

Bias of Student Evaluation of Teaching

(Dr. Birgit Hawelka in conversation with Prof. Dr. Bob Uttl)

Birgit Hawelka:

Hello and welcome to today's episode of our podcast series. I'm Birgit Hawelka and with me today is Bob Uttl. Bob is a professor of psychology at Mount Royal University in Calgary, Canada. His expertise lies in cognitive psychology, focusing on areas such as memory, intelligence, aging, assessment, and psychometrics. Bob has contributed to several high impact papers, particularly examining the validity and biases of student's evaluations of teaching effectiveness, in short, SET ratings. His work on SET-related research has caught the attention of various news outlets, including Inside Higher Ed, University Herald, The Guardian and more. Bob, a warm welcome, good to have you here.

Bob Uttl:

Thank you for invitation.

Birgit Hawelka:

So let's start directly. Bob, in your work on student evaluation of teaching, you've been criticizing the use of standardized instruments for measuring teaching effectiveness. Could you expand what arguments and findings led you to this conclusion?

Bob Uttl:

Yes. So the first problem which we have is that there is no accepted definition of effective teaching. So we do not know what it is that we actually want to measure. And people have thought about it, there's lots of experts discussing this issue. And basically, they all agree that the only thing which we know is that the effective teaching ought to lead to increased student learning. And so from that, people came up with the idea of these multi-section studies which were designed to sort of examine the validity of SETs, that is student evaluation of teaching ratings, relative to professors' contribution to student learning. So basically, what you have is you have some large course which you split into a number of sections. And professors get randomly assigned into sections, and then they do the teaching using the same syllabus, same book, same exams, etc. and then, towards the end of the course you give them SETs. And if professors are effective in increasing student learning, that professor's class ought to do

better than other professors', so there should be correlation between SETs, average SETs of the class and the average final exam grade.

Birgit Hawelka:

Okay. And is it the case?

Bob Uttl:

Well, Cohen, in 1981, in the famous meta-analysis said that that is the case. He said that the correlation of something like 0.41 or 0.43, you can wonder whether that's a sufficient correlation to evaluate professors on. Because it of course explains only about 16% of variance. But there's a fundamental problem with that meta-analysis: It did not consider bias of small studies. And so if you actually redo the meta-analysis, which is what we did, you find out that the correlation between SET and student learning attributable to the professor in multisection-studies is zero, after you control for students' prior knowledge and intelligence.

Birgit Hawelka:

Ok, so intelligence is the most important factor that influences the learning outcomes. Would you agree?

Bob Uttl:

I am not sure if it's a most important factor. It is certainly a very important factor. But there's also another important factor, and that is student motivation. So, how much is a student interested in actually studying the particular topic in a particular course? It is a well established finding in cognitive psychology that if you are interested in learning something, we learn it much faster than if we don't want to learn it.

Birgit Hawelka:

However, would it be possible that lecturers might influence the students' interest in the subject they're teaching and so they would indirectly be kind of responsible or able to influence the learning outcomes?

Bob Uttl:

It might, but we know from research now that prior student interest in a course before they take it, is one of the best determinants of what student evaluation of teaching ratings students actually give.

Birgit Hawelka:

So as we were talking about, biases lie in students' abilities or motivation. Are there are some other bias that could influence or distort the results of student evaluation?

Bob Uttl:

Yeah. So, you know, basically we know that the validity is zero, right? And the second way how to get into validity is to look at whether these SETs are influenced by various factors

which have nothing to do with the professor. So, for example, you can have a class size, you get assigned to teach small classes, let's say up to 20 people. You will have more time to interact with each student. You will be able to learn every student's name. Whereas if you are assigned a class with 300 people, good luck to you. Unless you are some kind of a special person who can memorize 300 names, you stand no chance, right? If you have a class with ten people then, of course, you will know everybody and you will have interaction with everybody, right? So the class size makes the difference.

Another big effect is - we talked about student abilities, too, and motivation, but also field. So today if you are assigned to teach some kind of a quantitative class, good luck to you. Because almost no one in your class is actually interested in taking it. So in psychology we have, you know, we have a requirement for statistics and research methods because of course, we rely on that. And we have done a survey of students asking them how interested they are in taking various classes, which our department offers. And classes which involve statistics or research methods were six standard deviations below the mean of our normal content classes. And of course, people are most interested in taking something like abnormal psychology, social psychology, and so on. So now a professor teaching abnormal psychology is facing students who are, by and large, very interested in the topic. A professor facing introductory statistics is facing students who don't want to be there.

Birgit Hawelka:

That would apply to mandatory courses, I think. They should be poorer rated than courses the students can choose by themselves and by their interest, right?

Bob Uttl:

That's right, yeah. There is some research looking into that, right? And so of course, courses which are elective, which you want to take, are typically rated higher than courses which are required.

Birgit Hawelka:

Okay. So we have some student characteristics, we have some course characteristics that might influence the student ratings. In one of your texts, you mentioned a study suggesting that incentives like giving chocolate can influence student evaluation. How would you explain such strange findings? And what psychological factors could be swaying students' opinions in this particular instance.

Bob Uttl:

I don't think it's a very strange finding, right? So we have two randomized studies showing that chocolates and chocolate cookies improve student evaluation teaching ratings by about 0.3 to 0.5 standard deviation, right?

Birgit Hawelka:

This is a lot.

Bob Uttl:

That's a lot, yes. It makes the difference, right? And it's a simple explanation. You bring something which students like, they will give you higher ratings. If you say that SETs are basically measures of student satisfaction and dictionary definition of that is, you know, happy, pleased feelings because of something that happened to you, right? So anything which increases happy, pleased feelings on the part of the students is likely to increase student evaluation of teaching. So low workload, high grades, chocolates, home baked cookies, anything which you can do to increase student satisfaction.

And so SETs – we didn't talk about this – but one question, of course, is: How are these used, right? And so in some cases or in many cases, some kind of committee or perhaps departmental chair decides that anybody who's above the average is satisfactory and anybody who's below the average is not satisfactory. So now you set up a race, you have to be in the top half of a racing professoriate to be satisfactory and not having to go to talk to your chair or to your dean about your low student evaluation of teaching.

In fact, in one commercial system, they recommend that only the top 30% of professors are satisfactory. So 70%, by definition, are not satisfactory no matter what they do. You may be an Olympian, but you are still unsatisfactory because you are in the bottom 70%, right? And so that sets up a race of really, really high stakes and anything goes: So cookies, lower the standards, give grades for nothing. We have a massive increase in grade inflation. Today the average grade is basically A- which would be unheard of 40 years ago, right? And so all these factors influence SETs. Professors learn what influences SETs. So, before we had the SETs online, you could see professors going around with bags of chocolates or chocolate bars, distributing SETs in paper and giving chocolate bars with it in order to make students happy.

Birgit Hawelka:

If I summarize for the moment, it can be said that there is no objective method for measuring teaching quality. However, SETs then perhaps measure how well teachers manage to meet students' expectations in good quality teaching?

Bob Uttl:

I would not say that it measures student expectations in quality of teaching. I would say that it measures student satisfaction. And the student satisfaction is encompassing, right? So if I want to get an easy grade and you satisfied me, then I will be happier than if you give me a D.

Birgit Hawelka:

Do students know their grades in Canada or at your University before they are doing the SETs?

Bob Uttl:

No, they don't. But basically they do because, of course, you typically have at least one mid-term. And so of course, SETs are after that midterm so that it gives you some idea. And for example, if you are a teacher or professor who assesses the students by multiple assignments or perhaps multiple tests, then the students will basically know what their grade is by the time they get into SETs.

Birgit Hawelka:

The American as well as the German Sociological Association, suggest using SETs not for summative evaluation but for formative evaluation. They recommend it as an opportunity to gather student feedback, their point of view on different aspects of a course. Would you agree to such scenarios?

Bob Uttl:

Yeah, I don't think there's anything wrong with a professor surveying his or her own students to find out how they perceive various aspects of your course. That seems not only reasonable but it also seems desirable. But it has to be the survey by the professor of the students in a professor's class which no one else sees. Because then the professor who is actually teaching the class knows what was going on, knows what questions he or she wants to actually ask, which are somehow important. Maybe you are teaching the particular class for the first time. You will not know what the students had most trouble with, so you may actually ask them. You may use a different software, get a feedback on that software. So those are the important uses of student feedback, but not for the evaluation of the professor's teaching effectiveness.

Birgit Hawelka:

So as we learned, and to sum up the gist of your findings, it seems you've concluded that for different reasons, measuring a teacher's effectiveness is hardly possible. But flipping the script: Can we identify and measure teaching behaviors that clearly signal ineffectiveness? So are there some signs you could use to say: Oh, this is a critical point, this might be quite ineffective in teaching.

Bob Uttl:

Yeah. So, a number of people actually point out that, you know, when we are trying to think about what is effective teaching, that people disagree, right, that they just cannot come to any agreement. You know, to give you an example of... somebody may say, well, it's an effective teaching to call on students during the class to answer some question. Other professors will say, well, no, that's anxiety producing, it's putting the student on the spot. It's demeaning. It's negative. Right? However, almost everybody would agree that not showing up for your classes is not effective teaching. Showing up in your class, reading a chapter from the textbook verbatim is also not effective teaching. So there's a number of these things which lots of people would agree, right? So even if you say, giving students feedback. Giving students feedback two months later, that's not very effective. And so you could have some kind of a survey, you know, I propose one survey like that in the latest paper, where you ask students to say whether certain things, objectively observable things, actually happened in the class. Let's say I got a feedback or response to my email within three business days. And so now, if 50% of the students say, well, you know, the professor didn't show up in class and the other 50% say he did, well, somebody is not telling the truth, so to speak. And now you can actually go and verify what actually happened. And so use it as an alarm for, you know, ineffective teaching. And in fact, if you know, if you are around university for a decade or two, you will realize that those ineffective teaching really fall into those categories. You know, people not showing up for

classes, dismissing the classes halfway through or something of that sort, right? Not providing the feedback, students don't get the grades ever, maybe the final one, but that's it. So those are ineffective teaching behaviors.

Birgit Hawelka:

Okay. And according to your findings, it might be helpful to identify such points or signals. In your latest text you summarize some of them as an indicator for ineffective teaching. That might be the basis for the deans to talk about what happens in this class, or to identify ineffective lecturers, right?

Bob Uttl:

That's right. It would simply be as an alarm, right? It should not be used as an evaluation. But simply, okay, students are in the class. We also need to consider the fact that, you know, when we say students are in a class, some are and some are not. So it's an important thing to sort of remember, right? But then when you have some of these objectively observable behaviors and 90% of the students say, yeah, he didn't show up most of the time. Well, you should probably look what happened and then, you know, determine what actually happened. And then you have some basis for some kind of evaluation or corrective action or something like that.

Birgit Hawelka:

Okay, so instead of trying to measure teaching effectiveness, you would suppose to identify more ineffective teaching. So to flip it around.

Bob Uttl:

That's right, yeah. So you basically look for ineffective teaching and to what you could do about it, right. There's another method which I discussed in that paper, which was used in one university in Alberta, where they basically use a statistical definition of an outlier. So as long as you are within the bounds of their racing professoriate and you are not an outlier, you are not the straggler back in the end, you are fine. But if you are an outlier, then that causes an alarm, right? It doesn't mean that you are ineffective, but it causes alarm. Where a chair ought to be looking at, you know, what's going on with this course. And maybe you taught it first time. Maybe you had some really bad class. Maybe they put you in the class with 20 chairs and you had 30 students, so ten of them had to stand. Even stuff like that happens, right?

Birgit Hawelka:

Okay, Bob, thank you so much for making the time to be here with us and for sharing your valuable insights on student evaluations.

Bob Uttl:

Thank you for having me here.