-0.02, 0.19)	0.12 (-0.02, 0.26)
4. I had a strong desire to take this course.	0.04 (-0.16, 0.25)	0.01 (-0.11, 0.13)	-0.05 (-0.22, 0.13)	-0.02 (-0.21, 0.17)	0.20 (0.06, 0.34)	0.07 (0.03, 0.11)	-0.04 (-0.16, 0.08)
5. The instructor taught in a way that served students' needs.	-0.02 (-0.15, 0.11)	-0.14 (-0.50, 0.23)	-0.16 (-0.36, 0.04)	0.09 (-0.02, 0.20)	0.09 (-0.21, 0.39)	-0.10 (-0.30, 0.10)	0.00 (-0.21, 0.20)
6. The instructor gave clear explanations.	0.05 (-0.08, 0.17)	0.01 (-0.23, 0.24)	-0.12 (-0.34, 0.10)	0.10 (-0.18, 0.38)	-0.06 (-0.55, 0.43)	0.07 (-0.07, 0.21)	0.08 (-0.14, 0.29)
7. The instructor was enthusiastic.	-0.02 (-0.53, 0.48)	0.20 (-0.04, 0.43)	-0.21 (-0.74, 0.31)	0.31 (0.10, 0.52)	0.30 (-0.10, 0.70)	0.10 (-0.27, 0.47)	0.06 (-0.21, 0.33)
8. The instructor acknowledged all questions insofar as possible.	0.12 (-0.20, 0.45)	0.16 (-0.11, 0.42)	-0.24 (-0.66, 0.18)	-0.03 (-0.58, 0.52)	0.01 (-0.38, 0.40)	-0.21 (-0.40, -0.02)	0.00 (-0.23, 0.22)
9. The instructor treated students with respect.	0.23 (-0.23, 0.68)	0.00* (-0.27, 0.27)	-0.22 (-0.63, 0.20)	0.30 (0.12, 0.48)	-0.53 (-2.44, 1.38)	-0.09 (-0.25, 0.08)	-0.16 (-0.53, 0.21)
10. The instructor was willing to meet and help students outside of class.	-0.04* (-0.34, 0.26)	0.08 (-0.10, 0.25)	0.05 (-0.12, 0.21)	0.11 (-0.24, 0.46)	-0.04 (-0.37, 0.28)	-0.01 (-0.31, 0.28)	0.15 (-0.12, 0.42)
11. The instructor kept students informed of their progress.	0.03 (-0.25, 0.31)	0.09 (-0.17, 0.36)	-0.07 (-0.49, 0.34)	0.20 (-0.05, 0.46)	0.01 (-0.50, 0.52)	0.22 (-0.01, 0.45)	0.21 (-0.05, 0.48)
12. The instructor used class time well.	-0.45 (-1.33, 0.42)	0.03 (-0.10, 0.15)	-0.11 (-0.34, 0.12)	0.02 (-0.36, 0.39)	0.06 (-0.26, 0.37)	-0.12 (-0.27, 0.04)	-0.02 (-0.28, 0.23)
13. The instructor seemed well prepared for each class.	-0.22 (-0.63, 0.18)	-0.10 (-0.44, 0.23)	-0.58 (-1.21, 0.05)	0.25 (-0.05, 0.55)	-0.84 (-3.23, 1.56)	-0.39 (-0.74, -0.04)	-0.07 (-0.36, 0.22)
14. Work requirements and grading system were clear from the beginning.	0.02 (-0.25, 0.28)	-0.08 (-0.35, 0.18)	-0.03 (-0.16, 0.09)	0.08 (-0.13, 0.28)	-0.25 (-0.90, 0.41)	0.00 (-0.18, 0.18)	-0.31 (-0.81, 0.19)
15. The amount of work required was appropriate for the credit received.	-0.22 (-0.55, 0.10)	-0.40 (-1.37, 0.57)	-0.10 (-0.33, 0.13)	0.01 (-0.19, 0.21)	0.03 (-0.20, 0.25)	0.10 (-0.04, 0.24)	-0.19 (-0.61, 0.24)
16. The instructor set high standards for students.	-0.01 (-0.12, 0.10)	0.12 (-0.10, 0.33)	0.21 (0.10, 0.32)	0.22 (0.08, 0.37)	0.24 (0.17, 0.30)	0.01 (-0.20, 0.23)	0.02 (-0.15, 0.19)
17. The instructor used techniques that fostered class participation.	-0.04 (-0.21, 0.14)	0.04 (-0.10, 0.17)	0.02 (-0.20, 0.25)	0.26 (0.14, 0.37)	-0.05 (-0.61, 0.52)	0.06 (-0.28, 0.39)	-0.06 (-0.33, 0.21)
Number of items with positive gain	8	11	4	15	10	10	7
Number of items with significant positive or negative gain	0	0	1	6	3	3	0
Number of items with significant positive gain	0	0	1	6	3	1	0

* For these entries, *N* is one fewer than listed at the top of the column because the average midterm rating for one course was 5.0, and this course was excluded from calculation of the average normalized gain to avoid a "divide by zero" error.

Table C2. Average normalized gain score for each item by intervention group. The top entry for each item is the average normalized gain, and the bottom row is the 95% confidence interval. Items for which the average normalized gain is statistically significantly different from zero are shaded. Student responses were on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).