Making Sense of Course Evaluations and Midterm Feedback from Students:
A Quick Guide for Instructors

Office of Assessment of Teaching and Learning, WSU

Contents:

1. Introduction

By using a few relatively simple methods of analysis, the results of course evaluations and midterm feedback from students can provide valuable information to support you in your teaching. As with any other source of data, course evaluations have limitations. They can be a useful tool for collecting student input on their own learning experiences in their courses. Students are uniquely qualified to provide feedback about their individual experiences of instructors and courses but not necessarily to provide an *evaluation* of the instructors and courses (despite the commonly used term "course evaluation"). Student feedback should be treated as one among several other potential sources of information for instructor and course evaluation. Understanding how student experience translates into usable data for improving teaching and learning requires some work to make sense of results. This quick guide offers a few methods for individual instructors to organize, analyze, and interpret results.[1]

2. Making Sense of Results

A quick read-through of your course evaluation results will give you an initial overview, but don't stop there: quick read-throughs can leave an impression that one aspect dominates out of proportion to its real significance—a particularly strong comment from one student, for example, can color your reading of the entire set of results. Taking a more purposeful approach can help you put the different aspects of the results in perspective and find usable information for your teaching. In general, analyzing the results means looking for patterns in responses and asking yourself questions, such as *why, how, when,* and *in what ways is this connected to learning, assignments*, *etc.?*

2. A. Making Sense of Likert-scale Responses

When examining responses to Likert-scale questions (questions with a rating scale such as Strongly Agree/Agree/Disagree), look for patterns. For example:

- Combine positive and negative response groups: What's the combined percentage of positive choices (i.e., agree and strongly agree) and of negative choices (disagree and strongly disagree)

- Are there two adjoining response groups that total more than 50% of the students?

- What are the two largest responses groups?

---

[1] Course evaluations/student ratings have been the subject of more than 50 years of research into their validity. While studies can be found demonstrating sometimes contradictory findings, one recent summary of the research and literature found that the research overall argues for the validity of student ratings: student ratings are "significantly and consistently related to student achievement, teacher self-ratings, administrator and colleague ratings, ratings by trained observers, and student written comments." The IDEA Center (2012). Student Ratings of Teaching: A Summary of Research and Literature. Retrieved from http://ideaedu.org/research-and-papers/idea-papers. Manhattan, KS: The IDEA Center. See Page 5 for quotation.

LIKERT SAMPLE

The following elements of the course helped me learn:					
	Strongly Agree	Mildly Agree	Neutral	Mildly Disagree	Strongly Disagree
Group work	**54% (228)**	34% (142)	8% (32)	3% (11)	2% (7)
Textbook and other readings	**51% (214)**	37% (154)	8% (35)	3% (11)	1% (6)

Interpreting this data:

See where the two largest response categories are. In this case, they also adjoin. So these can be grouped, allowing you to say that 88% of the students responding agree that the group work helped them learn. Likewise you can say that only about 5% of students responding disagree.

When using numerical values assigned to Likert categories such as Strongly Agree (4)/Agree (3)/Disagree (2)/Strongly Disagree (1), be aware that applying means, variances, or standard deviations may be statistically inappropriate and therefore can create a misleading analysis of your student rating results. While numbers assigned to Likert categories convey "greater than" or "less than" relationships, the differences between values are not necessarily constant. The difference in value between Strongly Agree and Mildly Agree and between Mildly Agree and Neutral, for example, are not clear nor is there a shared understanding of these values among raters.[2] For this reason, it's best to avoid averaging numerical values assigned to Likert categories; however, it would be appropriate to calculate the median (numerical middle point of scores) or the mode (the most common number or value).

Note: The distribution of responses to Likert-scale questions on course evaluations are typically skewed (i.e. the majority of responses are mildly agree and strongly agree, but there are a few mildly disagree or strongly disagree responses) as in the above example. Due to the skewness of the distribution, the mean misrepresents the actual ratings by portraying lower ratings than actually occurred as the mean is drawn toward the few low ratings. On the other hand, the median is not influenced by these few low ratings and remains in the middle.

---

[2] Mitry, D.J. and Smith, D.E. (2014) Student evaluations of faculty members: A call for analytical prudence. *Journal on Excellence in College Teaching*, 25(2), 56-67

2. B. Making Sense of Check-box Data

Check-box data offers another opportunity to look for patterns and what kind of combinations or trends they might show, but be aware of over-interpreting or assuming causal relationships. Also, look for results that might serve as entrees into conversations with students.

CHECK-BOX SAMPLE

What are your approaches / habits in this large class setting? (Please indicate what is true for you.)

	Response Percent	Response Count
I attend classes regularly.	92%	390
I tend to skip classes more often in a large class than a small class.	8%	34
I sit in the front row.	34%	162
I sit in the back row.	17%	72
I find a friend to sit with.	59%	251
I often feel lost.	5%	21
I generally come to the professor's office hours, to establish a personal connection.	4%	19
I feel pretty disconnected from the professor.	17%	70
I take notes.	87%	369
I try to study with friends.	43%	181
I tune out, no matter how hard I try to pay attention.	7%	31
Honestly, I have a hard time staying awake.	11%	45
When I'm not sure about something, I ask the professor questions after class or in office hours.	17%	72
When I'm not sure about something, I ask the TA later.	15%	64
I find something on my own (online or elsewhere) to help me understand the material.	28%	119

Interpreting this data:

What questions arise from these results? Consider, for example, that the results suggest that students are attending class, sitting with friends, and taking notes. How could they make the most out of those habits, and how could the instructor leverage them? Partway through a class, could you ask students to work with their neighbor to identify the main point or core concept of the first portion of the class period, or connect it to the readings? At the end of class, could students compare notes and have a chance to ask questions based on discrepancies, or could they write a review question together to submit for a class study guide? Also, what are they actually *doing* when they say they "study with friends"? What suggestions could they offer other students who are studying together, as it's likely that not all groups are equally effective? (Next semester could you add a question asking the students how they prepared for the midterm exam, specifically what they did, and share those results with the class, combined with grade?) Similarly, could you find a place in Blackboard or elsewhere for students to share helpful resources they are finding on their own, online or other?

Most students aren't feeling too lost or disconnected in this large class. Good! What could be contributing to that? Awareness of what you and others are doing can help you maintain and replicate this positive situation in the future.

2. C. Making Sense of Open-box Responses (Comments)

As qualitative data, student comments need to be sorted and analyzed like any other data to make sense of them and to get the most value out of them. As is true for the results as a whole, comments should always be considered in conjunction with other sources of information. Many instructors find the comments to be the most useful part of student feedback in supporting improvements to teaching and courses. At the same time, they can be difficult to interpret at first glance. Because end-of-term ratings are anonymous for students, sometimes they feel free to leave a nasty comment that is not well-considered or a broad laudatory comment without specifics—and sometimes the results can contain both negative and positive comments about the same topic, leaving the instructor wondering how to make sense of the seeming contradiction. Additionally, extreme comments, both good and bad, can have an emotional impact on an instructor out of proportion to their value in terms of evaluation. Consider using some of the tips below to organize and analyze comments so that you can make sense of them.

- For very large classes, it may be prohibitive to read every student comment. One approach is to read the comment feedback from a random sample of students. For a 95% confidence interval with an error level of 10%, choose the following sample sizes: for a class of 100, randomly select 49; for a class of 200, select 65; for a class of 400, select 78.

- Be aware that the comments may or may not be representative of students in the class. Typically only about 10% of respondents leave comments.[3] While several research studies show a strong correlation between student written comments and the global instructor rating,[4] one university study found that it was typical for their students with strong feelings (positive or negative) to respond in greater number to open-ended questions.[5]

- In general, don't give a lot of weight to "stray" responses, feedback which is very unusual (in research, we would consider these "outliers"). You can ask yourself if there is something useful behind a stray response--try forming it as a question which could connect to other issues or students. Otherwise, just throw out the strays.

- Avoid over-interpreting written comments. Look for the more specific comments as they tend to be more useful when you are choosing where to make improvements and where to continue doing what has been working well.

- It can be especially challenging to know what to do with contradictory comments, such as "the readings were terrific" and "the readings were awful." First, contradictory comments make sense; after all, your students are all different and likely have different learning styles. Look for repeated comments and patterns:

  o Compare comments in one category at a time. If you have similar sentiments expressed by more than just one or two students, look to see if there are any corresponding ratings about that topic. For example, if you have several comments that the course seemed disorganized but responses to one Likert question shows high ratings for organization, you can put those comments in better perspective: most students did not experience the disorganization articulated in a few comments. This understanding can drive your next questions: *Is it likely that a small group of students could benefit from a different approach? Is it possible or advisable to change my practices?* In this example, you might also ask, *have I clearly and frequently articulated the learning outcomes for the course and explained how assignments are designed to help students work toward those outcomes?* This practice can reveal the organization of the course that inexperienced students may have more difficulty recognizing on their own.

---

[3] Theall, M. and Franklin, J. (1991). "Using Student Ratings for Teaching Improvement," In M. Theall and J. Franklin (eds.), *Effective Practices for Improving Teaching*. New Directions for Teaching and Learning, no. 48. San Francisco: Jossey-Bass, 1991. See page 87.

[4] The IDEA Center (2012). Student Ratings of Teaching: A Summary of Research and Literature. Retrieved from http://ideaedu.org/research-and-papers/idea-papers. Manhattan, KS: The IDEA Center. See page 5.

[5] Winer, L. Di Genova, L. Vungoc, P. -A., & Talsma, S. (2012). Interpreting end-of-course evaluation results. Montreal: Teaching and Learning Services, McGill University

- In looking for patterns, it can be useful to informally "code" responses, compiling answers using whatever categories emerge. S e e t h e s a m p l e b e l o w .

CODING COMMENTS SAMPLE

What else, if anything, are you doing to help yourself learn in this course?

# of Students	Comment Categories
36	Reading all material, keeping up / staying ahead
25	Meeting with instructor or TA
21	Ask a classmate for help
19	Start my essays early
18	Take extensive notes
11	Cram before a test
10	Flash cards
10	Ask questions
8	Take notes on discussions in class
	Potentially interesting other single responses: I know it would be frowned upon, but during lecture I discuss material quietly w/ the student next to me. Discovery Channel

Interpreting this data:
Look for patterns in responses -- make grouping and counts. In this case, the responses were informally coded, compiling answers based on the categories that emerged. Perhaps you think more students would benefit from using flash cards or taking notes on discussions, so you can focus ways to communicate the importance of these activities to students.

---

3. Reflect and Create an Action Plan

- Spend some time reflecting on the feedback and your responses/questions. You may want to categorize feedback into *strengths, challenges, questions, and insights*.

- Be sure to identify what's working well, what you want to replicate or build on.

- Discuss the results with a trusted colleague; bring general questions to your faculty meeting, etc.

- What change(s) might you try? It may not be exactly what students suggest, but their comments might allow you to rethink something.

- Identify a few possible changes. Prioritize them, and decide on one or two to try.

- Share survey results with students this semester (if midterm) or next semester (if end-of-term feedback).

  - Present some (not necessarily all) the results in class.

  - Optional: Note that since students are different, you can't make the course perfect for each of them, you are balancing lots of needs and styles. For example, it's useful for students who don't like discussions to see that many other students do.

  - Identify one or more changes that you are making as a result of student input; let students know you'll

want their feedback about this change at the end of the term. Bonus: This approach generally increases response rates, since students see the instructor cares about their input.

- Consider this process like any other research: keep collecting feedback, asking questions, making adjustments, and getting more feedback.   (Suggestion: Use midterm feedback to get student input and make adjustments during a semester – this helps you experiment with new teaching techniques and refine them.)

- **Instructor's Course Evaluation**: Consider filling out an instructor's course evaluation at the end of each course, where you can summarize the results of the student ratings along with your own thoughts about how the course went--what worked and what didn't from your perspective--along with any notes about the context of the course that may have had an impact, such as new techniques, technology, the size of the class, characteristics of the students, etc.. This sort of reflection and record-keeping can help you track the development of your teaching over time and provide good material for a teaching portfolio. Questions on a form could include[6]:

  1. What changes did you make in the course this year? Were any changes based on previous student feedback? What were the results?
  2. Were the course materials effective? (readings, computer programs, clinical observations, group panels)
  3. Did any teaching approaches work particularly well or not work well?
  4. What did you notice about the students as a group? (characteristics, attitude, size of the class)
  5. Did the students have the prerequisite knowledge and skills that you expected for this course?
  6. Did you conduct a midterm evaluation and make any changes based on the results?
  7. Summary of student rating results and any comments you have about those results
  8. Strengths/weaknesses of course
  9. Recommended changes to course
  10. Any items you would like the curriculum committee to review

- Last but not least: Do <u>not</u> lose sleep or otherwise torment yourself over negative comments or low scores from 1 or 2 students.

4. References

Berk, R. (2006). *Thirteen strategies to measure college teaching a consumer's guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians*. Sterling, Va.: Stylus Pub.

The IDEA Center (2012). Student Ratings of Teaching: A Summary of Research and Literature. Retrieved from http://ideaedu.org/research-and-papers/idea-papers. Manhattan, KS: The IDEA Center.

Mitry, D.J. and Smith, D.E. (2014) Student evaluations of faculty members: A call for analytical prudence. *Journal on Excellence in College Teaching*, 25(2), 56-67

Nuhfer, Edward B. *Of What Value are Student Evaluations?* Center for Teaching and Learning, Idaho State University, 2003.

Karron G. Lewis, *Making Sense of Student Written Comments*. New Directions for Teaching and Learning Volume 2001, Issue 87, Article first published online: 4 FEB 2002

Theall, M. and Franklin, J. (1991). Using Student Ratings for Teaching Improvement. In *Effective Practices for Improving Teaching*. New Directions for Teaching and Learning, no. 48. San Francisco: Jossey-Bass.

Winer, L. Di Genova, L. Vungoc, P. -A., & Talsma, S. (2012). Interpreting end-of-course evaluation results. Montreal: Teaching and Learning Services, McGill University

---

[6] Adapted from Ithaca College's Instructor's Course Evaluation Form (2002)