

Can students evaluate teaching quality objectively?

By [Anne Boring](http://www.anneboring.com), OFCE-PRESAGE-Sciences Po and LEDa-DIAL, www.anneboring.com.

The author will present her work at the International Symposium on Gender Bias in the Governance and Evaluation of Research Bodies organized by the EGERA ([Effective Gender Equality in Research and the Academia](#)), which will take place on 23 February 2015 at Sciences Po, on the CERI premises in Paris.

Anglo-American universities generally rely on the evaluation of teaching by students to measure teaching quality. They hypothesize that students are the best placed to judge the quality of teaching in that they observe the teachers throughout a course. The evaluations usually serve two purposes. First, they are used as a tool for pedagogical management for the teachers themselves, by providing them with suggestions for improving their teaching; and second, these evaluations are also often used by the administration to make decisions about promotions or the extension of teaching contracts. The evaluations then act as incentives: they encourage teachers to give the best of themselves so as to be rehired the following semester or to obtain a promotion.

In France, the practice of evaluating teaching is still not very widespread, but many higher education institutions are planning to develop it. Some private schools already use it in their recruitment policy or to extend the contracts of supply teachers. As for the public institutions, they use the evaluations of teaching only to help teachers improve their pedagogical methods. Public institutions are obliged to comply with a directive from the French Ministry of Higher Education and Research which states that “evaluation results” can be

disclosed “only to the teacher concerned and not the head teacher or principal of the institution”.[\[1\]](#) This Directive upholds a 1997 decision of the French State Council, which indicates that the procedure for evaluating teaching should “simply allow teachers to have a better understanding of how the educational dimensions of their teaching are appreciated by the students”, and “it does not include or imply any impact on teachers’ prerogatives or careers”. Thus, only the teacher concerned may have “knowledge of the elements of this type of evaluation”.[\[2\]](#)

Regardless of whether the end use of this supervisory tool is the improvement of teaching or the management of the teaching teams, universities need to be sure that student evaluations are an objective measure of the quality of teaching. To do this, at least three conditions need to be verified:

- 1) that the students know how to measure the quality of teaching, that is to say, they are able both to establish criteria that define teaching quality and to use these criteria to judge the teacher;
- 2) that the students are not biased in their judgments and assessments; and
- 3) that the teachers cannot adopt strategic behaviours to secure good evaluations; in other words, that efforts to obtain good evaluations do not lead teachers to engage in behaviour that could undermine educational quality.

Do students know how to judge the quality of teaching? (Condition 1)

What teacher has not been in a discussion with colleagues where everyone defended his or her own teaching method as being “the best”? These discussions generally centre on the content of teaching and how to transmit this content, as well as on different ways to check on students’ learning. It is not easy to determine the criteria that define good teaching

quality, and the professionals themselves disagree. Yet the system of evaluation assumes that students are able to do this to some extent at least.

In the students' view, what criteria are important for determining the quality of teaching? The literature suggests that students believe that one essential criterion is the teacher's extroversion and dynamism, that is, their ability to capture attention (e.g. Radmacher and Martin, 2001). Several research studies tend to confirm that students seem to give priority to how a lesson is taught, rather than to the educational quality or the content what is being taught.

Consider the "Doctor Fox" effect (Naftulin, Ware and Donnelly, 1973), which makes reference to friendly teachers who can get good ratings by giving the impression of being competent, without however teaching relevant or good-quality content. In this example, which is well known in the United States, researchers hired an actor to teach a lesson on a fictitious subject. The course featured numerous neologisms and meaningless assertions, and the idea of the three researchers who hired the actor was to determine whether people attending it were able to detect this without being blinded by the lecturer's flair, self-assurance and academic authority (he was given a false resumé: a full range of prestigious fake diplomas and fake research papers). At the end of Dr. Myron Fox's course, those who attended gave him a positive evaluation. This experience shows first that the students' perception of a teacher's academic authority matters, and, second, that students are not always able to judge the content of what is taught.

Likewise, according to Carrell and West (2010), the perception that students have of teaching quality is not necessarily correlated with the actual quality of the course, when the latter is measured by long-term success. These authors show that evaluations are correlated positively with the students' short-term success, but not with longer-term success. Their

results suggest that teachers whose pedagogical techniques encourage cramming might be better assessed than teachers who use more demanding and difficult teaching techniques but promote the long-term learning of knowledge. Indeed, students are often primarily concerned with their success on final exams, rather than the future usefulness of the knowledge acquired during the semester. Universities need to develop incentives for teachers to use teaching methods that promote long-term learning, methods that do not always seem to be rewarded by students in their evaluations.

Are students' judgements on teacher quality unbiased? (Condition 2)

The evaluation of skills can be subject to bias on the part of the evaluators. The literature on social psychology in particular suggests that it is more difficult for people from minority backgrounds to be perceived as competent (even if they are), while it is more difficult for people from majority backgrounds to be perceived as incompetent (even if they are). Stereotypes and double standards for evaluation have an impact once it comes to determining individual competence (e.g. Basow, Phelan and Capotosto, 2006; Foschi, 2000). This impact can have especially negative consequences for certain minorities, in particular women university professors, who are still in a minority.

A study of evaluations by freshmen at a French higher education institution [\[3\]](#) showed that students do in fact apply many gender stereotypes in the way that they assess their teachers. The results of this econometric analysis show that male students tend to give better evaluations to male professors than to females. Male professors on average benefited from a bias on the part of male students in almost all the dimensions of teaching, in particular the quality of the presentation, the ability to be in touch with the latest developments, and participation in the student's intellectual development. The female students also tend to evaluate men

more favourably on these criteria, but give more favourable evaluations to women on other teaching dimensions, including the preparation and organization of the lessons, the usefulness of the class materials, the clarity of their evaluation criteria and the relevance of their corrective comments. The bias in the responses of the male and female students in favour of men on the criteria related to the presentation of the lessons in particular led to higher overall satisfaction scores for the male professors. However, other measures of teaching quality (such as exam results) tend to show that the education provided by women was as good as that provided by men. Furthermore, some teaching tasks for which women professors were more highly valued (only by women students) tend to be time-consuming. The women professors then find themselves with less time for other professional activities, such as research, for example.

Do teachers adopt strategic behaviours that undermine teaching quality (Condition 3)

Finally, numerous studies show that teachers can adopt strategic behaviours to improve their scores. Indeed, with the introduction of student evaluations, teachers are faced with the problem of the multitasking agent (Holmstrom and Milgrom, 1991; Neal, 2013): they must teach well, while getting good evaluations – goals that are not necessarily compatible, as Carrell and West (2010) demonstrate. The two strategic behaviours studied in the literature are a teacher's capacity for demagoguery (the Dr. Fox effect), on the one hand, and generosity in scoring student work, on the other. Although there is no consensus as to the causal link between good scores given by teachers and good ratings given by students, it has been shown that the two are correlated (e.g. Isely and Singh, 2005).

Conclusion

Evaluations by students do not seem to meet the three

conditions for an objective measurement of teaching quality. The question can also be raised as to whether the nature of educational activity can be measured objectively at all. But does this mean we should not set up systems for student evaluations? These evaluations can be useful, but they should be interpreted with caution and be taken for what in all likelihood they actually are: a measure of the pleasure that students have in going to the lesson rather than a single, objective measure of the overall quality of teaching. The pleasure that a student feels in going to class is just one ingredient among many in good quality education. It is also necessary to try to take into account and correct the biases that students express in these evaluations by weighting the evaluation criteria so as not to discourage or unfairly penalize certain categories of teachers, especially women, whose evaluations are not as good simply because of gender stereotypes.



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 612413.

References

Basow, S. A., Phelan, J. E., & Capotosto, L. (2006). Gender patterns in college students' choices of their best and worst professors. *Psychology of Women Quarterly*, 30(1), 25-35.

Carrell, S. E., & West, J. E. (2010). Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors. *Journal of Political Economy*, 118(3), 409-432.

Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology*, 21-42.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 24-52.

Isely, P., & Singh, H. (2005). Do higher grades lead to favorable student evaluations?. *The Journal of Economic Education*, 36(1), 29-42.

Naftulin, D. H., Ware Jr, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Academic Medicine*, 48(7), 630-635.

Neal, D. (2013). The consequences of using one assessment system to pursue two objectives. *The Journal of Economic Education*, 44(4), 339-352.

Radmacher, S. A., & Martin, D. J. (2001). Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. *The Journal of Psychology*, 135(3), 259-268.