Perspectives: Teaching Legal Research and Writing | Vol. 21 | No. 2 | Spring 2013

Cite as: David I. C. Thomson, *Using Student Evaluation Data to Examine and Improve your Program*, 21 Perspectives: Teaching Legal Res. & Writing 115 (2013).

Using Student Evaluation Data to Examine and Improve your Program

"'[S]tudent evaluations of teaching are likely to be much more useful than their critics typically believe.'"

By David I. C. Thomson

David I.C. Thomson is LP Professor and Director of the Lawyering Process Program at the University of Denver, Sturm College of Law in Denver, Colo.

At many schools, directing a legal writing program today is quite different than it was even 10 years ago. As LRW faculties mature and the individual faculty members grow in the profession, the need for a "top-down" director is lessening or going away in many programs. However, in many schools there remains a valuable leader/coach sort of role for a director, whether that person rotates, coordinates, or however it works in practice that is best for the school. This new sort of director is ideally someone who is able to encourage and support a culture of programmatic excellence and is willing to ask questions about how the program is doing as a whole—understanding, of course, that a culture of excellence and examination is created and given life by the faculty members in the program, not by the director.

One way to encourage a culture of excellence is to start with some measure of how you are doing and discuss whether there is room for improvement. Because we did not have that measurement, I looked around for what we did have, and of course—just like every program—we have student evaluations. They are certainly not a perfect measure, but they do have some validity.[1] "In general, student evaluations of teaching are likely to be much more useful than their critics typically believe."[2] With respect to student evaluations of legal writing professors, it is a commonly held belief that we inevitably receive lower scores on evaluations than those received by teachers of casebook courses. But this view has been challenged[3] and, with effort, can in large part be addressed.[4]

Of course, whenever you start poking around in student evaluations, folks get nervous. The worry is that such an effort is secretly about going after someone, even if it is not. It is very important that it is clear to everyone up front that the purpose of gathering and examining the evaluation data is solely for program-wide assessment. Assuring everyone that the data would not be used for individual faculty assessment was an important step. Student evaluations are certainly used in the various review processes for all faculty members, but that is not what this sort of programmatic study is about.

Because I am not an expert empiricist, I sought out help from those who are. Every university has an assessment person somewhere. I found ours, and studied as much as I could. I learned quickly a fairly basic point: that such a study would be more reliable if it had a lot of data in it, rather than a little. So the first study we conducted includes five years of student evaluation data across 16 sections of the course, both semesters each year, between the Fall 2005 semester and the Spring 2010 semester. That is 160 sets of evaluations covering nearly 3,000 individual evaluation forms. Finding someone to collect and chart that amount of data is not generally easy, but at the time we had an administrative support person who was good at this, and enjoyed doing it. When I asked her to do this work, I gave her some specific parameters to look at.

1 *See,* Arthur Best, *Student Evaluations of Law Teaching Work Well: Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree*, 38 Southwestern U. L. Rev. 1 (2008).

2 *Id.* at 34.

3 Julia Glencer, Erin Karsman, Jan Levine & Tara Willke, *The Fruits of Hope: Student Evaluations*, 48 Duq. L. Rev. 233 (2010).

4 *Id.* at 260-261.

Our student evaluation instrument is pretty detailed, arguably too detailed. It contains 18 statements, and students are asked to rate our work with them as earning a rating of Strongly Agree, Agree, Neutral, Disagree, or Strongly Disagree for each statement. Students are then invited to write textual comments about the professor and textual comments about the course materials. The student evaluation participation rate runs about 60 percent school-wide, and is generally over 70 percent in our course. Of the 18 statements, it seemed to me that eight of them were most relevant to our course and important to study. These questions were:

#7: I found this course to be well organized.

#8: The professor held my attention in class.

#9: This professor is always prepared prior to class.

#10: This professor was willing to assist me outside of class.

#12: This professor made good use of class time.

#14: This professor effectively communicated the content of the course to me.

#16: This professor motivated me to do my best work.

#18: I would enjoy taking another course from this professor.

I selected these statements to focus on, in large part, because they seemed most aligned with the guidance provided in a leading article on the subject of student evaluations in legal writing.[5] In his article, Professor Walter suggests that the best use of student evaluations in legal writing is to focus on those questions that relate to "professionalism and respect for students."[6] While all of our questions do not align perfectly with Professor Walter's guidance, statements #8, #9, #10, and #12 do quite directly, and #16 arguably does as well. Examples of two statements on our evaluation that I did not include in the study were #15: I would recommend to others that they take this class, and #17: I was able to keep up with the workload for this course. I left these out of the study because, with respect to statement #15, our students have to take the course, and with respect to statement #17, our students routinely complain about how hard the course is, because—generally speaking—it has to be. It seemed to me that including data from both statements would skew the results by including information that does not relate to the professionalism of the professor and his or her respect for students.

To keep the study at a manageable size, I decided also to reduce the number of responses to each statement that we would study from five to three. I excluded the Strongly Agree and Agree responses from the study, instead only including Neutral, Disagree, and Strongly Disagree answers. I did this because it seemed to me that the most useful information would be that which revealed those who were unhappy with our work, rather than those who were happy.

This was a fairly fundamental decision in this study. I decided that what we most wanted to know was where the problems were. I was thinking, I suppose, of the first line of Tolstoy's *Anna Karenina*: "Happy families are all alike; every unhappy family is unhappy in its own way."[7] What we most wanted to know was how and in what ways our family—our students—were unhappy with the program. That would tell us more than just studying the students who were fully satisfied with what we were doing.

Further, I decided that when a student provides an answer of "neutral" to one of the statements, it is not a bad response, exactly, but it is not a happy one either. So I wanted to include the neutrals in the study as well.

The result of the study can be found in figures 1, 2, and 3. These charts show the mean number of Neutral (fig. 1) responses, Disagree (fig. 2) responses, and Strongly Disagree (fig. 3) responses over the period of the study. Given small class sizes, one would hope that the number of students using these response options would range from zero to five. As you can see, over this lengthy period of

"What we most wanted to know was how and in what ways our … students—were unhappy with the program. That would tell us more than just studying the students who were fully satisfied with what we were doing."

5 *See*, David D. Walter, *Student Evaluations – A Tool for Advancing Law Teacher Professionalism and Respect for Students*, 6 J. Leg. Writing Inst. 177 (2000).

6 *Id.* at 191-92.

7 Leo Tolstoy, *Anna Karenina*.

“I have long held a theory that teaching a small class such as legal writing in a large law school classroom might be just a little bit more difficult than teaching in a smaller classroom … But I had no data to back that up.”

the evaluations we have a reduction in the mean number of students who responded on all three measurements. That is, the number of students who are unhappy has reduced over that time. Merely a reduction alone in the Neutral numbers, or in one of the other categories alone, would be interesting and good information to have. But it is the similar reduction across all three categories that make these charts compelling to us. We took this to mean that over the 2005–2010 period, students experienced an improvement in our program. Or at least they were less unhappy with it.[8]

Once we had this data, we could use it in other ways beyond the rough-cut "improvement" conclusion. We noticed that there was an obvious division between two statements and the rest. One of the two statements on which we were scoring the lowest dissatisfaction rate is fundamental to legal writing pedagogy—Statement #10: is "willing to assist me outside of class." If we were doing very poorly on that measure, it would be something to be concerned about as a program. It was encouraging that we were not.

We also noticed that the unhappy mean numbers generally decreased between fall semester and spring semester each year. This is as we would expect, since students often do not fully understand the sometimes painful and difficult work we are assigning them in the fall semester, while by the spring semester they often understand much better what they have learned. It was somewhat affirming that this phenomenon appears in our evaluations as well, but it also led us into good discussions about why such a phenomenon exists, and how it might be addressed in the future.

This data also allowed us to notice that, for example, we were not doing as well with student responses to statements #12 (good use of class time) and #14 (effectively communicated content), particularly among those who selected "Neutral" as a response. This led to a healthy discussion of the issue, and it is something we often talk about in our biweekly faculty meetings during the semester. Further, once we had this data, we were able to use it to cross-reference with other data points.

I have long held a theory that teaching a small class such as legal writing in a large law school classroom might be just a little bit more difficult than teaching in a smaller classroom that is more conducive to the kind of community that we try to foster in our classes. But I had no data to back that up.

I asked our data specialist to look at the evaluation data again, but to cross-reference it by classroom. I wanted to know whether there was a relationship between the "unhappiness" factor and the classroom in which the class was taught. I separated the rooms in which we have taught our LRW course in those years into two groups: larger rooms vs. smaller rooms. I also reduced the number of evaluation statements from eight to the three that seemed to me would be most directly affected by the dynamic created by teaching in a smaller or a larger room. The three statements that were examined in this study were: #8 (held my attention in class), #12 (made good use of class time), and #14 (effectively communicated content).

What we learned is that indeed there does seem to be some correlation between the classroom assigned and student unhappiness with the course. As you can see in fig. 4 (Neutral), fig. 5 (Disagree), and fig. 6 (Strongly Disagree), those smaller rooms (the blue bar) all have a lower "unhappiness" mean on each rating. I have used this data to request of our Associate Dean and Registrar that our Lawyering Process classes be taught in the smaller rooms whenever possible.

All of this work is based on student evaluations, which are certainly not a perfect instrument. Further, we are all quite sensitive about our student evaluations on an individual basis, and so we usually focus on them for individual improvement. Grouping a large data set of the evaluations and just examining the quantitative data that they contain helped us as a group of faculty to examine the program as a whole. It also led us to be more willing to examine program-wide data, rather than focusing on our individual scores. This has since

[8] You might notice that we experienced a brief spike in the Strongly Disagree category in the spring semester of 2009. The source of that problem had already been addressed by the time we reviewed this data, so conducting this study did not lead to its resolution. But it did confirm the nature of the problem.

led us into a more rigorous assessment process that is based on measurable student learning outcomes, a standardized rubric, and program-wide assessment based on improvement in actual student work, rather than merely student perceptions in their evaluations of our work with them and of the program. This study will be the subject of the next article on our experience at the University of Denver with program assessment.

Figure 1

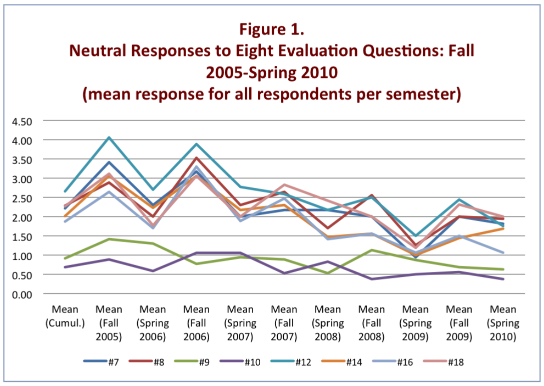

Figure 2

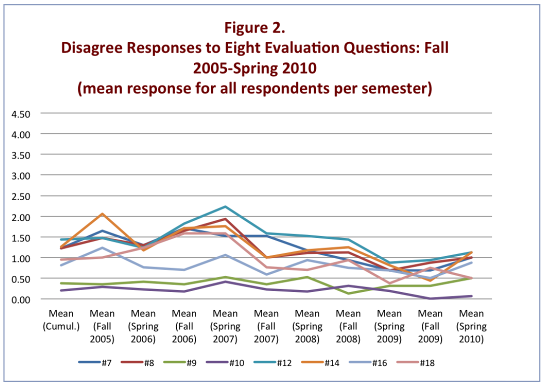

Figure 3

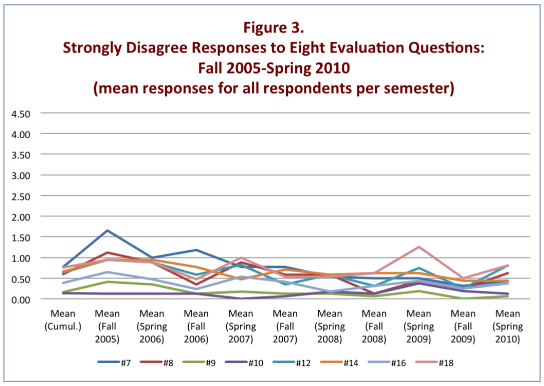

Figure 4

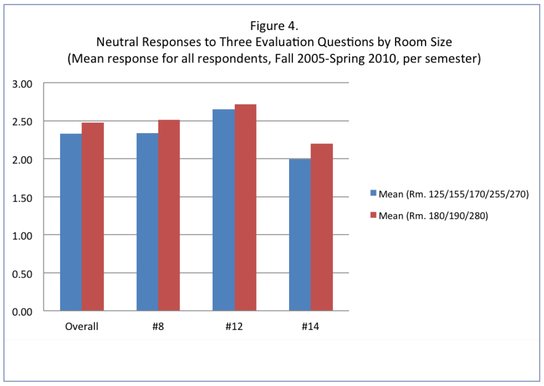

Figure 5

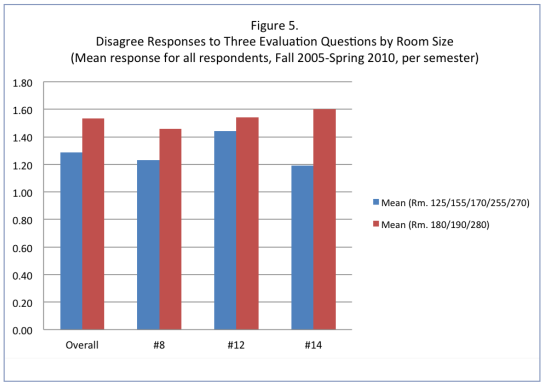

Figure 6