

TITLE: Students Helping Students Provide Valuable Feedback on Course Evaluations

1. RESEARCH QUESTION(S) & WHY THEY ARE IMPORTANT TO THE FIELD:

Research Questions

1. Does the helpfulness of feedback that students provide to their instructors on the student evaluations of teaching (SETs) improve after participating in a peer-led videopresentation?
2. Do students find the peer-led video presentation and the feedback rubric useful in composing potent responses for their instructors?

Significance

One of the purposes of the SET instrument is to help instructors enhance the teaching and learning experience in their courses; however, student feedback can often be more unconstructive than useful because students are usually asked to evaluate instruction with little or no formal training. As a result, these evaluations become missed opportunities for students to effectively communicate their learning needs and for instructors to collect actionable information about how the course is perceived. This project aims improve the quality of student responses to open-ended questions that instructors receive by partnering with students in demonstrating to their peers the importance of SET and how to compose potent answers for instructors. This project provides a research and professional development opportunity for students and faculty and develops a rubric and video presentation that can be easily transferred to other programs and institutions to support teaching, learning, and course development.

2. DESCRIPTION OF RESEARCH DESIGN:

This study used a mixed method approach where quantitative and qualitative data were collected, analyzed separately, and reported in the results (Creswell and Creswell 2018).

Undergraduates from an assessment support program developed a peer-led presentation on the importance of the SET instrument and on how to leave detailed and useful feedback for instructors per a rubric (Appendix A). This peer-led presentation was delivered in 11 classes to 205 students in AY 2016-2017, midway through the Fall 2016 and Spring 2017 semesters, and at the end of the Fall 2016 semester, immediately before students completed end-of-semester SETs. Based on the positive impact resulting from this preliminary phase of the project (Appendix F), this peer-led presentation was turned into seven- and five-minute videos as a resource to be made available for instructors willing to participate in further research. The following research design description represents the study implementing these seven- and five-minute peer-led videos.

Group	Pre	Mid	Final
Semester(s)	Fall 2013 – Spring 2015	Fall 2018 - Spring 2019	Fall 2018 - Spring 2019
SET Responses	205 (10 courses)	250 (13 courses)	112 (10 courses)
Peer-led video presentations	None administered	7-mins video administered mid-semester	5-mins video administered before final evaluations
SET timing	Final, online	Mid-semester, paper	Final, online

The seven-minute video was delivered in 15 classes in AY 2018-2019, midway through the Fall 2018 and Spring 2019 semesters, and the five-minute video at the end of the Fall 2018 and Spring 2019 semesters, immediately before students completed the end-of-semester SETs. The mid-semester SETs were used for formative assessment only and comprised the “Mid” group. The “Final” group was composed of official SETs completed online at the end of the Fall 2018 and Spring 2019 semesters. Students in their courses completed a Mid and a Final SET immediately after the video presentations. Mid SETs were completed on paper, while Final SETs were completed online.

A total of 567 SET responses were gathered from 33 writing classes taught by six volunteering instructors during the academic years 2013-2014 through 2018-2019. The SET responses consisted of three open-ended questions from the official SET currently used by the writing program:

Q1. “How would you describe your writing ability now compared to the beginning of the semester?”

Q2. “Identify and evaluate aspects of this course that have been especially helpful to you.”

Q3. “Describe aspects of this course that you would change if you had the opportunity.”

SETs collected from 18 classes prior to AY 2018-2019 were used as controls (the “Pre” group).

Participating faculty defined helpful feedback and designed a rubric accordingly (Appendix A).

Faculty underwent a norming session prior to rating the quality of student response to these questions in their own SETs as “H” (highly useful), “S” (somewhat useful), or “N” (not useful).

Courses were matched across groups; each instructor submitted rated SETs for at least two sections of a given course in the Pre-group and another two in either the Mid and/or Final groups. Each section included up to 20 students.

Data were analyzed for group differences in the quality of feedback provided, in aggregate as well as by question and course division (upper or lower). Pearson’s chi-square or Fisher’s exact test were used to calculate significance as appropriate. A bias-corrected form of Cramer’s V was used to calculate effect size from frequencies.

To assess net change in feedback quality, scored sums were calculated by weighting responses such that highly useful responses ranged from 70-100% useful, somewhat useful from 1-69%, and not useful as 0%. H responses were weighted at 0.85, S responses at 0.35, and N responses at -0.2.

Findings were robust to a wide range of weights, of which the weights above represent the mean. The negative weight of N responses represents the negative utility of reading and sorting through

SETs that provide no actionable feedback. The percent change in scores was used to measure effect size.

Participating faculty and students' perception on the experience were collected through post surveys (Appendix D)

3. LITERATURE REVIEW & THE RELATIONSHIP OF THE LITERATURE TO YOUR RESEARCH QUESTION(S):

SETs for formative assessment

Since their inception in the 1920s, SETs are understood to be a vital part of a university since they provide faculty with valuable feedback that can help improve teaching practices and give students an opportunity to share their perspectives in order to improve student learning (Algozzine, Gretes, Flowers, Howley, Beattie, Spooner, Mohanty, & Bray, 2004; Clayson 2009). When student feedback does not indicate what students need and how the instructor can help, all parties that engage with SETs (students, instructors, and administrators) suffer.

SET Validity

Student evaluations can often be judgements about performance rather than insight into what the instructor can do to improve the student experience (Clayson 2009; Clayson & Haley 2011). Student evaluations have been shown to be poor at correlating with learning outcomes, addressing specific aspects of the course or teaching, and being objective about the instructor (Wieman 2015). Wieman points out that "it is difficult to know what you do not know." Students are not the best judges of their own learning, which results in feedback that does not address the questions on evaluations clearly (Wieman 2015; Lauer 2017), so the accuracy of this evaluation relies on the expertise of the respondent.

The ability to evaluate is a higher-order skill, and based on the literature on learning, the best ways to learn a skill (and giving feedback is a skill) is to observe a model (Bandura as cited in Svinicki 2001); however, real learning does not occur in one-time events but needs spacing and repetitions to move content into long-term memory (Thalheimer 2006).

A New Approach to Addressing SET Validity

Research by Clayson (2009) and Price, Handley, Milla., & O'Donovan (2010) supports that instructors need to invest time in specifically teach students about the feedback process, why it is important, how it is related to course evaluations, and model good examples.

Across the globe, Students as Partners programs (SaPs) place students, faculty, and staff as co-collaborators in the teaching and learning process (Mercer-Mapstone, et al., 2017, 1; Healey, Flint, & Harrington, 2014; Cook-Sather, Bovill, and Felten 2014).

If student partnerships between staff and faculty can be used to further subject-based research inquiry, curriculum design, pedagogic consultancy, and assessment (Healey et al. 2014), then why not partner with students to help improve the quality of feedback that instructors receive on SETs?

4. FINDINGS, INCLUDING THEIR SIGNIFICANCE & LIMITATIONS:

Key Findings

SETs improved on all questions from Pre to Mid based on aggregated data (Appendix B, table 1). However, controlling for class division reveals that this improvement was not uniform. Class division was a highly significant and moderately strong predictor of feedback quality, with upper division students providing more useful feedback ($p < .001$, $V = .212$). Additionally, upper division students' SETs improved significantly from Pre to Final ($p < .001$, $V = .424$), but not from Pre to Mid, except on Q3, where they improved moderately ($p < .027$, $V = .242$). The lack of improvement in upper division students' feedback from Pre to Mid stands in stark contrast to their sizeable improvements from Pre to Final. This may reflect a more serious attitude towards formal SETs at the end of the semester than to the informal mid-semester feedback.

Lower division students improved slightly from Pre to Mid on Q1 and Q2 (Appendix B, table 3), but not from Pre to Final on any of the questions or in aggregate. Their responses did not improve on Q3 in either the Mid or Final conditions.

Analysis of scored sums supports these findings. Upper division students provided more useful responses to all questions in all conditions (Appendix B, table 2), and they derived sizeable benefit from the video training while lower division students did not (Appendix B, table 3). The difference in quality of feedback across divisions may be the result of greater experience in college courses, or simply of survivorship bias; students in UD courses are far more likely to have opted to take the course.

A total of 276 students or 70% rated the video as highly effective and effective. Upper and lower division students rated their skill development equally. However, only students in UD courses performed better in their SET responses. Most of the students found the provided rubric useful (88% or 345 students) (Appendix C).

Most students (318 or 81%) recommend the video be delivered in other courses. A total of 174 students or 44% mentioned they prefer a peer-led video to a faculty-led video, while 86 or 22% has no preference in this regard.

Based on the faculty survey results, participating faculty would recommend the video to other instructors, and upon analyzing students' comments, faculty identified concrete ways to enhance their courses for content and instruction.

Limitations

- Pre SETs were gathered from years prior to the Mid and Final SETs, and thus we could not control for longitudinal effects such as changes in instructor skill over time. Because evaluations are anonymous, we could not control for the possibility that multiple SETs from a single student were included in analysis. Furthermore, some students may have participated in the peer-led training twice.
- No blinding procedure was implemented – faculty knew in advance which group data would belong to, and their expectations could have influenced the ratings they gave to their evaluations.
- SETs in the Mid group were completed with a limited amount of time and after students

had only completed the first half of the course.

- Question order was not randomized and thus might play a role in the difference in response usefulness across questions, because students may have left too little time to answer later questions.

Implications

- The peer-led video is a very effective training tool as-is when administered in UD courses just before final course evaluations.
- UD students have been exposed to more college teaching, and thus to a greater variety of course structures and activities. This could explain the exceptionally large difference in feedback quality between UD and LD students on Q3, which is the only question that requires students to draw upon experiences from outside the course. Q3 could be revised for LD students in the future.
- The difference in treatment response between divisions may be because the video gives them a mental model to integrate skills and knowledge that UD students already possess. However, LD students may not possess the same foundation and thus would not show the same improvement. Tailoring the video to LD students might elicit more useful feedback from them.
- Multiple exposures to the video content may improve feedback from LD students since it models good examples for them to follow.
- Peers could be significantly more effective at helping fellow students understand the purpose of SETs as well as why it is essential to complete them and compose thoughtful answers to open-ended questions; and therefore, potentially improve the validity of SETs.
- Faculty agreed that the study recognizes students' role in the SETs process to receiving more quality and useful answers to the open-ended questions on.

Further intervention and research

- Qualitative analysis of SETs may yield insight into why UD SETs were more useful than LD SETs, especially on Q3.
- The video was administered shortly before students completed their SETs. How long do improvements in skill at providing feedback acquired through the video persist?

REFERENCES

- Algozzine, B.; Gretes, J.; Flowers, C.; Howley, L.; Beattie, J.; Spooner, F.; Mohanty, G.; and Bray, M., (2004). Student Evaluation of College Teaching: A Practice in Search of Principles. *College Teaching*, 52 (4), 134-141.
- Clayson, D. E., (2009). Student evaluations of teaching: Are they related to what students learn? *Marketing Education Review*, 31 (1), 16-30.
- Clayson, D. E., Haley, D. A. (2011). Are students telling us the truth? A critical look at the student evaluation of teaching, *Marketing Education Review*, 22(2), 101-112.
- Clayson, D. E. (2006). Personality and the student evaluation of teaching. *Journal of Marketing Education*, 28(2), 149-160.

- Cook-Sather, A. B. (2014). *Engaging Students as Partners in Learning and Teaching. A Guide for Faculty*. San Francisco: Jossey-Bass.
- Creswell J. W., and Creswell J. D. (2018) *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches*. Fifth Edition. Los Angeles, CA: SAGE.
- Felton, J. M. (2004). Web-based student evaluations of professors: the relations between perceived quality, easiness and sexiness. *Assessment and Evaluation in Higher Education*, 29(1), 91-108.
- Fraile, R. and Bosch-Morell F., (2014) Considering teaching history and calculating confidence intervals in student evaluations of teaching quality. An approach based on Bayesian inference. *Higher Education* 70, 55-72.
- Healey, M., Flint, A., Harrington, K. (2014) Engagement through partnership: students as partners in learning and teaching in higher education. *The Higher Education Academy* (July)
- Lauer C. (2017) A comparison of faculty and student perspectives on course evaluation terminology. *To Improve the Academy*, 31(1)
- MacNell, L. D. (2014, August). What's in a name: exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291-303. Retrieved from <https://doi.org/10.1007/s10755-014-9313-4>
- Mercer-Mapstone, L., Dvorakova, L.S., Matthews, K.E., Abbot, S., Cheng, B., Felten, P., Knorr, K., Marquis, E., Shammas, R., & Swaim, K. (2017). A Systematic Literature Review of Students as Partners in Higher Education. *International Journal for Students as Partners*, 1 (1).
- Price, M., Handley, K., Millar, J., and O'Donovan, B. (2010). "Feedback: All That Effort, but What is the Effect?" *Assessment and Evaluation in Higher Education* 35 (3): 277-289.
- Povlacs J. T., (1984) Reading students' written comments on evaluations of teaching. *To Improve the Academy*, 3(1) DOI: 10.1002/j.2334-4822.1984.tb00055.x
- Spooren, P. B. (2013). On the validity of student evaluation of teaching: the state of the art. *Review of Educational Research*, 83(4), 598-642.
- Wieman, C. (2015) A Better Way to Evaluate Undergraduate Teaching, *Change: The Magazine of Higher Learning*, 47:1, 6-15, DOI: [10.1080/00091383.2015.996077](https://doi.org/10.1080/00091383.2015.996077)
- Svinicki, M. (2001) Encouraging Your Students to Give Feedback. *New Directions for Teaching and Learning*, 87, 17-24, p. 20.
- Thalheimer, W. (2006, February). *Spacing Learning Events Over Time: What the Research Says*. Retrieved June 31, 2018, from <http://www.work-learning.com/catalog/>

Appendix A: Rubrics

Instructors collected and rated the usefulness of student comments according to the following criteria:

Highly Useful	I clearly understand the experience the student is having, what I am doing well, or what I could do better. I know what I should continue doing in this class, and exactly what I can do to improve my course and/or instruction. Any improvements that need to be made are plausible and are within my control
Somewhat Useful	I have a general or vague idea of what is going well or what I should change to improve my course, but it is not completely clear. I can make a change to my course or instruction, but I may not get the result this student is looking for. I may not have the ability to completely make this change
Not Useful	I don't know what I can do to improve my course at all based on this answer. It tells me nothing about my class or pedagogy. I can't tell if the student is having a positive learning experience or negative experience, and/or exactly why. I have no control over making this change
0	No response.

Students receiving feedback training were given the following instructions and rubric:

You are welcome to address any aspect of the course you wish, but I would particularly appreciate your feedback about the following:

- Giving and attending to feedback
- Analyzing readings
- Developing a topic
- Composing an argument and integrating evidence
- Crafting an essay

How to Provide Valuable Feedback on Course Evaluations

Criteria	Highly Useful	Somewhat Useful	Not Useful
1. Offer commentary on attributes of the learning environment.	"I find the instructor very caring and that motivates me to try harder in this class"	"The instructor cares about my learning."	"My instructor's hair is cool."
2. Answer all parts of the question focusing on description rather than judgment.	"My writing ability now is better than at the beginning because now I am more confident in my work and writing based on the feedback I received from instructor and peers."	"It improved a lot. I noticed that my critical thinking ability has improved a lot."	"Hard class."
3. Attribute positive or constructive feedback to specific aspects of the course. Use examples that support your answer to the question.	"Before this class I was every unsure on how to do a research paper, now that I have taken the class I am more confident in my writing skills. I understand how to format a research paper correctly and how to follow MLA."	"Instructor sometimes describes things unclearly, but I always ask questions if I am confused about anything."	"Research projects are stressful"
4. Focus on the course and the quality of instruction given regarding the course learning outcomes.	"I loved the projects, in particular group discussions were very important to understand the readings."	"Peer review, presenting, and office hours helped me with learning."	"I wish that Cat Courses told us when assignments are due"
5. Offer suggestions that are relevant and plausible to the course or instruction and why you think they would help your learning.	"If I had the opportunity, I would include more journal writings or just open ended writing assignments so students could grow more."	"I wouldn't change anything."	"This class is too early."

Appendix B: Findings from Faculty-rated SETs

Table 1. Differences in Response Quality by Condition and Question

Question	Pre vs. Mid		Mid vs. Final		Pre vs. Final	
	<i>p</i> <	<i>V</i>	<i>p</i> <	<i>V</i>	<i>p</i> <	<i>V</i>
Q1	.001	.168	.001	0.00	-	-
Q2	.001	.302	.001	0.00	-	-
Q3	.001	.192	.001	.169	.02	.128

Pre vs. Final calculations were not computed when Mid and Final did not differ significantly.

Table 2. Scored Sums of Response Quality by Condition, Question, and Division

Division	Question	Pre	Mid	Final
UD	AllQuestions	54.5%	64.9%	68.3%
	Q1	50.8%	62%	69.5%
	Q2	60.6%	70.5%	76.9%
	Q3	51.9%	62.2%	58.5%
LD	AllQuestions	43.1%	56.5%	59%
	Q1	43.0%	50.9%	59.7%
	Q2	54.4%	70.4%	62%
	Q3	32.1%	48.2%	59%

Table 3. Differences in Response Quality by Condition, Question, and Division

Division	Question	Pre vs. Mid		Pre vs. Final		
		<i>Increase</i>	<i>V</i>	<i>Increase</i>	<i>V</i>	
UD	AllQuestions	10.4%	0.320	13.8%	0.424	***
	Q1	11.2%	0.224	18.7%	0.504	***
	Q2	9.9%	0.218	16.3%	0.343	***
	Q3	10.3%	0.242	*	6.6%	0.312
LD	AllQuestions	13.4%	0.093	**	15.9%	0.000
	Q1	7.9%	0.048		16.7%	0.000
	Q2	16%	0.180	***	7.6%	0.056
	Q3	16.1%	0.044		26.9%	0.000

p* < .05, *p* < .01, ****p* < .001. Significance and Cramer's *V* were based on raw data while percent increases

were calculated from scored sums. UD Q3 Pre vs. Mid differed significantly because feedback quality became more polarized (more responses scored as H or N).

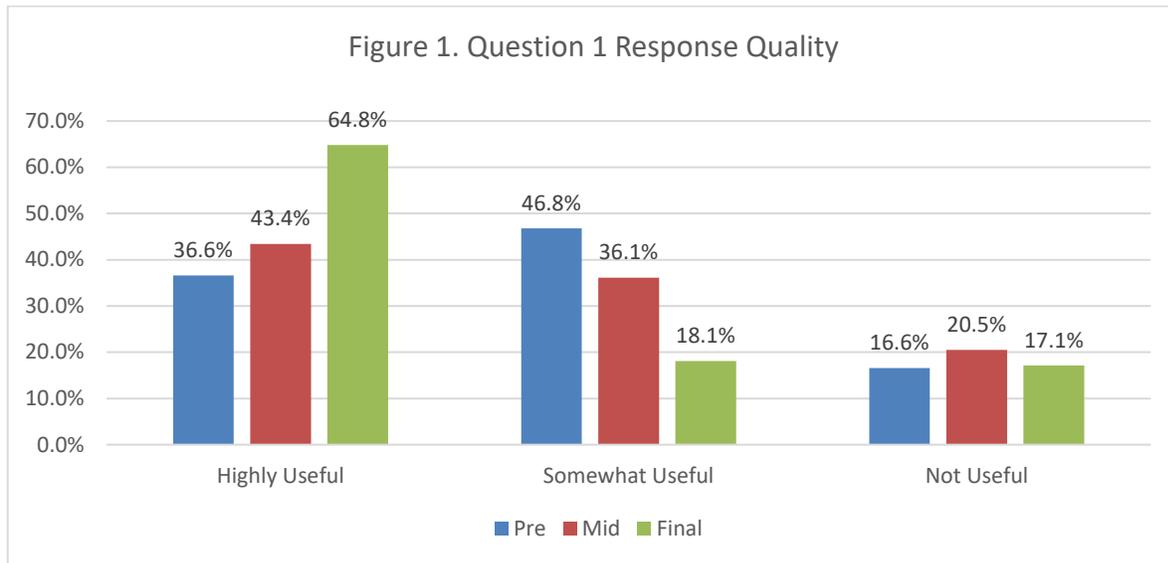


Figure 1 displays results for question 1: “How would you describe your writing ability now compared to the beginning of the semester?” The difference in feedback quality from Pre to Mid was significant ($p < 7.17E-05$, $V = 0.168$). The feedback received in the Final condition was not significantly different from in the Mid condition.

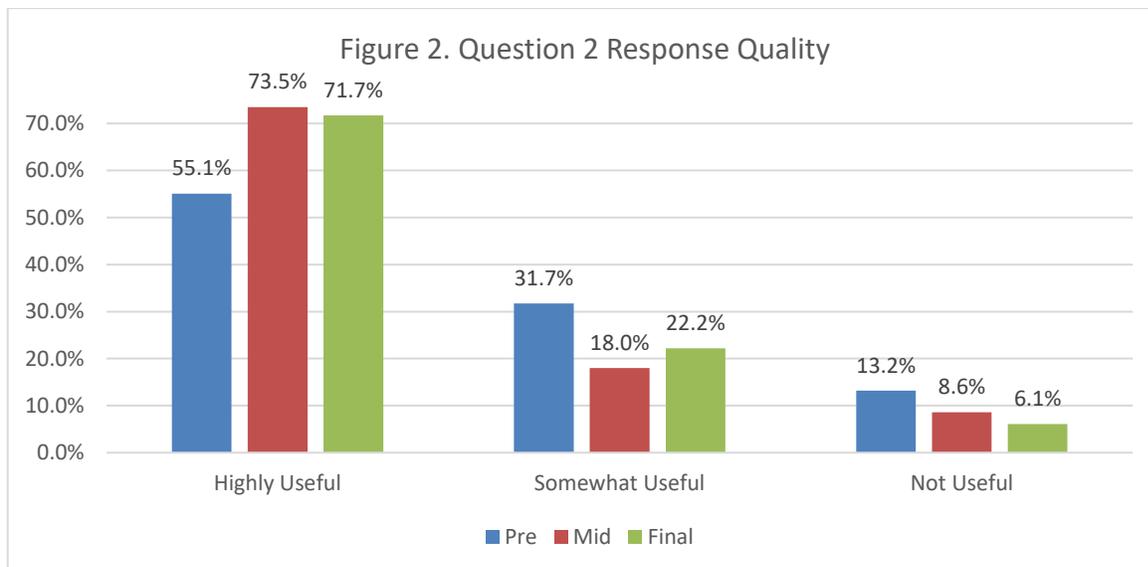


Figure 2 displays results for question 2: “Identify and evaluate aspects of this course that have been especially helpful to you.” The distribution of feedback across the three categories of usefulness changed significantly from Pre to Mid ($p < 1.16E-15$, $V = 0.302$). Feedback did not differ significantly between the Mid and Final groups.

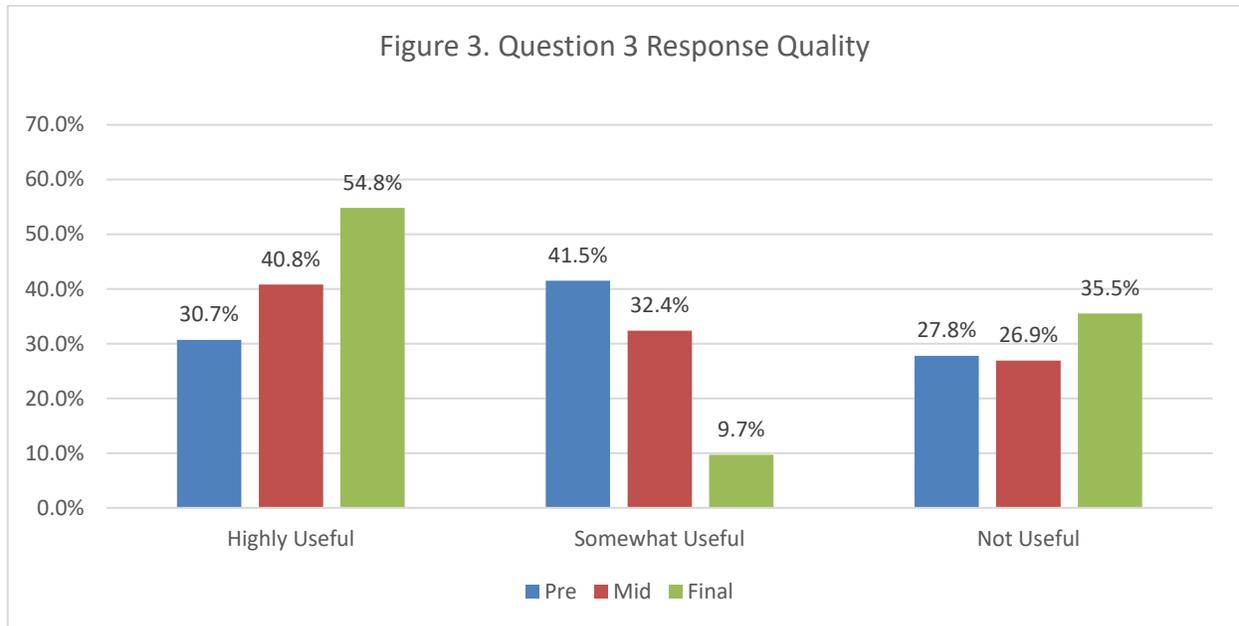
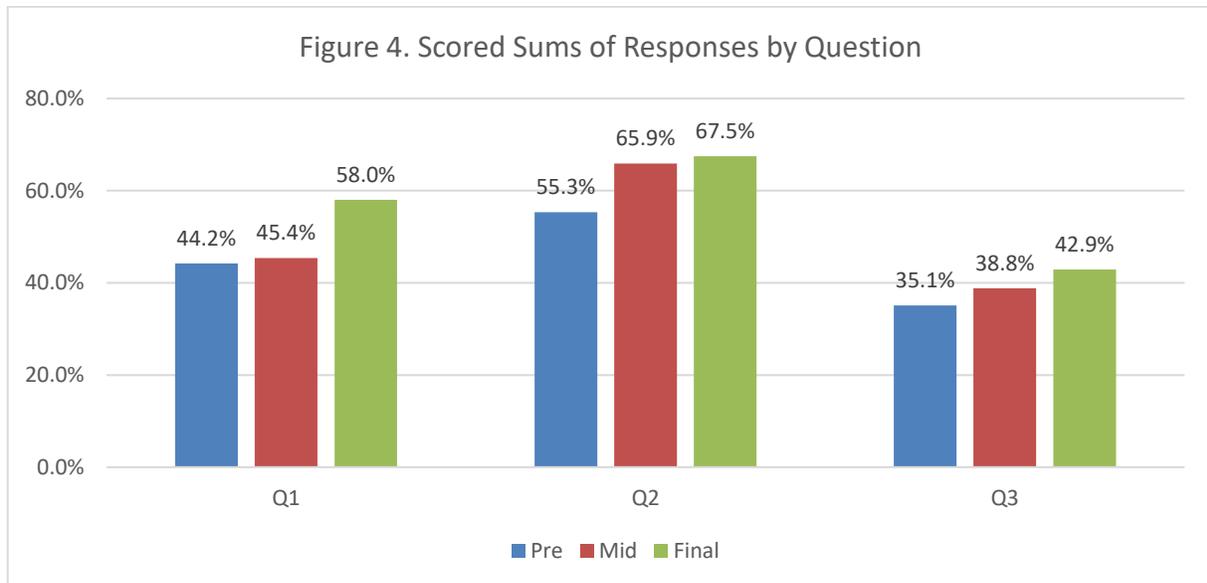


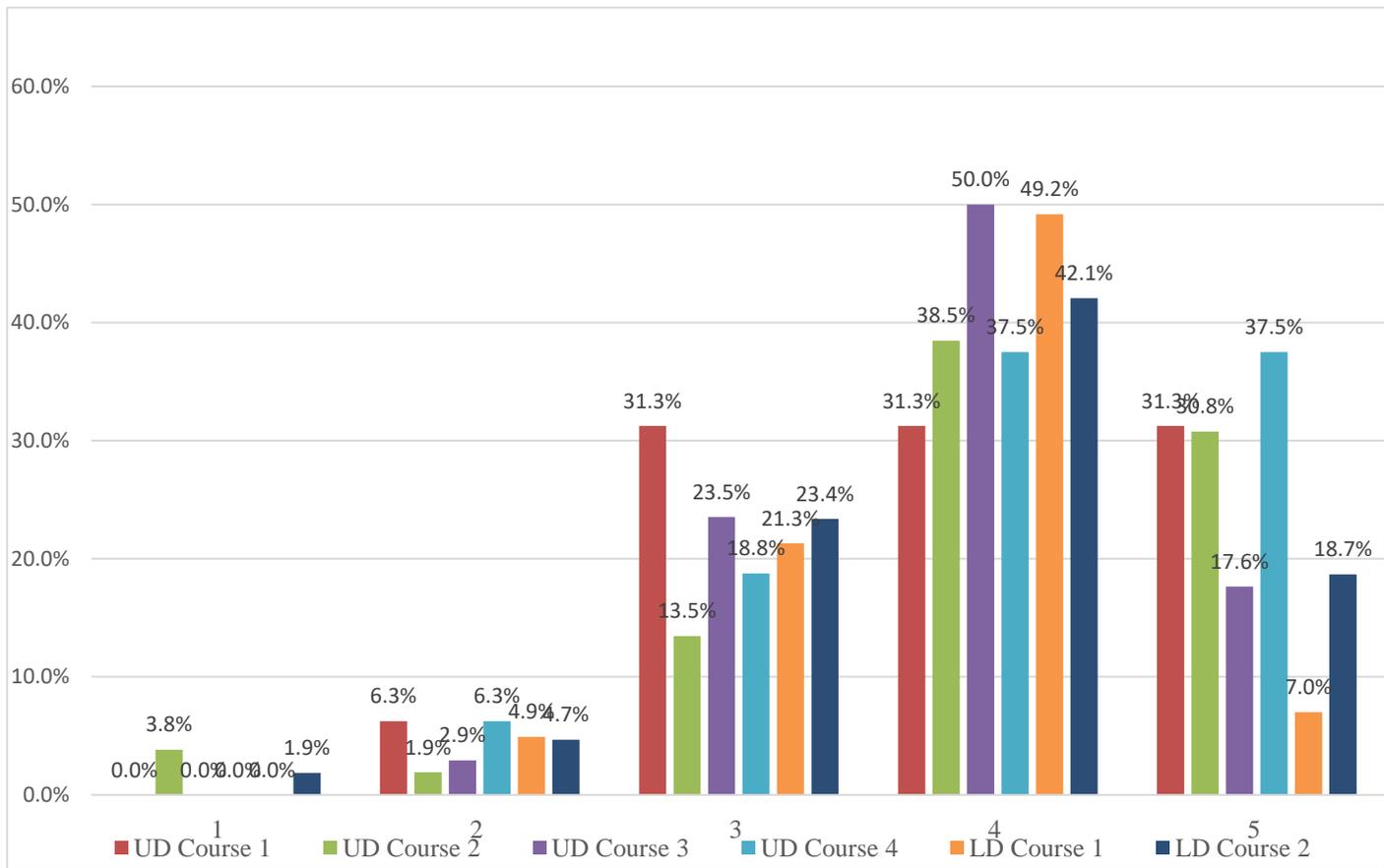
Figure 3 displays results for question 3: “Describe aspects of this course that you would change if you had the opportunity.” Faculty received a higher proportion of helpful feedback immediately following the video presentation ($p < 5.12E-06$, $V = 0.192$). Notably, the results for mid and final SETs were significantly different from each other ($p < 2.24E-23$, $V = 0.169$, unlike in the previous two questions).



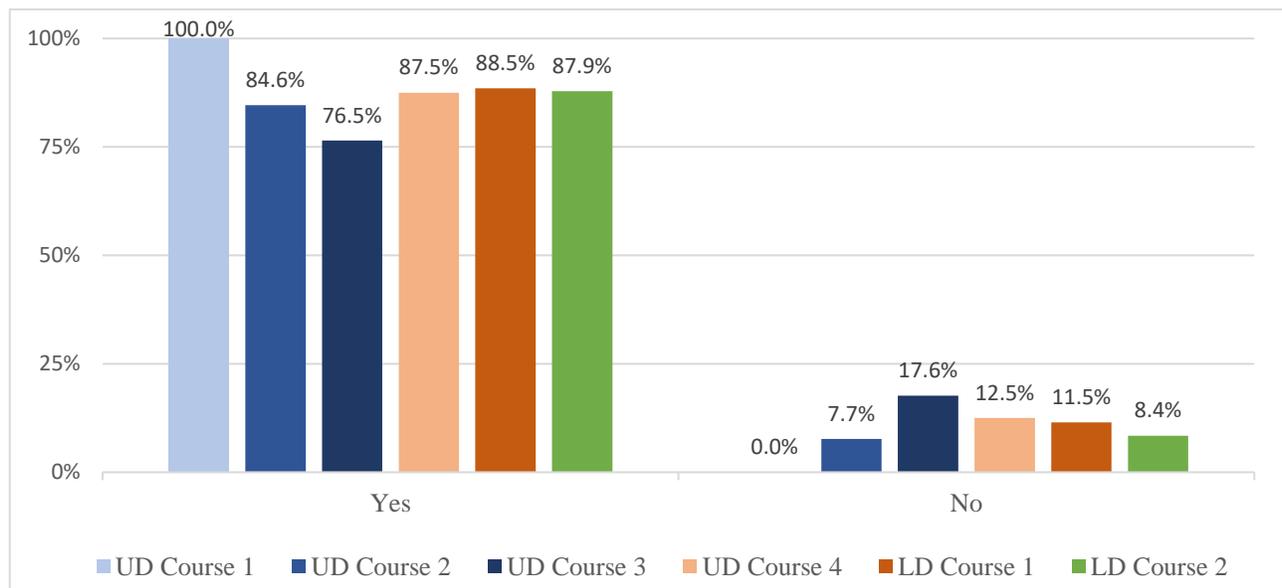
Feedback was scored by quality using a range of point values for H, S, and N and summed to create an aggregate score. Results were fairly similar across the entire range of reasonable values; moderate values of 0.85 for H, 0.35 for S, and -0.2 for N were used for Figure 4 above and subsequent analyses. As such, the maximum score is 85%.

Appendix C: Student Input on the Video Presentation

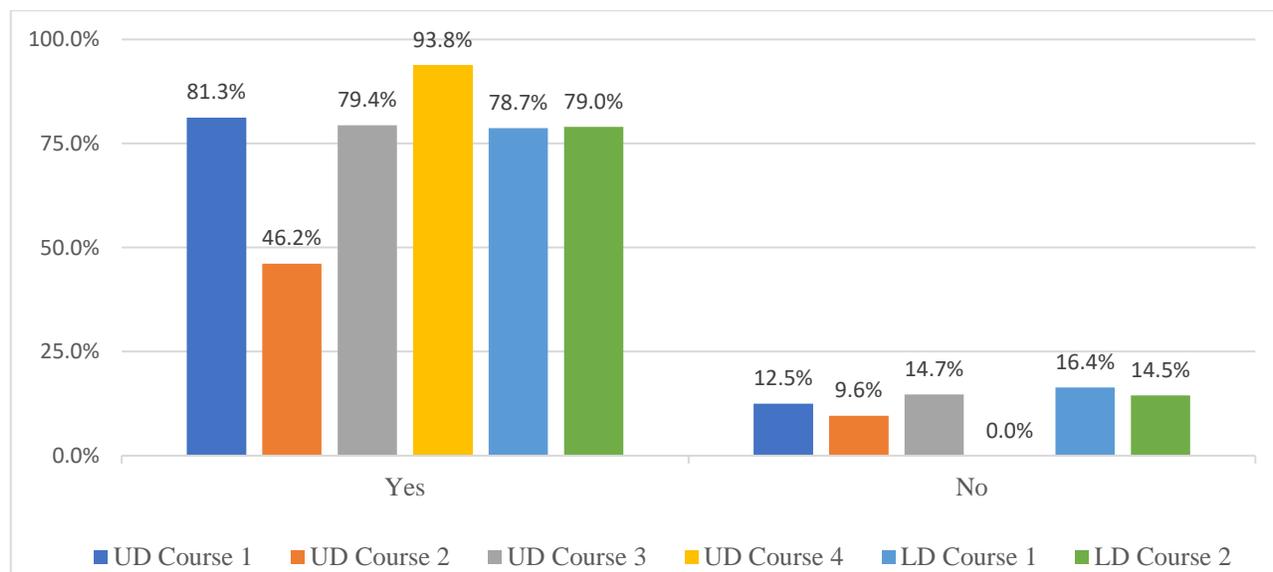
On a scale of 1-5 (5 being the highest), how would you rate your skills at providing valuable feedback on course evaluation after this presentation compared to before?



Question 2: Do you find the rubric useful?



Question 3: Would you recommend that this presentation should be facilitated in other courses?



Appendix D: Post Surveys for students and faculty

Videp Presentation: Post survey (for students)

-

About the video presentation ...

- On a scale of 1-5 (with 5 being the highest), how would you rate your skills at providing valuable feedback on course evaluation after this peer-led presentation compared to before the presentation?
1- 2- 3- 4- 5
Please explain why:
- Do you find the rubric useful? Yes/ No
If so, which criteria was the most useful.
If not, why not?
- Would you recommend any changes to the rubric? Please explain.
- How effective do you find the peer-led presentation? What suggestions do you have to improve this video presentation?
- Would you recommend that this peer-led presentation should be facilitated in other courses?
Yes _____ No _____
Please explain
- If the peer-led presentation was offered by faculty instead of students, what would have been different? Why?

Post Survey for Faculty

1. Do you recommend this "peer to peer" video presentation to other faculty prior to administering their mid or final course evaluations? Please explain.
2. For this question, please indicate whether you are referring to a lower division or upper division course: Did you come away with ideas for concrete changes to make after reflecting on the student feedback you received in your final course evaluations? If so, what changes did you implement? Please give at least one specific example.
3. Overall, did you notice any difference between the feedback you received from students after they participated in this project? No, in which way was it the same? If yes, in which way was it different?

Preliminary Phase: Peer-Led Presentations Results

Appendix E: Findings from Faculty-rated SETs

Table 1. Differences in Response Quality by Condition and Question

Question	Pre vs. Mid		Mid vs. Final		Pre vs. Final	
	<i>p</i> <	<i>V</i>	<i>p</i> <	<i>V</i>	<i>p</i> <	<i>V</i>
Q1	.001	.219	0.66	0.00	-	-
Q2	.007	.154	0.60	0.00	-	-
Q3	.019	.140	.005	.173	.02	.128

Pre vs. Final calculations were not computed when Mid and Final did not differ significantly.

Table 2. Scored Sums of Response Quality by Condition, Question, and Division

Division	Question	Pre	Mid	Final
UD	AllQuestions	54.5%	61.1%	41.6%
	Q1	50.8%	61.8%	78.1%
	Q2	60.6%	71.4%	78.0%
	Q3	51.9%	50.0%	73.9%
LD	AllQuestions	43.1%	53.6%	43.7%
	Q1	43.0%	56.2%	44.3%
	Q2	54.4%	62.9%	57.1%
	Q3	32.1%	41.6%	29.6%

Table 3. Differences in Response Quality by Condition, Question, and Division

Division	Question	Pre vs. Mid		Pre vs. Final		
		<i>Increase</i>	<i>V</i>	<i>Increase</i>	<i>V</i>	
UD	AllQuestions	11.9%	0.064	41.6%	0.429	***
	Q1	21.6%	0.224	53.8%	0.504	***
	Q2	17.8%	0.218	28.7%	0.347	***
	Q3	-3.7%	0.235	*	42.3%	0.312
LD	AllQuestions	25.4%	0.144	***	0.1%	0.000
	Q1	17.8%	0.203	**	3.0%	0.000
	Q2	15.7%	0.152	*	5.1%	0.056
	Q3	29.7%	0.079		-7.6%	0.000

* $p < .05$, ** $p < .01$, *** $p < .001$. Significance and Cramer's V were based on raw data while percent increases

were calculated from scored sums. UD Q3 Pre vs. Mid differed significantly because feedback quality became more polarized (more responses scored as H or N).

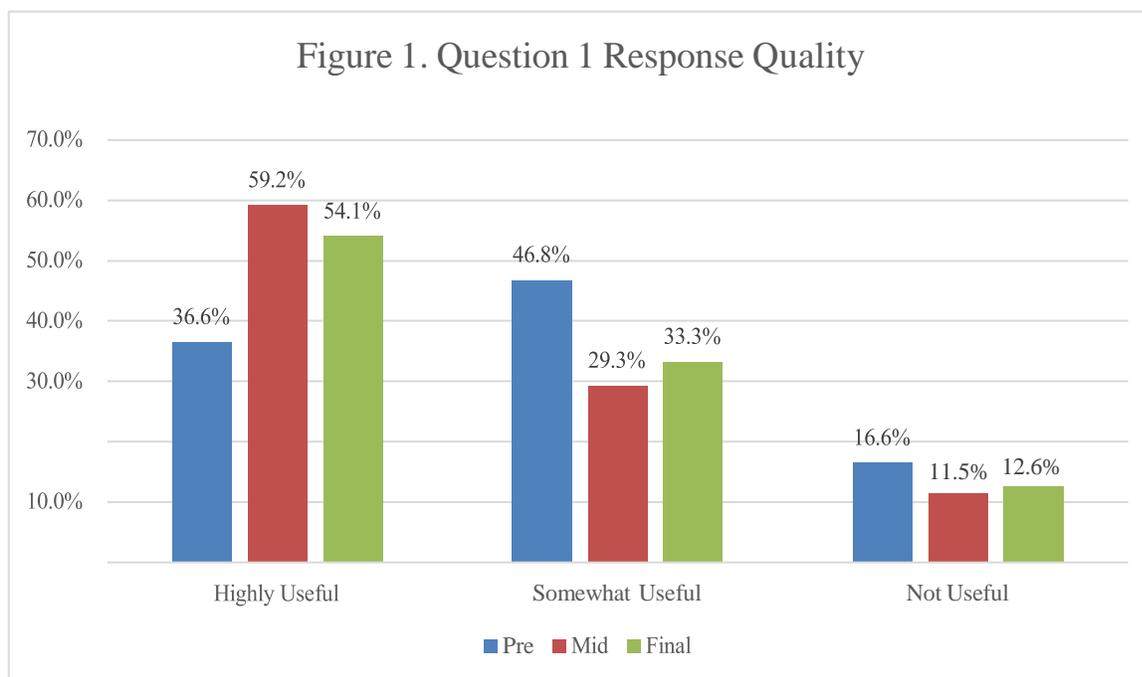


Figure 1 displays results for question 1: "How would you describe your writing ability now compared to the beginning of the semester?" The difference in feedback quality from Pre to Mid was significant ($p < 0.001$, $V = 0.22$). The feedback received in the Final condition was not significantly different from in the Mid condition.

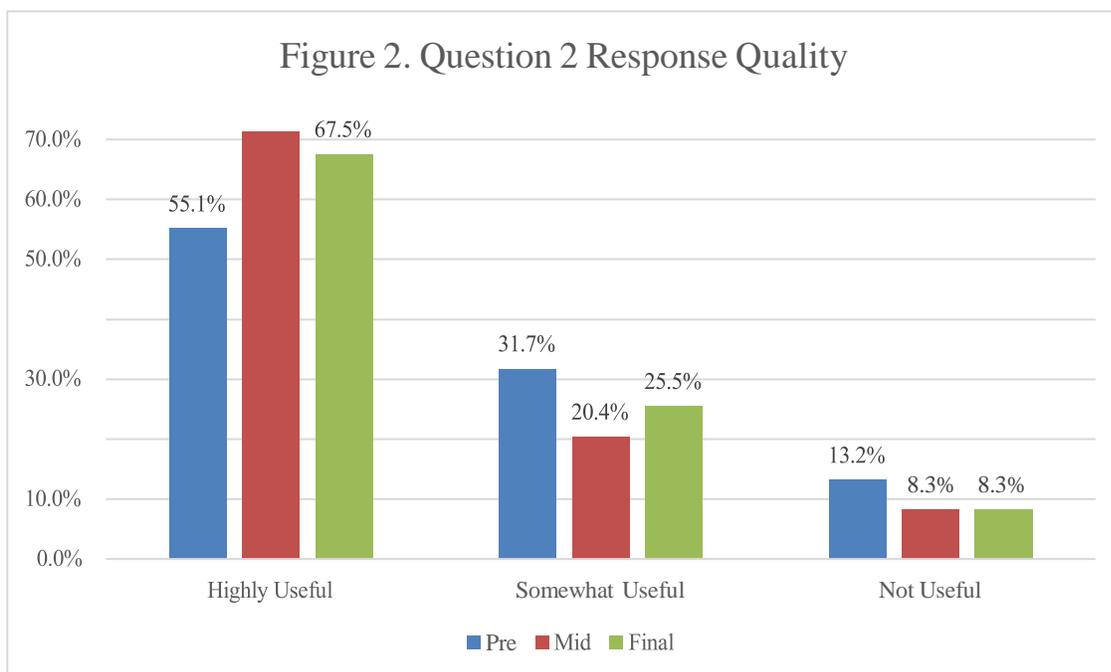


Figure 2 displays results for question 2: “Identify and evaluate aspects of this course that have been especially helpful to you.” The distribution of feedback across the three categories of usefulness changed significantly from Pre to Mid ($p < 0.007$, $V = 0.15$). Feedback did not differ significantly between the Mid and Final groups.

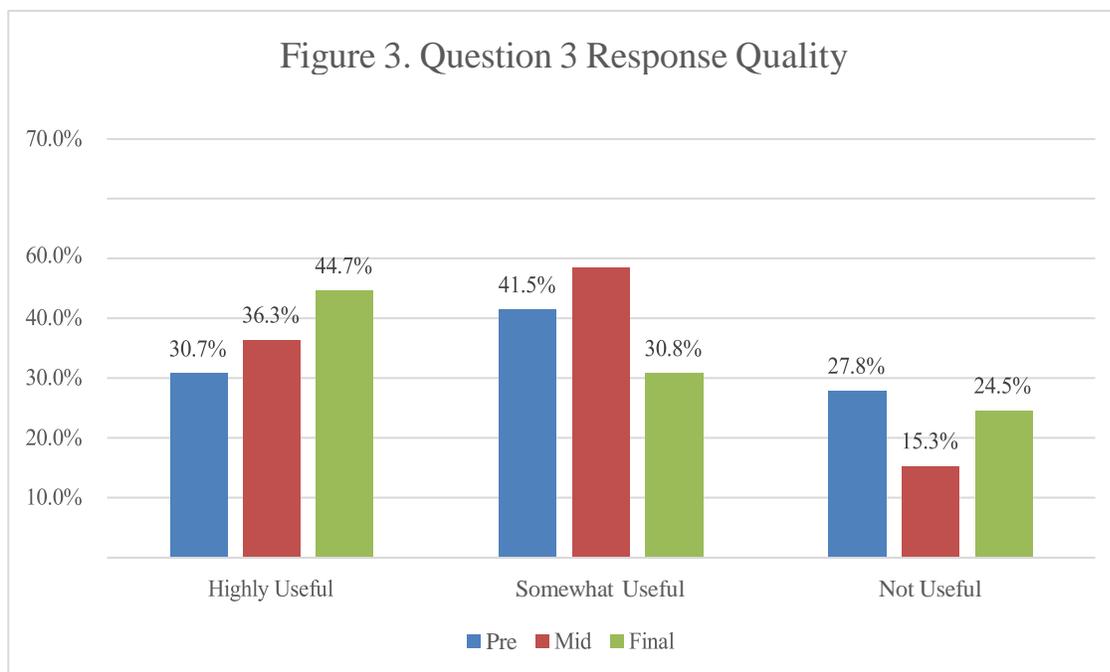
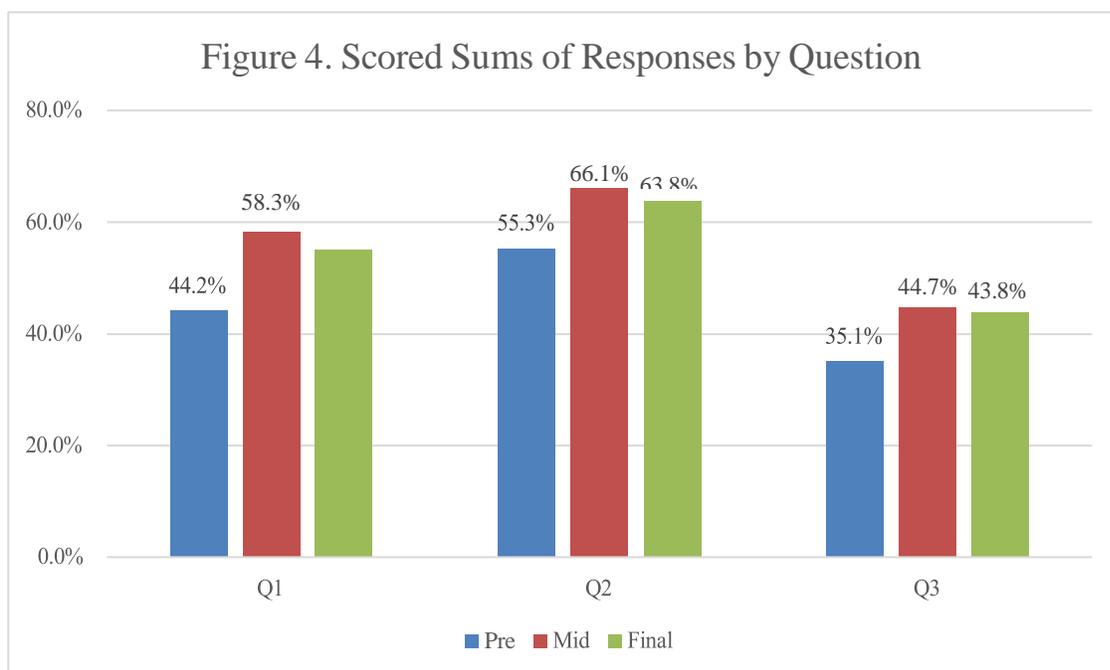


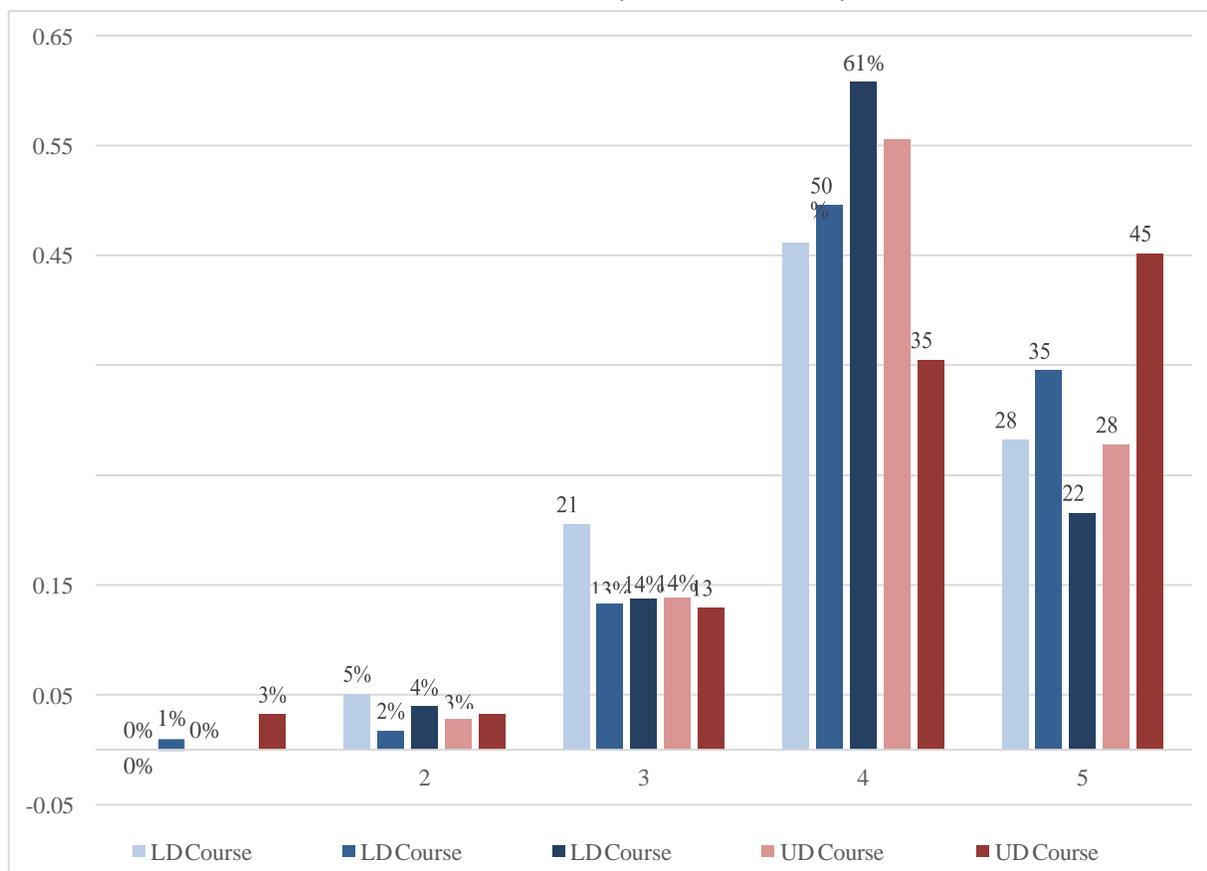
Figure 3 displays results for question 3: “Describe aspects of this course that you would change if you had the opportunity.” Faculty received a higher proportion of helpful feedback immediately following the live presentation ($p < 0.018$, $V = 0.14$). Notably, the results for mid and final SETs were significantly different from each other ($p < 0.004$, $V = 0.17$), unlike in the previous two questions.



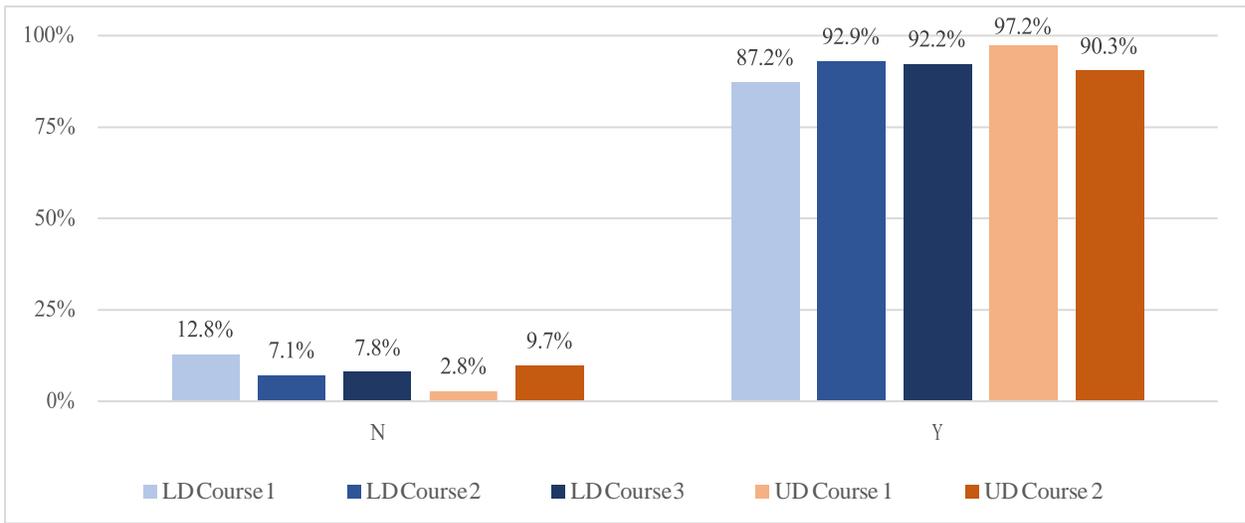
Feedback was scored by quality using a range of point values for H, S, and N and summed to create an aggregate score. Results were fairly similar across the entire range of reasonable values; moderate values of 0.85 for H, 0.35 for S, and -0.2 for N were used for Figure 4 above and subsequent analyses. As such, the maximum score is 85%.

Student Input on the Peer-Led Presentation

Question 1: On a scale of 1-5 (5 being the highest), how would you rate your skills at providing valuable feedback on course evaluation after this presentation compared to before?



Question 2: Do you find the rubric useful?



Question 3: Would you recommend that this presentation should be facilitated in other courses?

