2006-882: EVALUATING METHODS TO IMPROVE TEACHING IN ENGINEERING

Cynthia Finelli, University of Michigan

Dr. Cynthia J. Finelli (cfinelli@umich.edu)is Managing Director of the Center for Research on Learning and Teaching (CRLT) North and Associate Research Scientist of Engineering Education at University of Michigan (U-M). Her current research interests include evaluating methods to improve teaching, exploring ethical decision-making in engineering, developing a tool for comprehensive assessment of team-member effectiveness, and assessing the effect of the first year experience on under-represented student retention. She serves on the Executive Board of the Educational Research and Methods Division (ERM) of ASEE and was the ERM Division Program Co-Chair for the 2003 Frontiers in Education Conference and the 2006 ASEE Annual Conference and Exposition.

Amy Gottfried, University of Michigan

Dr. Amy C. Gottfried is the Chemical Sciences at the Interface of Education Postdoctoral Fellow in the Department of Chemistry at U-M as well as an instructional consultant for CRLT. She has led the development and implementation of a studio general chemistry course. Her research interests include understanding the manner in which students learn chemical concepts and preparing instructors at all levels of chemistry education.

Matthew Kaplan, University of Michigan

Dr. Matthew L. Kaplan is the Associate Director of CRLT at U-M where he focuses on external and university-wide initiatives. 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 issues of teaching evaluation, multiculturalism, and the use of interactive theatre for faculty development.

Vilma Mesa, University of Michigan

Dr. Vilma M. Mesa is Assistant Professor and Assistant Research Scientist of Mathematics Education at U-M's School of Education. Her research interests include undergraduate mathematics teaching, curriculum theory and evaluation in mathematics, and international comparisons of achievement.

Christopher O'Neal, University of Michigan

Dr. Christopher M. O'Neal is Coordinator of STEM Faculty Development at CRLT at U-M. His Ph.D. is in Ecology, and 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 TAs on student retention in the sciences and engineering.

Mary Piontek, University of Michigan

Dr. Mary E. Piontek is Assistant Research Scientist at CRLT at U-M where she works with individual faculty, departments/units, and schools/colleges that need assistance designing program evaluation and assessing the effectiveness of initiatives to improve teaching and learning. Her research and evaluation techniques capture the local context of an organization or program through individual and focus group interviews, in-depth participant observation, document and archival analysis, survey research, and qualitative/quantitative mixed designs. Her research interests include the changing roles of evaluators and their client/stakeholder relationships.

Evaluating Methods to Improve Teaching in Engineering

Abstract

Engineering faculty at a large research institution participated in a project for evaluating methods to improve teaching. Faculty were randomly assigned to one of four separate cohorts (each receiving a different type of feedback designed to improve teaching) and comparative data was collected on each of the four methods. Faculty in *Cohort 0: Control* served as the control population and did not receive formal feedback of any kind to improve teaching. Faculty in *Cohort 1: Ratings Report* received a report summarizing student ratings of teaching at midterm. For faculty in *Cohort 2: Feedback and Consult*, an instructional consultant facilitated a student feedback session at midterm (also known as a small group instructional diagnosis) and then conducted a follow-up consultation with the faculty member. An instructional consultant videotaped a class period for faculty in *Cohort 3: Videotape and Consult* and conducted a follow-up consultation.

To compare the four methods to improve teaching, data from three separate sources was analyzed. First, student ratings of teaching were collected in the middle of the academic term and again at the end of the term. The ratings were studied and the change in average ratings from the middle to the end of the term was compared to assess the level of teaching improvement. Second, all faculty completed an online survey to assess the method to improve teaching they completed, to rate their own teaching at the end of the term, and to describe their perceptions of the project. Faculty responses were analyzed and compared by cohort. Finally, a focus group for the instructional consultants was conducted to gauge their perceptions of each method, to ascertain the nature of the consultations, and to identify kinds of issues that arose in each consultation[i].

From this limited study, it appears that the student feedback and follow-up consultation may have the most positive impact on student ratings of teaching. However, having a class session videotaped and then having a follow-up consultation is also a promising method to improve teaching. Further work to study these methods more clearly is underway.

1. Experimental Design

Faculty teaching full-term, undergraduate, lecture courses in all engineering departments were invited to participate in the project[ii]. Those who participated were asked to follow a specific protocol for gathering feedback to improve teaching. Then, to evaluate teaching improvement, data was collected and analyzed from three separate sources. Both the protocols for gathering feedback to improve teaching and the methods for evaluating teaching improvement are described in this section.

1.1.   Methods to improve teaching

After faculty recruiting was complete, participants were randomly assigned to one of four cohorts (*Cohort 0: Control*; *Cohort 1: Ratings Report*; *Cohort 2: Feedback and Consult*; and *Cohort 3: Videotape and Consult*). Depending on their cohort assignment, faculty were asked to follow a specific protocol for gathering feedback to improve teaching (described in Table 1).

Table 1. Description of teaching improvement protocol for each cohort.

	Middle of term				End of term
	Student ratings survey	Report on student ratings	Student feedback session + consult	Videotape of class session + consult	Student ratings survey
Cohort 0: Control	✓				✓
Cohort 1: Ratings Report	✓	✓			✓
Cohort 2: Feedback and Consult	✓		✓		✓
Cohort 3: Videotape and Consult	✓			✓	✓

For faculty in all cohorts, student ratings of teaching were collected at the middle of the term and again at the end of the term for comparative purposes (the surveys are described later). A report containing details about the midterm and end-of-term student ratings was provided to all faculty after the term concluded.

The faculty in *Cohort 0: Control* served as a control group. Students in these classes completed the student ratings of teaching survey at midterm and again at the end of the term, and the faculty received the report summarizing student responses after the term was complete. These faculty received no intervention during the term and no additional summary reports during the term. To reiterate, these faculty did not receive results of the midterm student ratings of teaching survey until after the term.

For *Cohort 1: Ratings Report*, faculty received an individual report summarizing results from the midterm administration of the student ratings of teaching survey within one week of the survey. These faculty also received the summary report after the term concluded.

For *Cohort 2: Feedback and Consult*, an instructional consultant observed the first half of one class session during the middle of the term and collected feedback from students during the second half of the same class session (after the instructor left the room). The consultant then prepared a report about the observation and feedback session and met with the faculty member to discuss the report. This type of feedback, known as midterm student feedback or small group instructional diagnosis, is regularly implemented for purposes of teaching improvement through centers for teaching and learning (e.g., http://www.crlt.umich.edu/crlttext/feedbacktext.html), and it has been shown to be an effective catalyst for change in higher education classes[4]. The faculty member received details about the midterm and end-of-term student ratings of teaching survey only after the term was over.

For *Cohort 3: Videotape and Consult*, an instructional consultant videotaped one entire class session[iii] during the middle of the term, and then met with the faculty member to discuss the class session and review the videotape. The consultant showed portions of the videotape (for stimulated recall) and encouraged the faculty member to reflect upon the experience, a protocol

that has also been shown to be effective in improving teaching[11]. As with *Cohorts 0 and 2*, faculty in this cohort received details about the midterm and end-of-term student ratings only after the term was over.

1.2.   Approaches for evaluating teaching improvement

To compare and evaluate the methods to improve teaching, data analysis from three separate sources was conducted. These include student ratings of teaching surveys administered at midterm and again at the end of the term, an online faculty survey, and focus group discussions with the consultants who conducting the consultations.

First, to assess systematically the level of teaching improvement and to compare cohort groups, student ratings of teaching surveys were administered in every class at midterm and again at the end of the term (the surveys were administered by someone unaffiliated with the course). The surveys contained seventeen questions (Q1 through Q17) and were based on the *Seven Principles of Good Practice in Undergraduate Education*[3, 10], Murray's "effective classroom teaching behaviors"[8], and themes apparent from twenty-nine midterm student feedback sessions conducted for engineering faculty by one of the authors. Further, the questions represented traits that could potentially be changed over the short duration of six to eight weeks. A sample survey is included in Appendix 1. Consistent with the protocols to improve teaching, all faculty received a report summarizing and comparing responses on the midterm and end-of-term survey after the term was finished. Additionally, faculty in *Cohort 1: Ratings Report* (but not other faculty) received a summary of responses for their class in the middle of the term after the midterm survey.

Second, to assess faculty perceptions of each method to improve teaching, a short, four-part faculty survey was conducted online at the end of the term (after the end-of-term student surveys were collected). First, faculty were asked to indicate the degree to which they engaged in several activities related to teaching improvement, including whether or not they had previously had a teaching consultation. Next, to compare student ratings of teaching with faculty self-ratings, faculty were asked to rate themselves on thirteen of the same aspects that the students rated. Then, the faculty who received an intervention (report on midterm student ratings of teaching for *Cohort 1: Ratings Report*, student feedback session and follow-up consultation for *Cohort 2: Feedback and Consult*, and videotaped class session and follow-up consultation for *Cohort 3: Videotape and Consult*) were asked to rate the intervention, to describe the pros and cons of the intervention, and to discuss any changes they implemented as a result of the intervention. Finally, all faculty were given the opportunity to describe anything else they would like the research team to know about their experience with the project. Questions from the faculty survey are included in Appendix 2.

The third source of data was a discussion in which the instructional consultants talked about the project and shared their perceptions of each method. Four instructional consultants–all with backgrounds in science, math, or engineering–participated in the project, and each conducted at least one consultation pertaining to a student feedback session and at least one video consultation (for a total of three or four consultations per consultant). The consultants' perspectives provided valuable insight for the project.

2. Results

To evaluate the methods to improve teaching, a variety of analyses were conducted. First, to confirm that the random assignment of faculty into cohort groups resulted in similar groups, several faculty characteristics were compared across cohort. Second, responses from the midterm student ratings of teaching survey were compared across cohort, as were responses from the end-of-term student ratings of teaching survey and the change in response. Third, faculty responses to the online survey (including self-ratings of teaching and perceptions about the intervention) were compared. Finally, a focus group was conducted with the consultants to evaluate further the methods to improve teaching. Results form all of these analyses are described in this section.

2.1. Similarity of cohorts

The assignment of faculty into cohorts was purely random, and several comparisons of the cohorts indicate that they are similar in composition. In particular, the groups are comparable by demographics, by teaching ability (according to midterm student ratings of teaching), and in terms of prior level of engagement in activities related to teaching improvement.

Twenty-nine eligible faculty volunteered to participate, and 28 completed the project[iv]. The 28 participants represent all faculty ranks (three lecturers, one adjunct professor, eight assistant professors, six associate professors, and ten full professors) and ten separate departments in the College of Engineering. The participants' courses span all class levels (three 100-level, five 200-level, ten 300-level, and eleven 400-level), are assigned a range of credits (one 2-credit, seven 3-credit, and twenty 4-credit courses), and encompass a broad class size (average $= 52.6 \pm 37.7$ students). Four women faculty were among the 28 participants in the study.

Seven faculty were randomly assigned to each cohort, and Table 2 describes some basic faculty demographics by cohort. There was a broad distribution in each cohort by faculty rank, and there was one female in each cohort. Demographically, the cohorts are comparable.

Table 2. Faculty demographics by cohort.

	Gender		Faculty Rank				
	Male	Female	Lecturer	Adjunct Professor	Assistant Professor	Associate Professor	Full Professor
Cohort 0	6	1	0	0	3	4	0
Cohort 1	6	1	1	1	1	0	4
Cohort 2	6	1	0	0	2	2	3
Cohort 3	6	1	2	0	2	0	3

At midterm, 1015 individual student surveys were completed, and a statistical comparison of data from all 17 questions on the survey indicate that the four cohorts were also similar according to these ratings at midterm. For every one of the 17 questions, ratings from all six possible pairs of cohorts (Cohort 0 versus Cohort 1, Cohort 0 versus Cohort 2, Cohort 0 versus Cohort 3, Cohort 1 versus Cohort 2, Cohort 1 versus Cohort 3, and Cohort 2 versus Cohort 3) were studied, and statistically significant differences were identified (d.f.=6, $p<.05$, independent samples t-test). Of the 102 possible differences, there were only two significant differences, both

of which were in regards to Question 16 ("The instructor kept students informed of their progress."). For Cohort 1, the average student rating (3.94 ± 0.44) was significantly different from the average rating of both Cohort 2 (3.40 ± 0.33) and Cohort 3 (3.40 ± 0.36).

In addition, as part of the online survey administered at the end of the project, faculty were asked to indicate the degree to which they engaged in six activities related to teaching improvement and to indicate whether or not they had previously had a teaching consultation. Results are presented in Table 3. Although the number of faculty who did have a teaching consultation of some sort prior to this project is very different between cohorts, other baseline data is relatively stable across cohort.

Table 3. Faculty self-reported level of engagement in activities related to teaching improvement.
(Scale for questions a through f: 1=never, 2=occasionally, 3=often.)

	Cohort 0: Control	Cohort 1: Ratings Report	Cohort 2: Feedback and Consult	Cohort 3: Videotape and Consult
How often do you…				
a. Talk to others about your teaching	2.3	1.8	2.3	2.3
b. Solicit feedback from students about how they feel class is going	2.6	2.2	2.6	2.4
c. Spend time revising the content of your courses	3.0	2.8	2.4	2.6
d. Spend time revising the methods that you use to teach your courses	2.4	2.3	2.1	2.3
e. Attend programs on effective teaching	1.6	1.7	1.4	2.1
f. Read journals or books on effective teaching	1.7	2.0	1.4	1.7
Have you had a teaching consultation of any sort before this term?	4 yes 3 no	0 yes 7 no	4 yes 3 no	2 yes 5 no

2.2. Student ratings of teaching survey

Student ratings of teaching were collected for 17 course-related aspects at midterm and again at the end of the term. At midterm, 1015 individual student surveys were completed, and 952 were collected at the end of the term. To assess the level of teaching improvement and to understand better the differences between cohorts, the change from midterm to end-of-term in average student response for each question was computed. (Because of the limited data set, one must be careful to note *general trends* only rather than to draw conclusions.) Table 4 shows the average and standard deviation of the average change of all 17 questions for each cohort. The table also identifies changes that are statistically significant (d.f.=6, $p<0.05$, paired samples t-test).

Student ratings of teaching trends vary in the number of rating increases from midterm to end-of-term for each cohort. *Cohort 0: Control* exhibited an average increase on only six of 17 questions, and for *Cohort 1: Ratings Report*, there was an increase for seven of 17 questions. There was an increase from midterm to end-of-term in average student responses for 15 of 17

questions for *Cohort 2: Feedback and Consult* (Q4, desire to take the course, had no average change, and Q7 dropped slightly), and *Cohort 3: Videotape and Consult* had an increase on ten of 17 questions.

Table 4. Average change from midterm to end-of-term average ratings for each cohort.
Values listed represent the average of the differences between end-of-term and midterm ratings for each cohort. (Scale for responses: 1=strongly disagree, 2=agree, 3=neutral, 4=agree, 5=strongly agree.)

		Cohort 0: Control	Cohort 1: Ratings Report	Cohort 2: Feedback and Consult	Cohort 3: Videotape and Consult
Q1.	Overall this was an excellent course.	−0.02 ± 0.16	−0.05 ± 0.07	0.09 ± 0.10	0.04 ± 0.21
Q2.	Overall, the instructor was an excellent teacher.	0.06 ± 0.19	−0.10 ± 0.12	0.19 ± 0.17*	0.01 ± 0.29
Q3.	I learned a great deal in this course.	−0.13 ± 0.24	0.05 ± 0.05	0.17 ± 0.12*	0.08 ± 0.13
Q4.	I had a strong desire to take this course.	0.09 ± 0.34	−0.09 ± 0.17	0.00 ± 0.27	0.09 ± 0.03*
Q5.	The instructor gave clear explanations.	0.03 ± 0.11	−0.07 ± 0.21	0.12 ± 0.39	0.06 ± 0.14
Q6.	The instructor acknowledged all questions insofar as possible.	0.07 ± 0.18	0.05 ± 0.16	0.05 ± 0.27	−0.09 ± 0.07*
Q7.	The instructor used class time well.	−0.14 ± 0.18	−0.07 ± 0.10	−0.02 ± 0.31	−0.11 ± 0.15
Q8.	The instructor seemed well prepared for each class.	−0.12 ± 0.19	−0.03 ± 0.12	0.18 ± 0.28	−0.13 ± 0.14*
Q9.	Work requirements and grading system were clear from the beginning.	−0.01 ± 0.30	−0.11 ± 0.27	0.08 ± 0.20	−0.01 ± 0.16
Q10.	The amount of work required was appropriate for the credit received.	−0.16 ± 0.25	−0.13 ± 0.12*	0.03 ± 0.24	0.08 ± 0.14
Q11.	The instructor used techniques that fostered class participation.	−0.05 ± 0.22	0.08 ± 0.23	0.41 ± 0.22*	0.20 ± 0.29
Q12.	The instructor treated students with respect.	0.06 ± 0.20	−0.04 ± 0.14	0.18 ± 0.12*	−0.04 ± 0.08
Q13.	The instructor taught in a way that served students' needs.	0.00 ± 0.18	−0.17 ± 0.20	0.10 ± 0.14	−0.09 ± 0.25
Q14.	The instructor was willing to meet and help students outside of class.	−0.02 ± 0.23	0.07 ± 0.12	0.13 ± 0.39	−0.06 ± 0.26
Q15.	The instructor was enthusiastic.	−0.01 ± 0.17	0.09 ± 0.12	0.23 ± 0.17*	0.09 ± 0.14
Q16.	The instructor kept students informed of their progress.	0.03 ± 0.34	0.11 ± 0.45	0.39 ± 0.50	0.31 ± 0.33*
Q17.	The instructor set high standards for students.	−0.02 ± 0.11	0.09 ± 0.36	0.22 ± 0.15*	0.08 ± 0.20
# of questions with an average increase.		6	7	15	10

* Questions for which the average change is statistically significant (d.f.=6, $p < 0.05$, paired samples t-test) are indicated with an asterisk.

The number of *statistically significant* changes also varies by cohort. Overall, the change in average response from midterm to end-of-term ratings is statistically significantly different for 11 of 68 cases (again, because only seven cases compose each cohort, caution should be exercised in interpreting the findings). In *Cohort 0: Control* none of the changes was significant, as expected for the control cohort. For *Cohort 1: Ratings Report* one of the 17 questions demonstrated a statistically significant *decrease* in average response; however, the magnitude of the change is small for all of these. The two cohorts characterized by a consultation yielded the most significant change in response from midterm to end-of-term, hypothetically as a result of changes triggered by the consultation. *Cohort 2: Feedback and Consult* had a statistically significant increase for six questions, and *Cohort 3: Videotape and Consult* had two of 17 questions with a statistically significant increase and two with a statistically significant decrease. The magnitude of the changes for *Cohort 2: Feedback and Consult* is relatively large for many cases, while it is small for most changes for *Cohort 3: Videotape and Consult*.

The change in average response is notable for several questions. Question 7 ("The instructor used class time well.") decreased for all cohorts, but it showed the smallest average decrease for *Cohort 2: Feedback and Consult*, the one group where the most class time was taken up for the intervention. Responses to Question 11 ("The instructor used techniques that fostered class participation.") showed the highest average increases for *Cohort 2: Feedback and Consult* and *Cohort 3: Videotape and Consult* (0.41 and 0.20), and all faculty in these two cohorts received a consultant-driven intervention. The change in response to Question 12 ("The instructor treated students with respect.") is also of interest–only faculty in *Cohort 2: Feedback and Consult* demonstrated a statistically significant (and positive) change.

Overall *Cohort 2: Feedback and Consult* showed the greatest improvement in student ratings of teaching, but to study more clearly the differences in student ratings of teaching between cohorts, an independent samples t-test was conducted. Table 5 lists all questions for which the difference in the change from midterm to end-of-term response was statistically significant for each of six possible cohort pairs.

Table 5. Significant differences in student ratings of teaching between cohorts.
Questions having statistically significantly different changes from midterm to end-of-term are listed for each of six cohort pair combinations (d.f.=6, p<.05, independent samples t-test).

	Cohort 0: Control	Cohort 1: Ratings Report	Cohort 2: Feedback and Consult
Cohort 3: Videotape and Consult	No differences	4, 6, 10	8, 12
Cohort 2: Feedback and Consult	3, 8, 11, 15, 17	1, 2, 3, 11, 12, 13	
Cohort 1: Ratings Report	No differences		

There were no significant differences in the change on any question between *Cohort 0: Control* and *Cohort 1: Ratings Report* and between *Cohort 0: Control* and *Cohort 3: Videotape and Consult*; i.e., the change in average student response for Cohorts 1 and 3 did not differ significantly from the control cohort. On the other hand, *Cohort 2: Feedback and Consult*

showed statistically significant differences from *Cohort 0: Control* on five questions (Q3, Q8, Q11, Q15, and Q17). *Cohort 2: Feedback and Consult* also showed statistically significant differences from *Cohort 1: Ratings Report* on six questions (Q1, Q2, Q3, Q11, Q12, and Q13), and from *Cohort 3: Videotape and Consult* on two questions (Q8 and Q12). In all cases, average responses for *Cohort 2: Feedback and Consult* increased from midterm to end-of-term while the other cohorts' scores decreased, remained the same, or increased to a lesser extent. Statistically significant differences were also found between *Cohort 1: Ratings Report* and *Cohort 3: Videotape and Consult* on three questions (Q4, Q6, and Q10). For these differences, *Cohort 3: Videotape and Consult* had an increase and *Cohort 1: Ratings Report* had a decrease for two questions (Q4 and Q10) while for the other question (Q6) *Cohort 3: Videotape and Consult* had a decrease and *Cohort 1: Ratings Report* had an increase.

2.3. Faculty survey

At the end of the term, all 28 faculty completed a short online survey about the project. Besides describing their level of engagement in activities related to teaching improvement, faculty rated themselves on several aspects of teaching, described their personal experiences with the interventions for their cohort, and had an opportunity to provide general comments.

2.3.a Faculty self-ratings of teaching

To compare student ratings of teaching with faculty self-ratings, faculty were asked about thirteen of the same aspects that the students rated (four student-specific questions—Q1, Q2, Q3, and Q4—were not included). Table 6 shows the average student response for all 28 courses at the end of the term and the corresponding faculty self-ratings. The average correlation between the two ratings is also listed for each question. Rather than demonstrating differences between cohorts, these data are intended to identify differences between student and faculty perceptions. Questions for which the average student rating and the average faculty rating are statistically significantly different (d.f.=6, p<.05, paired samples t-test) are identified on Table 6. The average difference between faculty self-rating and student rating is plotted in Figure 1 (in that figure, the order of the questions is rearranged so questions with the highest difference are on the left side of the page).

Data presented on Table 6 and Figure 1 is consistent with previous research on student ratings that supports two important findings[1, 2, 5, 6]. First, faculty tend to rate themselves higher than their students do—especially in regards to the nature, quality, and frequency of feedback from the instructor to students; the instructor's friendliness, concern, and respect for students; and the instructor's availability and helpfulness (approximately Q16, Q12, and Q9). Second there is relative similarity between the two groups in the pattern of their assessment of the strengths and weakness of the teacher (i.e., there is profile similarity). In Feldman's 1989 meta-analysis[5], the average correlation for profile similarity across ten studies was extremely high (0.84), and in Cashin's 1995 revisitation of the issue[1], the average correlation between instructor self-ratings and student ratings was also high (0.29).

Table 6. Average end-of-term student and faculty ratings of teaching. Level of agreement with each statement. (Scale for responses: 1=strongly disagree, 2=agree, 3=neutral, 4=agree, 5=strongly agree.)

	Average student rating	Average faculty self-rating	Correlation coefficient (N=28)
Reflecting on my teaching this term, I feel that overall I…			
Q5. ... gave clear explanations.	4.03 ± 0.53	4.18 ± 0.48	0.365
Q6. ... acknowledged all questions insofar as possible.	4.41 ± 0.30*	4.79 ± 0.42*	0.000
Q7. ... used class time well.	4.07 ± 0.46	4.07 ± 0.72	0.505*
Q8. ... was well prepared for each class.	4.40 ± 0.30	4.32 ± 0.67	0.276
Q9. ... made work requirements and grading system.	4.09 ± 0.36*	4.46 ± 0.64*	0.177
Q10. ... required level of work appropriate for credit received.	3.95 ± 0.37	4.11 ± 0.63	0.304
Q11. ... used techniques that fostered class participation.	3.67 ± 0.63	3.82 ± 0.77	0.464*
Q12. ... treated students with respect.	4.48 ± 0.35*	4.82 ± 0.39*	0.473*
Q13. ... taught in a way that met all students' needs.	3.90 ± 0.49	3.68 ± 0.67	0.252
Q14. ... was willing to meet and help students outside of class.	4.23 ± 0.42*	4.71 ± 0.53*	0.596*
Q15. ... was enthusiastic.	4.38 ± 0.48	4.57 ± 0.50	0.332
Q16. ... kept students informed of their progress.	3.82 ± 0.43	3.96 ± 0.69	0.423*
Q17. ... set high standards for students.	4.12 ± 0.37	4.39 ± 0.63	0.390*

* Questions for which the average change is statistically significant (d.f.=6, $p<0.05$, paired samples t-test) are indicated with an asterisk.

Figure 1. Average difference in faculty and student ratings.
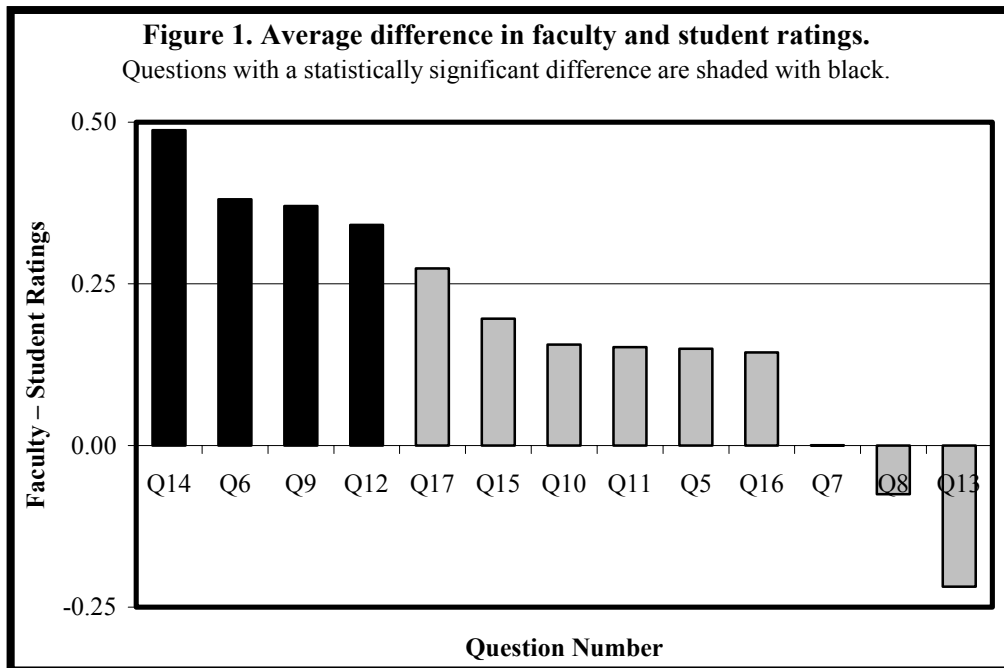Questions with a statistically significant difference are shaded with black.

Table 6 illustrates the first finding in that faculty self-ratings are higher than student ratings for all except three questions (ratings were the same for Q7 and student ratings were higher than faculty self-ratings on Q8 and Q13). This is more clear in Figure 1 from the ten bars that are

above the horizontal axis and just two bars below. In fact, faculty scored themselves higher than students did on five questions with statistical significance (shown by a black shaded bar in Figure 1). These questions relate to the degree to which faculty acknowledged questions, made requirements clear, respected students, were available to students, and set high standards, closely matching results described in Feldman[5].

The second finding consistent with research is that the ratings have profile similarity; i.e., when student ratings are high, faculty self-ratings are also high, while low student ratings correspond with low faculty self-ratings. This trend is apparent from Table 6; e.g., for Questions 6, 12, and 15, both student and faculty ratings were high while both ratings were low for Questions 11 and 16. In this work, the correlation between student and faculty ratings ranges from 0.00 to 0.596 (average=0.35), again showing consistency with research.

2.3.b. Faculty perceptions about interventions

The twenty-one faculty who received an intervention (report on midterm student ratings of teaching for Cohort 1, student feedback session and follow-up consultation for Cohort 2 and videotaped class session and follow-up consultation for Cohort 3) were asked to rate the intervention. Table 7 shows the average faculty response to statements about the intervention.

Table 7. Faculty level of agreement with statements about value of intervention.
(Scale for responses: 1=strongly disagree to 5=strongly agree.)

	Cohort 1: Ratings Report	Cohort 2: Feedback and Consult	Cohort 3: Videotape and Consult
a. The intervention was helpful in improving my teaching.	3.9	4.0	3.7
b. The intervention helped me identify areas of my teaching on which I could work.	3.7	4.1	3.7
c. It was easy to design and incorporate changes in my teaching based on the intervention.	3.4	4.0	3.6
d. It was worth the time it took for the intervention.	4.1	4.1	4.0
e. The consultant gave me ample suggestions about my teaching.	N/A	3.9	3.9
f. The consultant encouraged me to reflect on my teaching.	N/A	3.9	4.4
g. I would recommend this intervention to my colleagues.	4.0	4.1	3.7

Overall, faculty rated the interventions highly, and faculty in all three cohorts "agreed" that the intervention was worth their time. Faculty in *Cohort 2: Feedback and Consult* had a more positive (but not statistically significantly different) rating of the helpfulness of the intervention in improving teaching, the ability of the intervention to identify specific areas of teaching for improvement, and the ease of design and incorporation of changes based on the intervention than did faculty in both *Cohort 1: Ratings Report* and *Cohort 3: Videotape and Consult*. The faculty who received consultations (i.e., *Cohort 2: Feedback and Consult* and *Cohort 3: Videotape and*

Consult) agreed approximately equally that the consultant gave them ample suggestions to improve their teaching. Faculty who received a videotape-based consultation responded that the consultant encouraged them to reflect on their teaching to a greater degree than did faculty in *Cohort 2: Feedback and Consult*. This is the only statistically significant finding in the table, and it is in keeping with the protocol for the videotape consultations in which the faculty member was encouraged to be reflective.

Faculty who received an intervention were also presented a series of open-ended questions about the interventions. They were asked to describe the aspects about the intervention they found most and least helpful and to describe specific changes in teaching they made based on the intervention.

Faculty in *Cohort 1: Ratings Report* received a summary of the midterm student ratings of teaching survey for their class, and they may or may not have used the report. Four faculty replied to the questions about the utility of the report and they cited several useful aspects of the report: the questions that related to specific teaching techniques, feedback about the pace of the course, and the opportunity to adjust the teaching approach in the middle of the term. Comments about aspects of the report that were least helpful include the lack of written student feedback and the use of the standard college-wide questions (i.e., Q1 through Q4 on the survey). Three of the seven reported making changes in their class (attempts to foster more class participation and manage class time better) as a result of receiving the midterm report on student ratings.

Faculty in *Cohort 2: Feedback and Consult* received confidential student feedback and a follow-up consultation. Faculty reported that the summary of student comments and the consultant insight were especially helpful, and several commented on the ease with which student suggestions could be incorporated. The least helpful aspects of this intervention that were reported included: the difficulties scheduling the student feedback session, student comments about aspects of the course that could not be changed, and the short time during which changes could actually be implemented. Six of seven faculty members reported making changes in their classes as a result of the consultation (increased level of student interaction; provided more explicit outline of course and discussions about problem solving process; and introduced more opportunities to engage students in class).

Faculty in *Cohort 3: Videotape and Consult* had a class session videotaped and received a follow-up consultation about the videotape. Faculty noted several helpful aspects of the intervention including suggestions from the consultant and the opportunity to observe students from a different perspective. Faculty commented negatively that the time required to review the videotape was too lengthy and that the focus on *self*-reflection was less desirable than receiving an evaluation from the consultant. Three of seven faculty reported making changes (primarily in regards to more effective use of student questioning), and three reported plans to make changes in future classes as a result of the consultation.

Finally, all faculty were given the opportunity to describe anything else they would like the research team to know about their experience with the research project. Three faculty reported dissatisfaction with the amount of time that participation in the research project took from their class, particularly the time for the student surveys. One participant suggested that the college

should move towards midterm evaluations for all courses, and two noted that the overall reward structure of the college should be changed in an effort to improve teaching. Two faculty members also indicated that they enjoyed the project and would like to participate in the future.

2.4.   Consultant discussions

Four instructional consultants having backgrounds in science, math, or engineering participated in the project. Each had previous experience conducting student feedback sessions with faculty, but not everyone had experience conducting a video consultation like the one used for the project. Further, consultations about student feedback sessions are common at the Center for Research on Learning and Teaching at the University of Michigan (where the instructional consultants are employed), whereas video consultations are usually used to diagnose specific issues or when whey are specifically requested by the faculty. As such, the consultants were much more familiar with the one type of consult for this project than with another type. Each consultant conducted at least one consultation pertaining to a student feedback session and at least one video consultation (for a total of three or four consultations per consultant). The consultants also conducted the student feedback sessions and the class videotaping.

Since the consultants had different levels of experience and different approaches to the consultations, leading to potential consultant bias in the faculty perceived value of the intervention, an analysis was conducted to compare faculty responses about the value of the intervention according to consultant grouping. Of 42 possible differences (seven questions compared across four different groups – i.e., six different comparison pairs for each question), only two differences were significant. Both of the differences have to do with "It was easy to design and incorporate changes in my teaching based on the consultation." Otherwise, faculty responses to the value of the interventions are similar across consultant groups, indicating that there was effectively no consultant bias.

To ascertain the nature of the consultations and identify kinds of issues that arose in each consultation, a formal discussion with the consultants was conducted. Each consultant felt there were distinct differences between the quality and the nature of the two consultation processes.

2.4.a.   Student feedback sessions and follow-up consultations

The consultants noted several strengths about the midterm student feedback process: (1) the process captures both strengths and areas for improvement related to the course up to that point in time (not just related to one class period), (2) the process is guided by direct feedback from the students, (3) the student feedback data provides immediate talking points for the consultation, and (4) the process is also easily adaptable to classrooms of varied size and academic area. Further, the consultants noted the power of the student feedback session in creating the expectation and opportunity for change in the classroom during the semester. However, the consultants did comment that the intervention could result in students expecting change that the faculty member was unwilling or unable to make or that it could cause the faculty member to lose sight of what is going well while trying to implement changes in other areas.

Consultants noted weaknesses of the intervention that include student difficulty in identifying constructive suggestions for improvement and the potentially limiting effect of the feedback on a faculty member's reflection about his/her teaching. Further, consultant experience, skill, and style vary widely as do faculty members' receptivity and reflective abilities. As such, the consultation process is idiosyncratic for each student feedback session, making it difficult to assess the impact of the intervention on teaching and learning. Further, this approach is the one most "controlled" by students, so if the faculty member responds to all of the student comments, there is likely to be an increases in the student ratings of teaching. Results could be different if something other than student evaluations were used as a measure of teaching improvement.

2.4.b. Videotaped consultations

Positive aspects of the videotaped consultations that were described by the consultants include the ability to capture explicit visual data, thus allowing the consultant and faculty to view the same faculty/student behaviors and the same whole group and small group interactions simultaneously, while permitting "stop-action," "fast-forward," and "rewind" options. The consultants felt that the video data would be best combined with observational data (charts and notes) to provide a comprehensive portrait of a single class session. However, the process, while neutral in its expectations for what a faculty member will do with the information, lacks the expectation for improvement that is inherent in student feedback sessions. Other downsides include the limited data from videotaping only a single class, a faculty member's discomfort with watching him/herself on video (resulting in his/her becoming distracted by physical mannerisms unrelated to teaching behaviors), the time required to view the video, and the resource and supply intensiveness of collecting the video and transferring it to DVD.

The consultants suggested that good candidates for a videotape consultation would be faculty who had previously engaged in reflective practices about their teaching since they would be able to separate themselves from the video and comment on "behaviors" they observe from a relatively neutral perspective. Other candidates include experienced or highly skilled faculty who may rarely have the opportunity to see themselves "in practice," and faculty members who have specific events (activities or instructional methods) during a class period that they wish to have captured on video in order to see student response or review the level of clarity of their instructions. Consultants also suggested that several observational tools often used during student feedback sessions (i.e., classroom diagrams, maps of interactions, time logs, etc.) could be incorporated into the videotape protocol to capture a fuller picture of the classroom activities and aid in the consultation process.

3. Conclusions

Data presented here is based on the limited participation of 28 faculty. As such, it should be considered a *pilot* study and no true conclusions can be made. However, based on the *trends* in the data, it appears that the student feedback and follow-up consultation may have the most positive impact on student ratings of teaching and that the videotaped class session with follow-up consultation is a promising method to improve teaching. Other trends supported by this limited data include the following:

o The correlation between student ratings of teaching and faculty self-ratings is consistent with research on student ratings. Overall, faculty tend to rate themselves higher on specific teaching behaviors than their students do, and student and faculty ratings track moderately well (there is general agreement between faculty and student rating patterns).

o Most faculty who received an intervention made changes in their teaching. Several reported making specific changes, especially in the way in which they fostered class participation and managed class time. Of particular note is that six of seven faculty in *Cohort 2: Feedback and Consult* reported making changes while only six of fourteen faculty in the other two cohorts that received an intervention did.

o Faculty in the two cohorts having a consultation had the most change in student ratings. They had the greatest number of *statistically significant* changes in student ratings from midterm to the end of term, as well as the greatest number of *positive* changes. Changes were most notable in regards to using techniques to foster class participation, showing enthusiasm, and keeping students informed of progress.

o Faculty who received a consultation driven by student feedback (*Cohort 2: Feedback and Consult*) generally demonstrated the highest gains in student ratings. This cohort had significantly different changes on at least some student ratings when compared to all other cohorts. For all of the differences, the change in student rating was more positive for Cohort 2 than for the comparison cohort.

o Overall, faculty perceived that the interventions were helpful and were worth the time required. Most faculty indicated they would recommend the intervention to their colleagues.

o Discussions with the instructional consultants also indicated that there were distinct differences between the quality and nature of the two consultation processes (student feedback-driven versus focused on class videotape). In general, they were more comfortable with the consultations for *Cohort 2: Feedback and Consult*, but they made several suggestions to improve the consultations for *Cohort 3: Videotape and Consult*.

From this limited study, it appears that the student feedback and follow-up consultation may have the most positive impact on student ratings of teaching. However, the videotaped class session with follow-up consultation is also a promising method to improve teaching. Future work for this project includes the following two modifications to the protocol:

o *Cohort 0: Control* will be replaced with a protocol in which faculty receive a summary report on midterm student ratings of teaching (like *Cohort 1: Ratings Report*) and then have a follow-up consultation. In addition, all faculty will receive an individual report summarizing the midterm student ratings of teaching survey for their course at the middle of the term. This modification is consistent with research on student ratings suggesting that feedback from student ratings is much more effective in promoting teaching improvement when used in conjunction with consultation than when used alone[7, 9].

o The protocol for the videotaped consultations will be modified so that consultants view the videotape prior to the consultation and identify specific segments on which to focus during the consultation.

---

[i] No data was collected on faculty's teaching performance prior to the term in which the project was performed. As such, this project evaluates the impact of various methods to improve teaching within one term. It does not measure long-term improvement.

[ii] Before initiating the project, the project was approved for human subjects' research by the University of Michigan Behavioral Sciences Institutional Review Board.

[iii] Consent forms for students and faculty indicating their willingness to be videotaped were collected prior to videotaping.

[iv] One faculty member was unable to schedule the end-of-term student ratings of teaching survey. Therefore, although he was originally considered to be part of the project, he was unable to participate in the study.

References

1. Cashin, W. E. (1995, September). Student ratings of teaching: The research revisited. *Idea Paper No. 32*. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

2. Centra, J. A. (1973, Winter). Self-ratings of college teachers: A comparison with student ratings. *Journal of Educational Measurement, 10*(4), 287–295.

3. Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin, 39*(8), 3–7.

4. 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), 217–231.

5. Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education, 30*(2), 137–194.

6. Marsh, H. W. (1984). Students' evaluations of teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*(5), 707–754.

7. Menges, R. J., & Brinko, K. T. (1986, April). *Effects of student evaluation feedback: A meta-analysis of higher education research*. Paper presented at the 70th Annual Meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED270408)

8. Murray, H. G. (1985, September). Classroom teaching behaviors related to college teaching effectiveness. *New Directions for Teaching and Learning, 23*, 21–34. San Francisco: Jossey-Bass.

9. Piccinin, S. (1999, Fall). How individual consultation affects teaching. *New Directions for Teaching and Learning, 79*, 71–83. San Francisco: Jossey-Bass.

10. Sorcinelli, M. D. (1991). Research findings on seven principles. *New Directions for Teaching and Learning, 47*, 13–25. San Francisco: Jossey-Bass.

11. Taylor-Way, D., & Brinko, K. T. (1989). Using video recall for improving professional competency in instructional consultation. *To Improve the Academy,* 8, 141–156.

Appendix 1. Student ratings of teaching survey.

This form is a sample of the midterm and end-of-term student ratings of teaching survey.

E&E 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 midterm student rating of teaching survey, and these questions will be repeated on the standard end-of-term 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 used for comparative purposes ONLY--your course instructor will NOT receive any report about your responses or the responses of other students in this class until the end of the term. At the end of the term, a report summarizing overall student responses on this survey and the end-of-term survey will be given to your instructor.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Overall, this is an excellent course.	5	4	3	2	1
Overall, the instructor is an excellent teacher.	5	4	3	2	1
I am learning a great deal from this course.	5	4	3	2	1
I had a strong desire to take this course.	5	4	3	2	1
The instructor gives clear explanations.	5	4	3	2	1
The instructor acknowledges all questions insofar as possible.	5	4	3	2	1
The instructor uses class time well.	5	4	3	2	1
The instructor seems 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 uses techniques to foster class participation.	5	4	3	2	1
The instructor treats students with respect.	5	4	3	2	1
The instructor is teaching in a manner that serves my needs as a student.	5	4	3	2	1
The instructor is willing to meet and help students outside class.	5	4	3	2	1
The instructor is enthusiastic.	5	4	3	2	1
The instructor keeps students informed of their progress.	5	4	3	2	1
The instructor sets high standards for students.	5	4	3	2	1

Appendix 2. Online faculty survey.

The online faculty survey varied depending on the cohort in which the faculty member participated. The first set of questions (labeled common) included in this appendix was part of all surveys. Questions specific for Cohorts 1, 2, and 3 are included after that. The survey was completed by all faculty at the conclusion of the project.

Common questions for all cohorts
1. Using the scale provided, please indicate how often you engage in the following behaviors (ranged from never to often):
   - Talk to others about your teaching
   - Solicit feedback from students about how they feel a class is going
   - Spend time revising the content of your courses
   - Spend time revising the methods that you use to teach your courses
   - Attend programs on effective teaching
   - Read journals or books on effective teaching
2. Using the scale provided, please indicate your level of agreement with each of the following statements (ranged from strongly agree to strongly disagree):
   Reflecting on my teaching this term, I feel that overall I...
   - gave clear explanations
   - acknowledged all questions insofar as possible
   - used class time well
   - was well prepared for each class
   - made clear the work requirements and grading system from the beginning
   - required a level of work appropriate for the credit students received for the class
   - used techniques to foster class participation
   - treated students with respect
   - taught in a manner that met all students' needs
   - was willing to meet with and help student outside of class
   - was enthusiastic
   - kept students informed of their progress
   - set high standards for students
3. Have you had a teaching consultation of any sort before this term? Yes/No
4. Is there anything else you would us to know about your experience with the research project? (open-ended question)

Additional questions for *Cohort 1: Ratings Report*
5. Using the scale provided, please indicate your level of agreement with each of the following statements regarding the report on midterm student ratings of teaching that you received (scale ranged from Strongly Disagree to Strongly Agree):
   - The report on student ratings was helpful in improving my teaching
   - The report on student ratings helped me identify areas of my teaching on which I could work
   - It was easy to design and incorporate changes in my teaching based on the report on student ratings

- It was worth the time it took to collect the student ratings of teaching and review the report
- I would recommend collecting midterm student ratings of teaching and reviewing the report to my colleagues

6. Which aspects of the report on midterm student ratings of teaching did you find the MOST helpful, and why (open-ended question)?
7. Which aspects of the report on midterm student ratings of teaching did you find the LEAST helpful, and why (open-ended question)?
8. If there are specific changes you have made in your teaching based on the report on midterm student ratings of teaching you received, please describe them here (open-ended question).

Additional questions for *Cohort 2: Feedback and Consult*
5. Using the scale provided, please indicate your level of agreement with each of the following statements regarding the consultation you received about the confidential student feedback session (ranged from Strongly Disagree to Strongly Agree):
   - The consultation was helpful in improving my teaching
   - The consultation helped me identify areas of my teaching on which I could work
   - It was easy to design and incorporate changes in my teaching based on the consultation
   - It was worth the time it took to collect the confidential student feedback and to complete the subsequent consultation
   - The consultant gave me ample suggestions about my teaching
   - The consultant encouraged me to reflect on my teaching
   - I would recommend a confidential student feedback session and follow-up consultation to my colleagues
6. Which aspects of the consultation did you find the MOST helpful, and why (open-ended question)?
7. Which aspects of the consultation did you find the LEAST helpful, and why (open-ended question)?
8. If there are specific changes you have made in your teaching based on the consultation you received, please describe them here (open-ended question).

Additional questions for *Cohort 3: Videotape and Consult*
5. Using the scale provided, please indicate your level of agreement with each of the following statements regarding the consultation you received about the videotaped class session (ranged from Strongly Disagree to Strongly Agree).
   - The consultation was helpful in improving my teaching
   - The consultation helped me identify areas of my teaching on which I could work
   - It was easy to design and incorporate changes in my teaching based on the consultation
   - It was worth the time it took videotape the class session and complete the follow-up consultation
   - The consultant gave me ample suggestions about my teaching
   - The consultant encouraged me to reflect on my teaching

- I would recommend having a class session videotaped and having a follow-up consultation to my colleagues

6. Which aspects of the consultation did you find the MOST helpful, and why (open-ended question)?
7. Which aspects of the consultation did you find the LEAST helpful, and why (open-ended question)?
8. If there are specific changes you have made in your teaching based on the consultation you received, please describe them here (open-ended question).