Student Evaluations of Teaching and Teaching Effectiveness: Ryerson University's Arbitration Decision

An arbitrator ordered Ryerson University in Canada to amend its faculty collective bargaining agreement to ensure that student evaluations of teaching (SETs) are not used to measure teaching effectiveness for promotion or tenure. The following paragraphs speak to some of the reasons the arbitrator used to justify his decision:

> *While SETs are easy to administer and have an air of objectivity, appearances are somewhat deceiving. SET results make it easy to compare Professor A with Professor B, and to others in the Department, Faculty and University. SETs seem scientifically sound: the objective correlation of numerical data. But upon careful examination, serious and inherent limitations in SETs become apparent (discussed further below). The expert evidence led at the hearing persuasively demonstrates that the most meaningful aspects of teaching performance and effectiveness cannot be assessed by SETs. Insofar as assessing teaching effectiveness is concerned – especially in the context of tenure and promotion – SETs are imperfect at best and downright biased and unreliable at worst.*
>
> *According to the evidence, which was largely uncontested, and which came in the form of expert testimony and peer reviewed publications, numerous factors, especially personal characteristics – and this is just a partial list – such as race, gender, accent, age and "attractiveness" skew SET results. It is almost impossible to adjust for bias and stereotypes. Student and faculty gender affects outcomes, as does grade expectation. Other systemic problems were identified by the experts, and in the literature. One example, and there were many, is the reliability of SETs completed on line versus those completed in class. There are differences between the results of absent students who complete SETs online and those who complete the forms in class. These differences need to be understood. Overall response rates also need to be considered: the lower the response rate the less reliable the results. There is certainly no reason to believe that the views of responders can be extrapolated and applied to non-responders.*
>
> *Course characteristics – elective versus required – class size – small versus large – hard versus soft – quantitative versus humanities – traditional teaching versus innovative – these are just a few of the other factors that can impact results. Some traditional SET questions about teaching effectiveness are highly problematic, as are questions that seek to measure the breadth of the instructor's knowledge and scholarship, i.e. whether the curriculum was up to date. It is far from clear whether students have the expertise to comment on course content and teaching methods and assignments. Moreover, if faculty believe SETs will be used to assess performance, they may modify behavior – teaching to SETs – to achieve desired outcomes, ones that have little to do with what is supposedly being assessed. SET results also have a demonstrated correlation with student grade expectations, and the timing of the administration of the SET may influence its reliability. The list goes on. And this is just the briefest of summaries of a small part of the voluminous evidence adduced and filed at the hearing about the myriad problems in relying on SETs to evaluate teaching effectiveness.*
>
> *The fact of the matter is that SET results have demonstrable limitations that raise real issues about their use as a measure of teaching effectiveness in tenure and promotion decisions. A further complication is the practice of reducing the FCS results to averages and then comparing individuals with other individuals, the Department, Faculty and University. The evidence is clear, cogent and compelling that averages establish nothing relevant or useful about teaching effectiveness. Averages are blunt, easily distorted (by bias) and inordinately affected by outlier/extreme responses. Quite possibly their very presence results in inappropriate anchoring.*

Source: https://www.canlii.org/en/on/onla/doc/2018/2018canlii58446/2018canlii58446.html