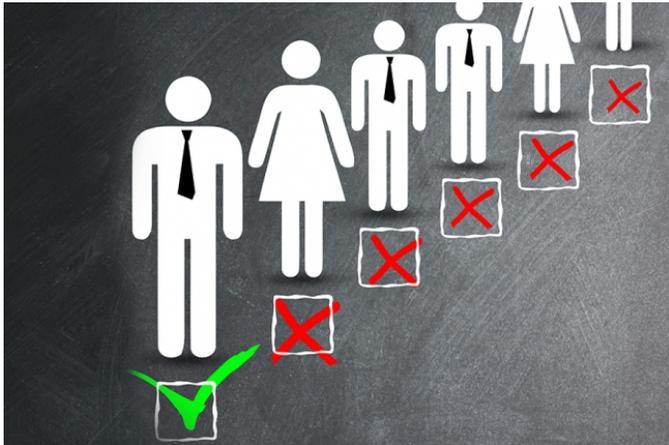# THE CHRONICLE
of Higher Education

In Defense (Sort of) of Student Evaluations of Teaching

By Kevin Gannon | MAY 06, 2018

iStock

A couple of weeks after the end of my first semester of teaching as the instructor of record, I received "the packet" in my campus mailbox — an interoffice envelope stuffed with course evaluations from my students. Those evaluations mattered a lot to me at the time, as I was still figuring out this whole teaching thing. Was I doing a good job? Did my students like the class? And, more selfishly, did they like me?

Well, in this particular batch, one student certainly did not like either the course or me. In the comments section, the student flatly declared: "He was a real ashole."

The spelling in that quote is sic. In that moment — as I wrestled with both the shame of being deemed an "ashole" and the urge to laugh at the absurdity of that being the sum total of this student's assessment — I had my first experience with a question that faculty members regularly confront:

Do student course evaluations even matter?

Of course, the short answer is that they do, at least to the people who make the decisions about our futures in academe. Department chairs, deans, promotion-and-tenure committees — all of them and more use student evaluations to determine whether or not we are "good teachers," and, more consequentially, whether we should continue to teach on the campus.

I've never been completely comfortable with the weight those forms are accorded — neither when I was a department chair evaluating part-time faculty colleagues nor in my current role helping colleagues interpret the feedback they receive in their own course evaluations. At best, student evaluations of teaching are a flawed instrument; at worst, they're a cudgel used against faculty members, many of whom already occupy precarious positions.

In some departments, student evaluations are one part of a larger set of evidence used to assess teaching performance. But elsewhere they make up the bulk — if not the totality — of that evidence. We all know that the race and gender of faculty members can affect how their teaching is evaluated. And we've all heard horror stories of instructors who had consistently good ratings, except for that one outlier comment seized on by the promotion committee and used against the instructor. Some of us have lived those horror stories ourselves. As the psychologist Abraham Maslow famously observed, if the only tool you have is a hammer, you tend to see every problem as a nail. This is a particularly apt description of the problems that can inhere in the faculty-evaluation process.

So we know student evaluations matter. Perhaps the better question is: Should they? Given their many demonstrable and potential flaws, why would we still use them to gather feedback on teaching and learning? It turns out the answer is more complicated than appearances suggest.

Certainly, students are not experts qualified to evaluate us on, say, whether we used the best and most applicable course readings. But they are experts on what they experienced and learned in a course, and they ought to have a voice. Just because their feedback is sometimes misused doesn't mean it's invalid or unnecessary.

In fact, course evaluations — despite their many problematic elements — may still provide the most accurate information available on teaching effectiveness. Elizabeth Barre, whose research into student evaluations — in particular, the metastudies of the subject — is essential reading, observed that "we have not yet been able to find an

alternative measure of teaching effectiveness that correlates as strongly with student learning. In other words, they may be imperfect measures, but they are also our best measures."

And therein lies the rub: We need to assess teaching, and we often have to rely on not the best, but the least worst, option.

What does that mean, though, for the individual instructor opening a packets of evaluations at the end of a semester? What can we do with the results of these flawed instruments, as they aren't going anywhere anytime soon (and certainly not before our next portfolio review)? What follows are some suggestions — for individuals and institutions — on how we can use these tools constructively and appropriately, as opposed to employing them like Maslow's hammer.

One's an accident, two's a trend, three's a problem. Whenever I receive a batch of course evaluations, I immediately scan the comments left by the students. Only later do I look at the quantitative data summarizing the entire class. I know I'm not alone in that habit, and it's a natural reaction to want to see what our students actually said about the course — and about us.

But it's also hazardous. If you're anything like me, once you see a negative comment, it's over. It's like when a toddler pees in the pool; it only takes a small amount to ruin the whole thing. I could get a whole section's worth of rave reviews, but if there was one nasty comment, that's what I'm going to obsess over. It's as if the rest of the ratings didn't exist.

We need to avoid that trap. We wouldn't want our department chair or the tenure committee to seize on one isolated data point to characterize our entire teaching performance, so we shouldn't do that to ourselves, either. Look for trends, not outliers.

And don't ignore the quantitative results, either, as they can tell us a lot. For example, if in response to a question on the overall quality of the instructor, you see that 90 percent of your class responded with "good" or "excellent," then that one comment about how "awful" you were gets put into its proper place. Anecdotal data are not representative. It's

the trends (perhaps most of the students said they would have appreciated more guidance on a particular assignment, for example) that tell us where to focus our energy. The whole point of feedback is to help us become better teachers, and we can't get there by obsessing about outliers instead of reflecting on the representative trends and aggregate results.

Don't take it personally. That's easy to say when you're not the one students are calling "ashole." As hard as it is to do in practice, though, we cannot let anonymous rants become a referendum on our personal worth.

Teaching is a difficult gig for many reasons, but chief among them is the degree to which we tend to tie our sense of identity and self-worth into our classroom performance. Because of that, we may lack the critical distance necessary to reflect honestly about our own pedagogy. Student comments can be mean — sometimes unintentionally and other times most definitely on purpose. Likewise, quantitative ratings can be lower on the Likert scale than we were prepared to see. And it hurts. Believe me, I know.

It's hard to take the long view in that situation, but it's essential to make the attempt. No one is a perfect teacher, and what works great one semester can bomb the next. The key to improving is to be, as Stephen Brookfield put it, a "critically reflective" practitioner. If all we take from a batch of evaluations is "I'm awful," or "I'll never be a good teacher," then we can't properly do the reflective work to diagnose what went awry and course-correct for the next time.

What story does this data *really* tell? Sometimes a course just goes badly. We don't teach at our best. Maybe there's a clutch of hostile students in class. Or maybe we got inserted as instructor at the last minute and couldn't prepare as well as usual. There are plenty of factors that can help explain negative evaluations, if we're honest.

We need to remember that data acquires meaning only through context. Every set of data has a story. When we interpret the results of our course evaluations, we should be thinking about the ways we'll tell our story — because if we don't, someone else will. That means taking advantage of whatever vehicles exist to help us both present our ratings and place them in the proper context. For example, as part of the evaluation process,

some institutions ask faculty members to write a narrative, which would be an ideal spot to discuss the factors shaping the course-evaluation results. Another opportunity may arise via direct conversations with whomever is conducting our evaluation.

At some point we all must advocate for our own teaching. Was a course particularly difficult to manage? Did students chafe at the amount of reading or writing involved? Were there issues that affected the classroom dynamic? Was it a class that students traditionally dislike, or see as intimidating? Those and a host of other factors can affect student perceptions of the course. It's critical that we understand and are able to articulate that larger context to others.

It may feel like we're merely rationalizing, but abandon that mind-set. Honest advocacy of our teaching is not an exercise in excuse-making; it's making sure that the process works like it's supposed to. The only way to accurately assess teaching performance — no matter if it's us or someone else doing the assessing — is to put our data in its proper context. I've focused mostly on negative evaluations here because they're the ones that can derail our confidence and careers, but I don't want to discount the importance of positive student feedback and the need to tell that story, too.

How do we evaluate teaching, anyway? Given the well-known limitations of student evaluations, it behooves every department or institution to be careful how they are used. The best faculty-evaluation systems are multilayered and employ a number of different measures.

To be honest, student evaluations of faculty instruction ought more properly to be referred to as "ratings," since "evaluation" connotes a more complex, informed process than what's possible via these instruments. In assessment terms, student evaluations are only indirect measures of teaching effectiveness, and any assessment process dependent on indirect measures will not produce accurate information.

Instead, student evaluations ought to be treated as supplemental material. They should complement — but never overshadow — faculty narratives, peer observations, reflective dialogue, and sample teaching materials. Even more important, course ratings should be used equitably; their documented bias against specific faculty groups has to be part of

the calculus. To assume that all student-evaluation data can be unproblematically used in the same way for every faculty member ignores substantial evidence to the contrary, and undermines the evaluation process.

Departments and institutions have an ethical obligation to be discerning in evaluating faculty members. Flawed processes create flawed results. It's incumbent on us to evaluate teaching with a process that centers faculty voices and experience, and assures that data will be interpreted with attention to context.

These suggestions won't make student ratings any less flawed but should help us reckon with the nature of those flaws and then proceed accordingly. As is often the case in teaching, the key is knowing how to use our tools appropriately. Well, that and not being an "ashole."

Kevin Gannon is a professor of history at Grand View University and director of its Center for Excellence in Teaching and Learning. Browse his previous columns here .