

Dead Ideas in Teaching and Learning Podcast Series

Season 7, Episode 6: Let's Stop Relying on Biased Teaching Evaluations with Joanna Wolfe

Center for Teaching and Learning, Columbia University

[00:00:00] **Catherine Ross:** Hello and welcome to Dead Ideas in Teaching and Learning, a higher education podcast from the Center for Teaching and Learning at Columbia. I'm Catherine Ross, the center's executive director. As a quick reminder for our listeners, in this podcast series, we are exploring dead ideas in teaching and learning. In other words, ideas that are widely believed, though not true, and that drive many systems and behaviors in connection to teaching, exercising what Diane Pike called the tyranny of dead ideas. Welcome, everyone. I'm speaking today with Dr. Joanna Wolfe, whose article, *Let's Stop Relying on Biased Teaching Evaluations*, offers some really helpful institutional strategies for addressing a well-documented problem with student course surveys.

[00:00:57] Dr. Joanna Wolfe is a teaching professor of English and affiliated faculty of mechanical engineering at Carnegie Mellon University. Her research focuses on engineering communication and diversity, equity, and inclusion. Her, her most current research, published in the journal *Sex Roles*, examines communication strategies that women can use to speak up effectively about problems they experience in male dominated settings.

[00:01:26] Her research tends to be very applied and action oriented, seeking to make specific and measurable changes in the world. She is working on a forthcoming book series titled *Modular Technical and Professional Communication* to be published by Macmillan.

[00:01:43] Welcome to our Dead Ideas podcast, Joanna. I'm so delighted to be talking with you today.

[00:01:49] **Joanna Wolfe:** Well, thank you so much for having me.

[00:01:51] **Catherine Ross:** So I'm going to set the stage a little bit here, um, to keep us sort of in a thread amongst our different episodes this season, which is targeting systemic change. Um, in my own experience working in teaching innovation and educational development, it's become very, very clear to me that one of the major systemic barriers to changing teaching is the ways in which, um, it's evaluated and particularly the ways in which student surveys are used by institutions for, um, some very high stakes decision making.

[00:02:29] And for our listeners who have listened to our previous podcast with Michelle Miller, teaching evaluation improvement is one of the markers of whether an institution truly values teaching or doesn't. So I'm very interested in highlighting Joanna's ideas about how institutions can make these changes.

[00:02:52] Um, there's in general, just so many dead ideas about teaching evaluation practices from, you know, questions about the forms and what should be on them and how they're administered, what gets reported back to instructors um, as well as a lot of dead ideas that instructors have about these, um, these student surveys in particular. Um, but I thought instead of unpacking all of the many dead ideas about evaluation of teaching it would be really helpful to offer some solutions. Um, rather than just naming the problems. So Joanna's article caught my eye because it offers really, um, implementable solutions and strategies that could mitigate the harm that these evaluations can cause when they are misused.

[00:03:42] Um, and I would say that although the article was spurred, um, at the time in 2022 by efforts that institutions were making, uh, to confront and address racism. The three suggestions in this article are still critically important and much needed today. So I really wanted to shine some light on, on this piece.

[00:04:06] So Joanna. Can you share with our listeners a little of the background that your article provided on why we should be making changes to these instruments, especially, you know, when we consider the unintended outcomes of these systems for faculty success and retention, especially.

[00:04:27] **Joanna Wolfe:** Yes. Well, thank you so much for the question. There's been lots of research on student evaluations of teaching, and multiple studies have uncovered gender biases and racial biases in these evaluations. Male teachers overall are perceived as more competent, as having better class leadership skills than female teachers. In terms of racial differences, faculty of color receive lower evaluations than white faculty overall, and again, larger cultural stereotypes feed into this. And at the same time that there's this well documented bias in the evaluations, they show little to no correlation with

learning. And that's learning measured in terms of test scores or in terms of how students do in a subsequent class in the curriculum. And one of my favorite studies, showing the bias, took an online class where students never saw their instructors.

[00:05:34] And this class was taught by two different graduate teaching instructors. One male in real life and one female and what the researchers did was they manipulated what the students saw. So that half of the students who had in reality the female teaching assistant believed that that teaching assistant was male and half of the students who had the male instructor in reality were led to believe that that instructor was female and so they were able to calculate the evaluations of teaching Both for the real female and real male and the perceived female and perceived male. In reality, the actual female on every measure did slightly better, but not significantly better than the male instructor, but in terms of the perceptions and the perceived gender. When students perceived the gender of the instructor as female, they had significantly lower evaluations on all sorts of measures, including objective measures, such as whether or not the assignments were returned on time or things like whether or not they liked the textbook. And so this really kind of shows the difference that the student perceptions can make these evaluations.

[00:06:57] So we have this well documented bias. And yet this is the primary way that we evaluate faculty teaching. I am a teaching professor at Carnegie Mellon, and the only way that my teaching is evaluated is by asking students and asking students on these evaluations. And we have no idea how many potential instructors of color are turned off by the perceived bias, or the actual bias, in these evaluations.

[00:07:35] Uh, we have no idea how much they are frustrated by, by this, or just declined to pursue a profession where there's so much bias baked into the evaluation system. And we know that women and faculty of color actually put more effort into their teaching, but for, for less reward. So, there's um, many reasons not to use these.

[00:08:05] In addition, I would argue that they actively hurt teaching by making instructors worry about things that might be good for the students, but that they know will negatively affect them on their teaching evaluations. So I know as a female faculty, that if a student asks me if they can turn in something late, and I refuse then that student will punish me on the evaluations in a way that they would not do for a male instructor and, uh, so that gives me an inclination to grant that request, even if it's maybe in the students best interest to hold them

accountable to deadlines. And I think that plays out in a myriad number of ways.

[00:08:52] **Catherine Ross:** Yeah, there is just so much research on this that supports everything you're saying. I've seen this in the work I've done over the years in this realm. And it's very, very hard to underestimate the impact that it has on faculty retention, particularly, I think, in stem disciplines is it really can be even harder because, um, there's research going back decades that in general, science instructors can get lower overall evaluations than instructors in the humanities. So there's all these sort of layers of problems. Um, and I'm really glad that you were able to, um, address that, you know, in a very clear way so that we can see I, you know, I think we do need to keep the student voices. We want this student voices, but there are ways to do that that don't jeopardize people's employment or people's career trajectory. Right? So that's what we need to be really thoughtful about getting that kind of information and how, and then how we use that kind of information for, um, teaching improvement, not for getting, not reappointing people or not giving people a promotion.

[00:10:10] So yeah. Okay. So your good recommendations, let's jump into those. Your first recommendation is as follows. To remove or replace vaguely worded questions prone to eliciting bias. Could you share some examples of how vagueness in a question can become a bias amplifier? That's a term I use. Um, I don't know if it's common. And what, what would a more concrete question look like?

[00:10:44] **Joanna Wolfe:** Yes, thank you. So probably the most consequential question on the evaluation, one that most people pay attention to, is worded something along the lines, of course, worded differently at different institutions, but overall, how would you evaluate this instructor?

[00:11:02] And if you were going to design a question to elicit bias, you could hardly do a better job than that question. Because first of all, there's no criteria associated with it. So what, what are the evaluation metrics? What does it mean? This question invites comments on personality. So we know that these questions will elicit bias on personality that students might comment on how abrasive someone is or condescending or insecure, and these are all words that get used much more frequently with women than they do with men. Instead of these broad, open-ended questions where we don't even know what teaching means to the student, we need to ask much more narrowly worded questions.

[00:11:59] So, for instance, another question that is typical on evaluations is, were assignments returned promptly? Well, what does prompt mean? That can

mean different things to different people. And that's such an easy one to revise. We can specify the time frame. We can ask students where assignments returned within one week, within two weeks, within three weeks.

[00:12:21] Uh, we can ask for a specific metric. Or we could even forego asking students because most faculty nowadays are, are returning assignments online. And so we could use the content management service to find out how long that is. We can ask students about their learning. So how much do you feel that you've learned about x, whatever the learning goal is. And that's how I do my own personal evaluations in my classes, is I go through my syllabus and take all of my learning objectives on there. And for each learning objective, as students, To rate how much they learned about it and how important they think it is for their future, future selves, and that gives me a lot of really good information about what I'm really nailing in the class and where I'm maybe not doing as well as I, I think I am. These questions again, when they're asked in very vague ways, we are asking students also to evaluate things that maybe they don't have the background or sufficient knowledge to evaluate. So, for instance, when we ask them to evaluate teaching, well, one of the reasons that more faculty do not adopt active learning principles in their classes, even though we have tons of evidence suggesting that active learning is much more effective than passive learning, is because to students that can look like the teacher isn't doing anything. They aren't doing what the student considers teaching. So even though the student learns more in that environment, when they're asked to evaluate the faculty's teaching, they might say, well, the faculty didn't really do anything. They just let us do stuff and they just let us talk to other students and engage them. I had to figure it all out myself. The faculty didn't do anything here. If we ask students what they learn, what the student is qualified to say, and can give at least their perceptions of what they learn, I think we will get much better results from these, from these questions.

[00:14:41] **Catherine Ross:** Yes, which is related to your next point, but before we jump to that, I do want to say that in my years and years of working with instructors, when they receive these evaluations and helping them parse what it means and figure out, you know, what they want to do about it. Um, I have actually seen those words written many, many times. We, if an instructor did use active learning, students would say, oh, um, we had to teach ourselves.

[00:15:13] And I was stunned the first couple times I saw that, and then I was like, right, because instructors may not realize that they need to talk about what they're doing and why they're doing it that way, right? To prepare students for a kind of learning that requires more from them than they're used to.

[00:15:35] Many students believe, this is a dead idea we tackled in last seasons, in the last episode with two students on the science of learning, um, You know, many students believe that sitting and listening to a lecture is really good and that's how they learn because there is that illusion of comprehension, right? Um, and they don't always like being asked to do something different, as one of our students, um, said in that podcast. So, um. Yeah, it's a real challenge, and then if you add into that mix the instructor identity, things get even worse for, as you've said, for women instructors and instructors of color. Um, it's, it can be tough going.

[00:16:23] So, um, I do want to come back to that overall question, though. Could we change it to something like, overall, how would you evaluate, um, this instructor's teaching in terms of how it helped you learn or something like that? Does that help?

[00:16:42] **Joanna Wolfe:** I still would not ask them to evaluate the instructor's teaching. I would ask them to evaluate their own learning.

[00:16:49] **Catherine Ross:** Yeah. Yeah, that makes sense.

[00:16:50] **Joanna Wolfe:** That removes some of the bias then by not asking about the instructor, by asking how much they, they learned, uh, that's going to be less influenced by the identity of the instructor.

[00:17:04] **Catherine Ross:** Yes. That's, that's one possible solution, um, if we could convince administration to give up that question, that number, that overall evaluation question. That's a different, we could talk all day. Okay. So let's look at the second recommendation, um, to educate students to be less biased evaluators, which is kind of related to what you were just saying about involving students in evaluating their own learning. So please, um, give some, give us our listeners some ideas about how they might, um, talk to students about these surveys.

[00:17:43] **Joanna Wolfe:** Yes. So we can remind students that they should be professional. We should encourage them to avoid implicit bias and let them know what the biases are. We can, more specifically, tell them to avoid personality comments because that's really not the goal and those comments tend to be so inflected with bias.

[00:18:08] **Catherine Ross:** Also, how people dress. I have seen those kinds of comments on women's, um, student surveys and sometimes really brutal and nasty. Um, so yes, personality and appearance.

[00:18:26] **Joanna Wolfe:** Absolutely. Uh, appearance is a, is a big one. And we could hold students accountable for that. Now, the instructor can't know who the student is. We want to preserve their anonymity in that case. But there's no reason why an advisor couldn't, for instance, review student comments. And if a student says something sexist or racist, an advisor could, at the very least have a conversation with students. And if students know that they're at least going to have to have a conversation about this, I think they'll think a little bit more. Or if a student does something inappropriate, such as comment on appearance, an advisor could have a talk with them and say, why, what made you think that this was that this was appropriate here.

[00:19:19] **Catherine Ross:** So what could instructors say to students? Um, you know, what could you tell your students that might help prevent this from happening?

[00:19:31] **Joanna Wolfe:** Well, what I do in my classes, which does tend to change the tenor of those open-ended comments, is I ask them to really kind of focus on the class itself. And I ask, I say, give me two or three things that I should keep that you thought were very effective about the way this class was structured and give me two or three things I should change. And that keeps the comments very constructive. I also, in my classes talk about, we do a lot of peer feedback in classes. And so I talk about giving people suggestions for improvement and I explain that I, uh, expect them to do that, to do that with me and that I value their feedback and that I take those constructive comments very, very seriously I think that really kind of focuses it.

[00:20:26] Now, one of the problems with this, though, is that currently at my institution, I get no control over the timing of those evaluations. So a lot of times before I can tell students how I would like what kind of feedback I would like them to give me, the evaluation has already been released and students have already filled it out. Of course, either the students who really love me or the students who have some kind of problem with my class, and so it's too late for me to give that kind of direction to their feedback.

[00:21:02] **Catherine Ross:** I think one potential solution that, um, I've recommended to instructors over decades, um, and it works every time that would sort of had that office to do a kind of early semester um, start, stop, continue, right? What could we start doing that would help you learn? What should we stop doing that doesn't help you learn? That kind of thing. Continue doing that does, you know, um. So I think if that happens, that's an opportunity early in the semester to have that conversation with students to talk about why you're going to do this quick survey, you know, to check in with them in the

beginning of the semester to make sure that they're learning is happening. And, and that this will carry over into the final surveys they're going to fill out, right? That you value feedback and I think that sends a message early on and also helps them, um, already learn how to think about your learning, giving information on their own learning, right, as opposed, you know, giving good feedback, concrete, specific examples of things that are helping them learn or not helping them learn. So, um, I, I often offer that to people because it's a very low stakes, easy way to kind of head that off, right? So you don't find yourself in, in that particular situation. But yeah, I think all of those things you said really are helpful and do change, um, student responses, um, in, I mean, I've worked with people and I've seen the changes in, from doing the very things that you advised.

[00:22:51] So the first step that universities could take would be to incentivize faculty members to study ways to mitigate bias in teaching evaluations. I, I really thought this was a great idea. And I did see it in action at one university that I worked at. It took five years, but it worked. Um, but yeah, using this in house expertise to develop better instruments.

[00:23:18] Um, would you see this work happening, like, do you think in a department, or is this a Senate thing, or, or like a task force kind of situation? The experience I had was all of, it started as a task force, and then it went to the, um... let's see, how did it go? It started as a task force, and then it went to a Senate committee, and then from there it finally went for approval through the Senate.

[00:23:48] **Joanna Wolfe:** I would think as a task force, but maybe one that is commissioned by the, by the senate, so that it goes through the faculty senate. But this just seems like such a no brainer, because so many schools, so many institutions have faculty in psychology or in the school of business who studied evaluations. It just doesn't make sense, especially when they are so instrumental and central to our lives as faculty.

[00:24:20] **Catherine Ross:** Yeah, absolutely. Totally. And I think in the case that I was involved in, um, we finally had a pilot ready to go. And we had faculty volunteer to do both the old form and the new form. And we also were able to ask students questions about the forms and, you know, any reactions they had to the new form, which was hugely valuable. And then over the summer, two faculty, I think from the school of education, went through all the results of multiple large classes. They, these were large intro classes and did all of the um, psychometric evaluation of the forms. So we knew in the end with their suggestions incorporated that we had a valid and reliable instrument.

[00:25:12] **Joanna Wolfe:** That makes so much, so much sense. We have so much expertise in house. There's also, I want to point out, there's no reason that there, there should be just one form or that forms can't be customizable. We obviously have the technology. Uh, and there's no no reason why every faculty should be evaluated on the exact same seven questions.

[00:25:38] **Catherine Ross:** Exactly. In, you know, at the time this happened, that was not easily done, but you're correct. Today, it can, these forms could be very customizable. And, and many times are. I think the thing that really impressed me was that none of your evaluation Um, your recommendations would be hard or costly to implement, right?

[00:26:00] That's usually a barrier if it's going to cost a lot of money. But why, so why do you think institutions aren't taking these steps? I know there is a group, there's a group called TVAL, um, and it's a consortium of institutions that have tackled this and have, have been doing really interesting work. Um, and we'll have a link to that. There's a link to that in the description of this episode. But why do you think so many universities aren't doing it?

[00:26:28] **Joanna Wolfe:** I have to say my answer is somewhat cynical, but I think it's something that you can't really brag about the way that you could brag about a new vice provost for DEI or new center, or even a new committee, uh, and this is really a small thing that should have happened years ago. So almost by tackling it and making it too public, the institution is kind of opening itself up for criticism of why, you know, why have you been functioning that way for, for all these years?

[00:27:08] **Catherine Ross:** Indeed. Hadn't thought of that.

[00:27:12] **Joanna Wolfe:** It's very similar to politicians don't get elected for fixing problems. They get elected for big ideas. It's a kind of unsexy problem to fix because it is. It's so easy and straightforward.

[00:27:26] **Catherine Ross:** Yeah. Yeah, definitely. And I think it also ties to how institutions are sort of evaluated, you know, whether it's in rankings or other things and teaching is. It's, let's just say it's often lip service. Um, there's no really robust, um, information about teaching and, and things like that. And so, as you said, it's not something big that they're going to get public acknowledgement for, although I know there are people who are thinking that needs to change. So we'll see if it, if it happens, but, um...yeah, I think you're spot on.

[00:28:11] **Joanna Wolfe:** Thank you. Also, I'll just say that sometimes people get a mentality because to really do a rich and full job of evaluating teaching would be a lot of effort and a lot of work. And so I think people get intimidated by that. And the case I want to make is let's just do the small thing but do it better. Let's focus on some incremental improvements, even if it's still not going to be our ideal.

[00:28:39] **Catherine Ross:** Yes, don't get me started on peer review. There's a lot of work that could be done there, but if we could at least, at least address the issues around the use of the student surveys, because those are pervasive, um, that would be a huge step forward, as you've suggested.

[00:28:59] Thank you so much, Joanna. I'm so grateful for you agreeing to participate in this seventh season of Dead Ideas and to tackle this big one.

[00:29:11] **Joanna Wolfe:** Thank you again for having me. I really appreciate this opportunity to talk more about this really important issue.

[00:29:22] **Catherine Ross:** If you've enjoyed this podcast, please visit our website where you can find any resources mentioned in the episode, ctl.columbia.edu/podcast. Please like us, rate us and review us on Apple podcasts or wherever you get your podcasts. Dead Ideas is produced by Stephanie Ogden, Laura Nicholas, John Hanford, and Michael Brown.

[00:29:47] Our theme music is *In the Lab* by Immersive Music.