Running Head: STUDY OF WRITTEN COURSE EVALUATION COMMENTS

A Study of Students' Written Course Evaluation Comments at a Public University

Dawn M. Zimmaro, Charles S. Gaede, E. Joel Heikes, Mi-Suk P. Shim, and Karron G. Lewis

University of Texas at Austin

Researchers have investigated end-of-course evaluations for decades, with most of the focus on how students respond to closed-ended survey questions. However, little attention has been paid to open-ended questions and the types and numbers of comments that students write. This study investigated how often students write comments, the topics of those comments, how students who wrote comments differed from those who did not, and how the tone of written comments related to overall instructor and course ratings. Results indicated that about 40% of students wrote comments on end-of-course evaluations, with most comments related to teaching skills and course quality. Additionally, students' comments tended to be positive and their ratings of the course and instructor tended to be more positive than those of students who did not write comments.

A Study of Students' Written Course Evaluation Comments at a Public University

The collection of student written comments in end-of-course evaluations is a widespread practice. The presence or absence and the merit of students' written comments are often mentioned in the literature but have seldom been studied systematically. In fact, the literature reviewed for this study yielded only two substantive articles and a few articles with indirect relevance.

Ory and Piper (1980) used three methods to collect student evaluation information: objective response items, written responses to open-ended items, and group interviews. They analyzed the data with regard to the quality and kind of information obtained, finding that all three methods provided the same quality of information and that the method was not an influence in the rankings of instructors. However, method of collection did influence the kind of information, in that the written comments and the group interviews provided specific comments for improving instruction that the quantitative information did not.

Theall and Franklin (1991) found that about 10% of students will provide written comments, unless something unusual occurs in the course of the instructional term. In a typical course, usually the most and least satisfied students write comments. When an abnormal positive or a negative event occurs, written comments tend to match the quantitative results in frequency and intensity.

In view of these findings, it is important to determine what students are saying in open-ended comments and how that information contributes to the evaluation of university teaching. The purpose of this study was to answer six research questions using end-of-course evaluations at a large public university.

1. What percentage of students provides written comments on course evaluation forms?
2. On what topics do students provide written comments?
3. How do students who write comments differ from those who do not, by college, course size, course level, course workload rating, self-reported probable course grade, and self-reported GPA range?
4. How do students who provide written comments differ from those who do not, by ratings given for overall effectiveness of instructor and course?
5. How do students who write positive comments differ from those who write negative comments, by college, course size, course level, course workload rating, self-reported probable course grade, and self-reported GPA range?
6. How do students who write positive comments differ from those who write negative comment, by ratings given for overall effectiveness of instructor and course?

Method

Sampling

A sampling size of 100 course sections—about 1% of the total of approximately 8,000 possible course sections—was chosen as the maximum number of courses and related forms that

could be interpreted qualitatively, given available resources. Courses were included in the study if there was a completed course evaluation form for spring 2005. The sampling plan shown in Appendix A indicates the courses included in the analysis.

One-hundred course sections were selected using a random stratified approach based on four strata:

College: The proportion of courses selected from each college was based on the percentage of students enrolled in each college in fall 2003, the most recent year for which institutional research data were available.

Department: An equal number of courses was selected from each department within a college.

Class size: An equal proportion of courses (33% of each size) was selected based on three enrollment ranges: small (35 or fewer), medium (36-100), or large (101 or more).

Course level: The proportion of courses selected from the undergraduate and graduate levels was based on the percentage of students enrolled in undergraduate and graduate programs for each college in fall 2003.

Researchers used an alphabetized list of faculty (by department) for whom there were course evaluation data as the source of courses for random selection for inclusion in the sample. Working from a list of sequentially numbered courses within each department, they used a random number generator to select course numbers within a given department to include in the sample, until the department's quota was filled in accord with the sampling plan. During the sampling process, researchers made adjustments to account for course sections lacking course evaluation data or course sections that were outside the appropriate stratum. In particular, only a few large graduate course sections were available, so they were underrepresented and small graduate course sections were overrepresented in terms of the original sampling plan.

A total of 103 course sections comprising 4,880 forms were included in the final sample. Refer to Appendix A for the original and final sampling plans.

Procedures

All faculty members at the participating university are requested to administer a course evaluation form at the end of each semester, assuring their students that responses are anonymous, with no personally identifying information. Every form contains a comments section with these instructions: "In many ways your written comments can be the most important part of your evaluation of the course and instructor. In the space provided, please indicate what aspects of the course content and instruction were best, how the instructor could improve his or her teaching, and how the content of the course might be improved. The instructor will receive this form after the semester is over."

The first research question was addressed using scanning technology to identify forms that contained written comments.

The second research question was addressed through systematic qualitative analysis of written comments from sampled course evaluation forms. Researchers assigned—from a list of codes adapted from the university's bank of course evaluation questions—codes to each student comment, first by general category, then by specific trait within a general category, and finally by whether the comment was neutral, negative, or positive. Unassigned codes were included in the event additional codes emerged from the data. See Appendix B for a copy of the code list.

Three graduate students with previous qualitative coding experience completed the coding following training by the two lead researchers. To ensure consistent application of the codes, the lead researchers reviewed the graduate students' work with a sample of the written comments, and the three coders also worked with a common set of surveys (25% of the total sample) to establish inter-rater reliability. Once consistency was established, each rater coded a third of the remaining comments, which were then analyzed for common themes.

College, class size, course level, course workload, self-reported probable course grade, and self-reported grade point average (GPA) range were used to analyze group differences among code themes. Class size was defined as *small*, *medium*, and *large*: enrollments of 1-35, 36-100, and 101 or more, respectively. Course level was defined as *undergraduate* or *graduate*. Course workload was defined as *1 = excessive*, *2 = heavy*, *3 = average*, *4 = insufficient*, or *5 = light*. Self-reported probable course grade was recorded as *1 = A*, *2 = B*, *3 = C*, *4* = D, or *5 = F*. Self-reported GPA ranges were recorded as *1* = less than 2.00, *2* = 2.00 – 2.49, *3* = 2.50 – 2.99, *4* = 3.00 – 3.49, and *5* = 3.50 – 4.00.

The third and fifth research questions were addressed using quantitative analysis techniques for the same sample of course evaluation forms. Logistic regression was used to compare students by college of the course, course size, course level, course workload rating, probable course grade, and GPA range. The fourth and sixth research questions were addressed using a t-test to compare overall course and instructor ratings between students.

Results

Inter-rater reliability was determined by calculating the percent agreement among the three coders. With 1,589 codes assigned by the raters in the common set of 500 surveys, the three raters had perfect agreement on 1,152 of the codes (72.5%). An additional 17.9% of the codes were commonly assigned by two of the three, and for 9.6% of the codes the raters did not have any agreement on a code.

1) What percentage of students provides written comments on their course evaluation form?

A total of 4,880 forms were returned for the 103 sampled course sections, with an average return rate of 67%. Of those, 1,993 forms (40.8%) contained some type of written comment from a one-word phrase or drawing to multiple sentences. Additionally, results from use of the scanning software indicated that approximately 40% of all spring 2005 course evaluation forms (54,753 out of 134,214 forms) contained a written comment, so the sample was representative of the larger university. Table 1 shows the return rate per course section and Table 2 shows the comment rate per course section.

Table 1
Descriptive Statistics for Percent of Forms Returned per Course Section

	N	Minimum	Maximum	Mean
Return rate	103	30.00%	100.00%	66.86%

Table 2
Descriptive Statistics for Percent of Comments per Course Section

	N	Minimum	Maximum	Mean
Comment rate	103	9.09%	100.00%	40.82%

2) On what topics do students provide written comments?

Tone of Comments

On the 1,993 forms with written comments, raters assigned 5,896 topic codes, judging almost two-thirds of the comments (65%) as positive. Table 3 shows the distribution of written comments by tone.

Table 3
Frequency and Percent of Written Comments by Tone

Comment tone	Frequency	Percent
Positive comments	3,859	65.4%
Negative comments	1,653	28.1%
Neutral comments	384	6.5%
TOTAL	5,896	100.0%

Descriptive Statistics for Category Codes by Tone of Comment

Patterns for the subjects of students' most common positive and negative comments were the same for four of the top five categories: teaching skills, course quality, short sentiments, and course organization, as shown in Tables 4-7. Table 4 shows descriptive statistics for the general category codes related to positive comments, and Table 5 shows the breakdown of specific comments for the top four category codes. Table 6 shows descriptive statistics for the category codes related to negative comments, and Table 7 shows the breakdown of specific comments for the top four category codes.

Table 4
Pattern of Positive Comments by Subject

Subject category code	Frequency	Percent
Teaching Skills	1,134	29.4
Course Quality	1,064	27.6
Short Sentiments	547	14.2
Interpersonal Skills	524	13.6
Course Organization	155	4.0
Course Impact	131	3.4
Drawing	79	2.0
Grading	66	1.7
TA	65	1.7
Other	54	1.4
Technology Use	32	.8
Course Goals/Objectives	8	.2
Total	3,859	100.0

Table 5
Pattern of Positive Comments by Specific Comment for the Four Most Frequent Subjects

Subject category code	Frequency	Percent
Teaching Skills		
General (includes any general comments about instructor)	631	55.6
Quality of lectures/discussions (content, method, how info is conveyed)	144	12.7
Clarity of presentation/explanations	94	8.3
Knowledge of subject matter	83	7.3
Use of examples/used real-world examples	37	3.3
Oral communication skills (speaking skills, speech patterns, delivery)	29	2.6
Check understanding	20	1.8
Student participation	16	1.4
Student recommendations	16	1.4
Encourage critical thinking	14	1.2
Response to questions in class	12	1.1
Student motivation	11	1.0
Quality of feedback provided to students	7	.6
Uses variety of teaching techniques	7	.6
Explain expectations clearly	5	.4
Identify important concepts	5	.4
Staying on topic/use of class time	3	.3
Total	1,134	100.0
Course Quality		
General (includes general references to course/class)	539	50.7
Value of course	213	20.0
Course content/material	106	10.0
Activities and assignments	74	7.0
Course supplements (handouts, PowerPoint slides, video)	71	6.7
Textbooks/Readings	49	4.6
Student recommendations	12	1.0
Total	1,064	100.0
Short Sentiments		
Brief student expressions of feelings (‘Thank You’)	547	100.0
Interpersonal Skills		
General	297	56.7
Enthusiasm	95	18.1
Approachability/Rapport	51	9.7
Availability	39	7.4
Class climate	34	6.5
Acceptance of other viewpoints	7	1.3
Student recommendations	1	.3
Total	524	100.0

Table 6
Pattern of Negative Comments by Subject

Subject category code	Frequency	Percent
Course Quality	390	23.6
Teaching Skills	337	20.4
Course Organization	267	16.1
Grading	222	13.4
Short Sentiments	139	8.4
Teaching Assistant (TA)	105	6.3
Other	64	3.9
Interpersonal Skills	61	3.7
Technology Use	52	3.1
Course Impact	9	.6
Drawing	6	.4
Course Goals/Objectives	1	.1
Total	1,653	100.0

Table 7
Pattern of Negative Comments by Specific Comment for the Four Most Frequent Subjects

Subject category code	Frequency	Percent
Course Quality		
Student recommendations	133	34.1
Course content/material	73	18.7
Activities and assignments	58	14.9
General (includes general references to course/class)	41	10.5
Course supplements (handouts, PowerPoint slides, video)	39	10.0
Textbooks/Readings	24	6.2
Value of course	22	5.6
Total	390	100.0
Teaching Skills		
Quality of lectures/discussions	91	27.0
Student recommendations	68	20.2
General (includes any general comments about instructor)	49	14.5
Oral communication skills (speaking skills, speech patterns, delivery)	32	9.5
Staying on topic/use of class time	30	8.9
Clarity of presentation/explanations	20	5.9
Explain expectations clearly	12	3.6
Identify important concepts	7	2.1
Response to questions in class	5	1.5
Use of examples/used real-world examples	5	1.5
Student motivation	4	1.2
Check understanding	3	.9
Encourage critical thinking	3	.9
Quality of feedback provided to students	3	.9
Student participation	3	.9
Knowledge of subject matter	2	.5
Total	337	100.0
Course Organization		
Student recommendations	114	42.7
Course workload	60	22.5
Format	31	11.6
Pacing	28	10.5
General	24	9.0
Organization of content	10	3.7
Total	267	100.0
Grading		
Exams	87	39.2
Student recommendations	69	31.1
Grading practices	35	15.7
Quizzes	15	6.8
General	8	3.6
Assignments/papers/presentations/essays	8	3.6
Total	222	100.0

1) How do students who write comments differ from those who do not by college, course size and level, course workload rating, probable course grade, and self-reported GPA range?

To determine who was writing comments, a logistic regression was performed with *writing comments* as the dependent variable and *course college*, *course size*, *course level*, student rating of the *course workload*, student self-reported *probable course grade*, and student *self-reported GPA range* as predictor variables. Dummy coding was used for the categorical variables of college, course size, and course level. For the college predictor, liberal arts courses were selected as the reference group because theirs was the college with the largest number of forms returned and the college in which most general education courses were offered. For course size, large courses (101 or more students) were selected as the reference group. For course level, graduate courses were selected as the reference group. The overall test of the model coefficients was significant, χ^2 (19) = 209.724, $p < .001$. The results of the logistic regression are shown in Table 8.

Table 8
Logistic Regression Predicting Those Who Wrote Comments by Course College, Course Size, Course Level, Students' Rating of Course Workload, Students' Probable Grade in the Course, and Students' Self-Reported GPA Range

Regression variables	B	S.E.	Exp(B)
College[1]			
Architecture	1.490*	.642	4.438
Business	-.250*	.110	.778
Communication	.366**	.114	1.442
Education	.038	.138	1.039
Engineering	.222*	.104	1.249
Fine Arts	-.041	.196	.960
Information	-1.014*	.393	.363
Natural Sciences	-.326***	.091	.722
Nursing	-.081	.256	.923
Pharmacy	-.978*	.448	.376
Public Affairs	2.214*	1.053	9.156
Social Work	.937	.486	2.552
Intercollegiate programs	-.633	.642	.531
Class Size[2]			
Small (1-35)	.439***	.094	1.551
Medium (36-100)	.343***	.074	1.410
Level[3]			
Undergraduate	-.123	.121	.884
Workload	-.198***	.044	.820
Probable course grade	-.201***	.050	.818
GPA range	.097*	.038	1.102
Constant	.086	.283	1.089

Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$

[1] Liberal Arts used as reference group for College variable

[2] Large class (100+) used as reference group for Class Size variable

[3] Graduate class used as reference group for Class Level variable

Students taking classes in architecture, communication, engineering, and public affairs were more likely to write comments than students taking liberal arts classes. Students taking classes in business, information studies, natural sciences, and pharmacy were less likely to write comments than students taking liberal arts classes. Students in small and medium sized classes were more likely to write comments than students in large classes. Students in undergraduate and graduate courses were equally likely to write comments. Students who rated the course workload near the insufficient or light end of the scale were less likely to write comments than students who rated the course workload near the heavy or excessive end of the scale. As probable course grade decreased from *A* to *F*, students were less likely to write comments. Finally, as GPA range increased, students were more likely to write comments.

4) How do students who provide written comments differ from those who do not on overall instructor and course ratings?

A comparison was made between students who wrote comments and those who did not on the basis of how they rated the instructor, where the range was from *1 = Very unsatisfactory* to *5 = Excellent*. An independent samples t-test was performed to compare the mean ratings for the two groups. Levene's test for equality of variances was significant, $F(1, 3903) = 45.648$, $p < .001$, so equal variances were not assumed. The results indicated a significant difference, $t(3903) = 5.426$, $p < .001$. Those students who wrote comments rated the instructor more positively than did those students who did not write comments. The descriptive statistics for overall instructor rating by presence of comments are shown in Table 9. The frequency distributions for those who did and did not write comments are shown in Table 10.

Table 9
Comparison of Overall Instructor Rating by Whether Students Wrote Comments

	Presence of comments	N	Mean rating	Std. deviation	Std. error mean
Overall instructor rating	Wrote comment	1,969	4.26	.999	.023
	Did not write comment	2,874	4.11	.890	.017

Table 10
Patterns of Overall Instructor Rating by Whether Students Wrote Comments

	Wrote comments		Did not write comments	
Overall instructor rating	Frequency	Percent	Frequency	Percent
Very unsatisfactory	38	1.9	23	0.8
Unsatisfactory	94	4.8	89	3.1
Satisfactory	281	14.3	593	20.6
Very good	457	23.2	1,012	35.2
Excellent	1,099	55.8	1,157	40.3
Total	1,969	100.0	2,874	100.0

Students who were more positive about their instructor were more likely to write comments. Students who wrote comments were more likely to rate the instructor as *excellent*.

A comparison was also made between students who wrote comments and those who did not on the basis of how they rated the course, where the range was from *1 = Very unsatisfactory* to *5 = Excellent*. An independent samples t-test was performed to compare the mean ratings for the two groups. Levene's test for equality of variances was significant, $F(1, 3931) = 13.147$, $p < .001$ so equal variances were not assumed. The results indicated a significant difference, $t(3931) = 6.077$, $p < .001$. Those students who wrote comments rated the course more positively than did those students who did not write comments. The descriptive statistics for overall course rating by presence of comments are shown in Table 11. The frequency distributions for those who did and did not write comments are shown in Table 12.

Table 11
Comparison of Overall Course Rating by Whether Students Wrote Comments

	Presence of comments	N	Mean rating	Std. deviation	Std. error mean
Overall course rating	Wrote comment	1,966	4.06	1.000	.023
	Did not write comment	2,868	3.89	.902	.017

Table 12
Patterns of Overall Course Rating by Whether Students Wrote Comments

	Wrote comments		Did not write comments	
Overall course rating	Frequency	Percent	Frequency	Percent
Very unsatisfactory	40	2.0	22	.8
Unsatisfactory	106	5.4	131	4.6
Satisfactory	365	18.6	811	28.3
Very good	633	32.2	1,073	37.4
Excellent	822	41.8	831	29.0
Total	1,966	100.0	2,868	100.0

Students who were more positive about the course were more likely to write comments. And while students who wrote comments were more likely to rate the course as *excellent*, this tendency was not as pronounced as it was for instructor ratings.

5) How do students who write positive comments differ from those who write negative comments by college, course size, course level, course workload rating, self-reported probable course grade, and self-reported GPA range?

To determine who was writing positive comments, a logistic regression was performed with *writing positive comments* as the dependent variable and *course college*, *course size*, *course level*, student rating of the *course workload*, student self-reported *probable course grade*, and student *self-reported GPA range* as predictor variables. Dummy coding was used for the

categorical variables of college, course size, and course level. For the college predictor, liberal arts courses were selected as the reference group because theirs was the college with the largest number of forms returned and the college in which most general education courses were offered. For course size, large courses (101 or more students) were selected as the reference group. For course level graduate courses were selected as the reference group. The overall test of the model coefficients was significant, χ^2 (19) = 131.246, $p < .001$. The results of the logistic regression are shown in Table 13.

Table 13
Logistic Regression Predicting Those Who Wrote Positive Comments by Course College, Course Size, Course Level, Students' Rating of Course Workload, Students' Probable Grade in the Course, and Students' Self-Reported GPA Range

Regression variables	B	S.E.	Exp(B)
College[1]			
Architecture	.117	.690	1.124
Business	-.475*	.201	.622
Communication	.356	.223	1.428
Education	1.610**	.473	5.002
Engineering	-.145	.188	.865
Fine Arts	.636	.489	1.889
Information	1.051	1.103	2.860
Natural Sciences	-.341	.175	.711
Nursing	-.176	.465	.839
Pharmacy	20.712	15079.334	988977459.474
Public Affairs	.714	1.081	2.042
Social Work	-.382	.572	.682
Intercollegiate programs	-22.661	28401.695	.000
Class Size[2]			
Small (1-35)	.765***	.188	2.149
Medium (36-100)	.475**	.145	1.608
Level[3]			
Undergraduate	.410	.222	1.506
Workload	.189*	.083	1.208
Probable course grade	-.519***	.093	.595
GPA range	-.178*	.077	.837
Constant	1.614**	.523	5.024

Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$
[1] Liberal Arts used as reference group for *college* variable
[2] Large class (100+) used as reference group for *class size* variable
[3] Graduate class used as reference group for *class level* variable

Students taking classes in education were more likely to write positive comments than were students taking liberal arts classes. Students taking classes in business were less likely to write positive comments than were students taking liberal arts classes. Students in small and medium sized classes were more likely to write positive comments than were students in large classes. Students in undergraduate and graduate courses were equally likely to write positive comments. Students who rated the course workload near the insufficient or light end of the scale

were more likely to write positive comments than were students who rated the course workload near the heavy or excessive end of the scale. As probable course grade decreased from *A* to *F*, students were less likely to write positive comments. Finally, as GPA range increased, students were less likely to write positive comments.

6) How do students who write positive comments differ from those who write negative comments on overall instructor and course ratings?

A comparison was made between students who wrote positive comments and those who wrote negative ones on the basis of how they rated the instructor, where the range was from *1 = Very unsatisfactory* to *5 = Excellent*. An independent samples t-test was performed to compare the mean ratings for the two groups. Levene's test for equality of variances was significant, F (1, 526) = 180.655, $p < .001$ so equal variances were not assumed. The results indicated a significant difference, t (526) = -23.682, $p < .001$. Those who wrote positive comments rated the instructor more positively than did those who wrote negative comments. The descriptive statistics for overall instructor rating by tone of comments are shown in Table 14. The frequency distributions for those who wrote positive and negative comments are shown in Table 15.

Table 14
Comparison of Overall Instructor Rating by Tone of Students' Comments

	Tone of comment	N	Mean rating	Std. deviation	Std. error mean
Overall instructor rating	Negative	424	3.24	1.122	.054
	Positive	1,412	4.60	.702	.019

Table 15
Patterns of Overall Instructor Rating by Tone of Students' Comments

	Wrote positive comments		Wrote negative comments	
Overall instructor rating	Frequency	Percent	Frequency	Percent
Very unsatisfactory	7	.5	30	7.1
Unsatisfactory	11	.8	76	17.9
Satisfactory	103	7.3	144	34.0
Very good	298	21.1	112	26.4
Excellent	993	70.3	62	14.6
Total	1,412	100.0	424	100.0

Students who wrote positive comments were overwhelmingly more likely to rate the instructor as *excellent* than were those who wrote negative comments.

A comparison was made between students who wrote positive and negative comments on the basis of how they rated the course, where the range was from *1 = Very unsatisfactory* to *5 = Excellent*. An independent samples t-test was performed to compare the mean ratings for the two groups. Levene's test for equality of variances was significant, $F(1, 566) = 27.536$, $p < .001$ so equal variances were not assumed. The results indicated a significant difference, $t(566) = -3.423$, $p < .001$. Those students who wrote positive comments rated the course more positively than those who wrote negative comments. The descriptive statistics for overall instructor rating by

presence of comments are shown in Table 16. The frequency distributions for those who wrote positive and negative comments are shown in Table 17.

Table 16
Comparison of Overall Course Rating by Tone of Students' Comments

	Tone of comment	N	Mean	Std. deviation	Std. error mean
Overall course rating	Negative	424	3.11	1.045	.051
	Positive	1,409	4.39	.765	.020

Table 17
Patterns of Overall Course Rating by Tone of Students' Comments

Overall course rating	Wrote positive comments		Wrote negative comments	
	Frequency	Percent	Frequency	Percent
Very unsatisfactory	6	.4	32	7.5
Unsatisfactory	22	1.6	74	17.5
Satisfactory	143	10.1	176	41.5
Very good	487	34.6	101	23.8
Excellent	751	53.3	41	9.7
Total	1,409	100.0	424	100.0

Students who wrote positive comments were overwhelmingly more likely to rate the course as *excellent* than were those who wrote negative comments.

Discussion

Review of the pattern of the results reveals a portrait of the type of students who write comments. Generally, about 40% of students write some form of comment on their end-of-course evaluation form, with two-thirds of those comments being positive in tone. Positive comments generally focus on teaching skills and course quality, with specific references to the general teaching skills of the instructor and the quality of the lectures, general course quality, and the value of the course. Negative comments tend to focus on the same two areas, teaching skills and course quality, with specific recommendations for improving the course, focused on course content and the quality of lectures. This is not surprising: the instructions for the written comments ask students to focus on these topics.

Interestingly, recommendations were one of the two most frequent comment types when students made a negative remark, indicating that students are not simply complaining about the course but are also providing suggestions for how it could be improved. Similar to Ory and Piper's (1980) findings, these results indicate that written comments can provide suggestions for improving instruction that quantitative questions do not.

Comparison of students who do and do not write comments indicates that those who do write them provide more positive ratings of the instructor and the course: over half of the students who wrote comments rated the instructor as *excellent* and over forty percent rated the course as *excellent*. While some students who express very negative impressions of the course or

instructor do provide written comments, it appears that typically it is students who view the instructor and course most favorably who provide comments, contrary to findings by Theall and Franklin (1991) that the most and least satisfied students write comments.

Among subgroups by college for students who did and did not write comments, students taking architecture, communication, engineering, or public affairs courses were more likely to write comments than were students taking liberal arts courses. The schools of architecture and public affairs were each represented by one small graduate course, so the difference may have been due to the particular courses sampled. Most of the engineering courses sampled were upper-division, and students in those courses may have been more likely to write comments than students in general education, lower-division liberal arts courses. Students in communication courses may have been more likely to write comments as an artifact of characteristics typical of students enrolled in such courses.

Students in small- or medium-sized courses were more likely to write comments than were students in large courses. Because students in large courses may have less personal interaction with the instructor, their anonymity may influence their motivation to provide feedback to the instructor. There were no differences between undergraduate and graduate students in terms of how often they wrote comments, suggesting that maturity does not influence whether students will write comments. As students' workload ratings decreased from *excessive* to *insufficient,* students were less likely to write comments, indicating that when the workload is greater students may write more, possibly to complain or provide suggestions for how to make the workload more reasonable. Students who tended to write comments reported higher probable course grades and higher grade point average ranges, indicating that the better students in the class and overall may be more motivated to provide written comments than those who are not doing as well.

Students who wrote positive comments rated the instructor and course more positively than did those who wrote negative comments. Understandably, students who perceive the course more favorably will write comments that align with that perception. Students taking classes in business were less likely to write positive comments than were students taking classes in liberal arts. Nine of the twelve sampled business courses were upper-division or graduate courses. Courses in the college of business are known for high standards and stringent requirements, perhaps leading to fewer positive comments than typical general education, lower-division liberal arts courses.

Students in small- and medium-sized classes were more likely to write positive comments than were students in large classes, who may be less favorable about the course because of less personal interaction with the instructor. Additionally, many large classes use a traditional lecture-style format, and students may not rate this type of instruction and course as favorably as discussion-oriented courses that are more common with smaller class sizes. Because large classes may be required courses, students may have less interest in the content.

Students in undergraduate and graduate courses did not differ in writing positive comments, suggesting that undergraduates and graduates are equally likely to be favorable or critical of the instructor and the course. As workload rating decreased from excessive to light,

students were more likely to write positive comments, indicating that courses with lighter workloads do tend to receive more positive comments. Students with lower probable course grades were less likely to write positive comments, indicating that students' expected grades in the course can influence their decision to write positive comments. As GPA range increased, students were less likely to write positive comments, which is somewhat counterintuitive in that students with higher university GPA ranges might be expected to write more positive comments. However, this tendency may result from the desire of students with the highest GPAs to provide suggestions for improvement, because they are interested in getting the most out of their instructor and their courses.

Implications

It seems clear that students do put forth effort in providing instructors feedback about their skill as an instructor and the quality of the course. So, while students who write comments tend to be more positive about the instructor and the course and are most likely the students who will earn better grades in the course, good students are writing negative comments as well. This is an important finding, because it suggests there is more complexity to the pattern of students' comments than simply that "good" students write positive comments and "poor" students write negative ones.

When students provide negative comments, they often provide suggestions for improvement. While comment tone is generally positively correlated with GPA, grades, and other student achievement indicators, good students do write negative comments. Recommendations from such students can be a useful for instructors when they are constructive suggestions for improvement, not simply complaints about the instructor or course. As noted by Ory and Piper (1980), written comments provide specific information for improving instruction that quantitative information does not.

Students in large courses tend to write less often than do students in small- or medium-sized courses, indicating that smaller course sizes may impact the relationship students have with their instructor and possibly their motivation for providing written feedback. Instructors in large classes may need to spend time developing more direct interactions and soliciting feedback throughout the term to increase student motivation to write comments on the end-of-course evaluation.

Limitations

One limitation of the study is the generalizability of the conclusions beyond the sample drawn. While all colleges and most departments were represented in the stratified sample, the final sample of 103 course sections represented only about 1% of the approximately 8,000 possible course sections. Because sampling was based partially on stratification by college, some of the conclusions about colleges are tentative, where only one or a few courses were included in the sample. Additionally, large graduate courses were underrepresented in the final sample—because very few such courses existed—and small graduate courses were overrepresented. As a result, subgroup analyses by college, class size, and class level should be interpreted with caution. Additionally, response rate ranged from 30% to 100% with an average response rate of

67%. For courses with lower response rates, there may be non-response bias in that students who were not present or elected not to complete the course evaluations may have differed from those students who did complete the course evaluations. Despite these limitations, some significant patterns arose that help clarify the types of comments students make on end-of-course evaluations and the types of students making those comments.

Another limitation is that the instructions to students for providing written comments may have influenced the comment topics and tone. The instructions read, "In many ways your written comments can be the most important part of your evaluation of the course and instructor. In the space provided, please indicate what aspects of the course content and instruction were best, how the instructor could improve his or her teaching, and how the content of the course might be improved. The instructor will receive this form after the semester is over." Once students were asked to comment on course content and instruction, is it surprising that these were the two most frequent topics about which students wrote? Additionally, the instructions asked students to provide suggestions for improvement, which may have influenced students to focus their negative comments on recommendations. If given a more generic prompt, such as "Please comment on the course and instructor," students may have written on different topics and may not have provided as many suggestions for improvement.

With so little information on the nature of student written comments on end-of-course evaluations, this study helps shed light on the frequency and content of student comments. The results of this research can be used to determine how students who provide written comments differ from those who do not and how they rate courses and instructors. This study also provides insight into the relationship between how students rate courses and instructors and what the tone of their written comments is.

Areas for Future Research

Understanding why students choose to write or not write comments was not addressed in this study. While Theall and Franklin (1991) found that about 10% of students provided written comments, this research found that 40% of students did so, both in the study sample and across the university. Future research should explore what motivates students to write comments on end-of-course evaluations. Additionally, although this study examined the nature of written comments, it did not investigate how instructors use them. It seems that students tend to point out what works well and do make recommendations for course improvements. Future research to understand how faculty members use written comments would help determine the full impact of this type of feedback on improving the quality of teaching and courses.

References

Ory, J. C., Braskamp, L. A., & Pieper, D. M. (1980). Congruency of student evaluative information collected by three methods. *Journal of Educational Psychology, 72* (2), 181-185.

Theall, M. & Franklin, J. (1991). Using student ratings for teaching improvement. In M. Theall & J. Franklin (Eds.), *New Directions for Teaching and Learning, 48,* (pp. 83-96). San Francisco: Jossey-Bass.

Appendix A
Original and Final Sampling Plan

COLLEGE	FALL 2003 Undergrad Enrollment	Fall 2003 Undergrad and Graduate Enrollment	# of courses sampled	# of courses by size Small class: 1 – 35 Medium class: 36 – 100 Large class: 101+	# of Undergrad (U) and Graduate courses (G)	# of course by size (S, M, L) and level (U, G)	Actual courses sampled
Architecture	357 1.0%	613 1.2%	1 1%	S = 1	U = 1	SU = 1	SG = 1
Business Administration	4,408 11.5%	6,015 11.7%	12 12%	S = 4, M= 4, L = 4	U = 9 G = 3	SU = 3, MU = 3, LU = 3 SG = 1, MG = 1, LG = 1	SU = 3, MU = 3, LU = 2 SG = 2, MG = 2
Communication	3,601 9.4%	4,217 8.2%	8 8%	S = 3, M = 3, L = 2	U = 6 G = 2	SU = 2, MU = 2, LU = 2 SG = 1, MG = 1	SU = 2, MU = 2, LU = 2 SG = 2
Education	2,049 5.3%	3,454 6.7%	5 5%	S = 2, M = 2, L = 1	U = 3 G = 2	SU = 1, MU = 1, LU = 1 SG = 1, MG = 1	SU = 1, MU = 2, LU = 1 MG = 1
Engineering	5,517 14.4%	7,476 14.6%	14 14%	S = 5, M = 5, L = 4	U = 10 G = 4	SU = 3, MU = 4, LU = 3 SG = 2, MG = 1, LG = 1	SU = 3, MU = 4, LU = 3 SG = 4
Fine Arts	1,268 3.3%	1,909 3.7%	3 3%	S = 1, M = 1, L = 1	U = 2 G = 1	MU = 1, LU = 1 SG = 1	MU = 1, LU = 1 SG = 1
Information		301 0.6%	1 1%	M = 1	G = 1	MG = 1	MG = 1
Law		1,492 2.9%	1 3%	L = 1	G = 1	LG = 1	LG = 1
Liberal Arts	11,481 29.9%	13,263 25.8%	28 28%	S = 10, M = 9, L = 9	U = 24 G = 4	SU = 8, MU = 8, LU = 8 SG = 2, MG = 1, LG = 1	SU = 8, MU = 8, LU = 9 SG = 3, MG = 1
Natural Sciences	8,496 22.1%	10,085 19.6%	21 21%	S = 7, M = 7, L = 7	U = 18 G = 3	SU = 6, MU = 6, LU = 6 SG = 1, MG = 1, LG = 1	SU = 7, MU = 5, LU = 6 SG = 2, MG = 1
Nursing	701 1.8%	943 1.8%	2 2%	S = 1, M = 1	U = 1 G = 1	SU = 1 MG = 1	SU = 2 MG = 1
Pharmacy	271 0.7%	626 1.2%	1 1%	L = 1	G = 1	LG = 1	LG = 1
Public Affairs		283 0.6%	1 1%	S = 1	G = 1	SG = 1	SG = 1
Social Work	234 0.6%	617 1.1%	1 1%	M = 1	G = 1	MG = 1	SG = 1
Intercollegiate Programs		132 0.3%	1 1%	L = 1	G = 1	LG = 1	SG = 1
TOTAL	38,383 100.0%	51,426 100.0%	100 100%	S = 35, M = 34, L = 31	U = 74 G = 26	SU = 25, MU = 25, LU = 24 SG = 10, MG = 9, LG = 7	SU = 27, MU = 25, LU = 24 SG = 18, MG = 7, LG = 2

Appendix B
Code List

Category Code		Specific Code	
1	Grading	01	General
		02	Grading practices
		03	Assignments/papers/presentations/essays
		04	Opportunity to revise work
		05	Exams
		06	Quizzes
		07	Student recommendations
		08	
		09	
		10	
2	Course organization	11	General
		12	Course workload
		13	Format
		14	Pacing
		15	Organization of content
		16	Student recommendations
		17	
		18	
		19	
		20	
3	Course quality	21	General (includes general references to course/class)
		22	Activities and assignments
		23	Textbooks/Readings
		24	Value of course
		25	Course supplements (handouts, PowerPoint slides, video)
		26	Course content/material
		27	Student recommendations
		28	
		29	
		30	

4	Technology use	31	General
		32	Impact of using technology
		33	Instructor skill at using technology
		34	Need for technology in the course
		35	Quality of course technology resources (website)
		36	Student recommendations
		37	
		38	
		39	
		40	
5	Interpersonal Skills	41	General
		42	Acceptance of other viewpoints
		43	Approachability/Rapport
		44	Availability
		45	Class climate
		46	Enthusiasm
		47	Student recommendations
		48	
		49	
		50	
6	Teaching Skills	51	General (includes any general comments about instructor)
		52	Check understanding
		53	Clarity of presentation/explanations
		54	Knowledge of subject matter
		55	Oral communication skills (speaking skills, speech patterns, delivery)
		56	Encourage critical thinking
		57	Explain expectations clearly
		58	Identify important concepts
		59	Student motivation
		60	Quality of feedback provided to students
		61	Response to questions in class
		62	Quality of lectures/discussions (content, method, how info is conveyed)
		63	Staying on topic/use of class time
		64	Summarize information
		65	Use of examples/used real-world examples
		66	Uses variety of teaching techniques
		67	Student participation
		68	Student recommendations
		69	
		70	

7	Course goals/objectives	71	General
		72	Student recommendations
		73	
		74	
		75	
8	Course Impact	76	General
		77	Impact on future (occupation, changed major)
		78	Impact on learning
		79	Impact on personal growth
		80	Student recommendations
9	Other	81	Nonspecific/Catchall
		82	Physical appearance
		83	Physical environment (room/location)
		84	Time of class
		85	Student self-description
		86	Size of class (# of students)
		87	Comments about other students/classmates
		88	
		89	
		90	
10	Short Sentiments	91	Brief student expressions of feelings (‘Thank You’, ‘You suck’)
11	Drawings	92	Any drawing or figure (smiley face, heart, symbols)
12	TA	93	Comments about the TA, lab section, or discussion section
		94	TA grading practices
		95	
		96	
		97	
		98	
99	Unintelligible	99	Can’t read/incomplete thought

Tone of code:

Positive = 2
Negative = 1
Neutral = 0