

## Why I have a problem with our use of Student Evaluations of Teaching (SETs) ...and what we can do about it

*Aziza Ellozy, Associate Provost for Transformative Learning and Teaching*

Last November, I attended a conference in Dubai the theme of which was “*Continuous Learning in the Journey of Education: Using Analytics to Measure Progress and Inform Strategic Decisions*”. My main interest in the conference stemmed from the effort we have undertaken to redesign the student evaluation survey and to find out how we can better leverage the new platform (Blue) that is being used for these evaluations.

Several of the sessions at the conference dealt with the theme of Student Evaluation of Teaching (SETs), and I was invited to participate in a debate on the motion “*Student evaluations of teaching are an important assessment tool for evaluating overall faculty performance*”. By choice, I debated against the motion and I had no problem building my argument because it is something that I have been discussing lately with other faculty at AUC.

But first let me try and put this discussion in its proper context and make a couple of assertions:

1. Although I have concerns about SETs in general, my main concern is that only very few departments at AUC triangulate these SET data with other sources and types of data (see #2).
2. Teaching and learning is a complex and multidimensional phenomenon, and assessment of teaching requires a multidimensional approach which would include ongoing formative (i.e. developmental) peer assessment, as well as summative peer evaluation, reflective faculty self-assessment, and teaching portfolios that could provide reinforcing or conflicting assessments.
3. Although there is continued debate over the merits of using SETs for renewal of contract, promotion and tenure decisions (Harrison, Douglas, & Burdsal, 2004), and although I will discuss in what follows what the limitations of SETs are, the **wrong conclusion would be to ignore or dismiss them.**

SETs will continue to be part of the summative evaluation of faculty teaching but my hope is: 1) that we can collectively talk about their limitations, 2) that chairs and committees understand the variables that might affect them and interpret the results adequately and 3) that chairs and faculty supplement them with multiple other sources and types of data.

### So what are the criticisms leveled against SETs?

Most of the debate that is taking place in the literature focuses around the usefulness and validity of the ratings of SETs (Benton, 2011; Stark, 2014; Horstein, 2017). A quick overview summarizes the problems as follows:

- Critics question the validity of SETs as the main tool to measure teaching effectiveness. For a tool to be **valid**, they say, it should measure what it is intended for, and the assumption that they measure **teaching effectiveness** is a wrong assumption. Unfortunately, there is no consensus in the literature about what “teaching effectiveness” is and there is no agreed upon single, all-encompassing yardstick to measure it. Different institutions and different disciplines tend to develop their own guidelines/framework and many do not. As far as I know at AUC, we do **not** have a *university* framework or rubric to determine teaching effectiveness. One such framework is included on p3 of this newsletter. This framework was presented by J. Todd, “Developing and Evaluating Teaching Effectiveness” at the ‘Leadership in Higher Education conference’ (October 2019) and has been adopted by her institution, Colorado State University. I have the permission of the author to share it, in the hope that departments or schools will be inspired to develop their own.
- One could argue that the most important indicator of teaching effectiveness is increased student learning, which unfortunately is a hard thing to measure given that grades are not necessarily a good gauge for learning. Recently a meta-analysis (Uttl et al, 2016) of multi-section studies showed that SET ratings are unrelated to student learning.
- Other limitations of SETs cited in the literature include:

- A bias with respect to gender, perceived age of instructor, perceived difficulty of the course, whether the course is required/elective, etc (Benton, 2011).
- Low response rates which have become a general phenomenon since switching from in-class surveys to online surveys. This affects sampling error: the smaller the sample size the higher the margin of error. Low response rates have also been attributed to other reasons including timing of the survey near finals, a perception of possible lack of anonymity, a perception that no action is taken based on the feedback, etc.
- Averaging SETs scores and comparing between courses of different types, levels, sizes, etc., can be statistically questionable (Stark, 2014)
- Using SETs as the dominant if not sole indicator of teaching effectiveness in tenure and promotion cases can lead faculty to focus on student satisfaction and not challenge their intellectual capabilities, which in turn may result in grade inflation. There is a concern that faculty on term contracts are particularly vulnerable to this pressure.

Given these limitations of the SETs we need to be thoughtful as to how we use them, and faculty/administrators should interpret the data with attention to context. They should take into consideration the possibility of the reasons behind the feedback and perhaps try to address them.

So one might ask what good are SETs? The simple answer is that they can definitely give us information on students' perceptions and experience in the course. Students, who are the recipients of instruction, are in a good position to give feedback on the management of the course, are able to judge if the lectures are organized and would know if the instructor is available outside of class, and whether or not feedback is given promptly etc. What they are not equipped to evaluate is course pedagogy nor necessarily how knowledgeable an instructor is about his/her subject area.

So, let me go back to my original concern: unless we have a solid peer evaluation process (which we do not have at present with the exception of a couple of departments) accompanied by a teaching portfolio with reflective faculty self-assessment, we will continue to use SETs as the sole criterion on which to base the decision to renew faculty contracts, or to decide whether they should get tenure and/or be promoted, and that, in my opinion, is a flawed approach.

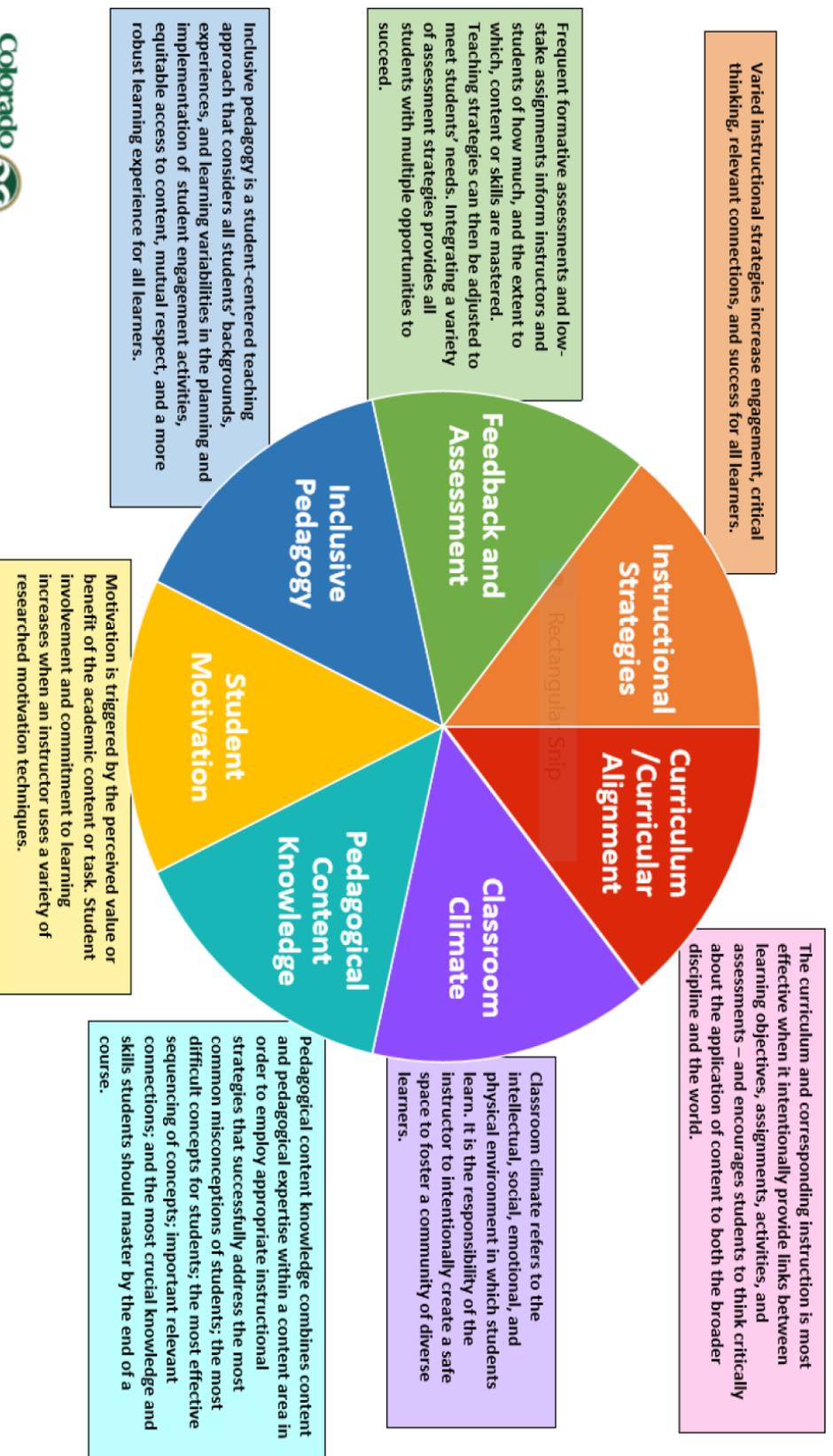
To sum up, this newsletter is a call to attention to the most important recommendations of the Task Force on the Quality of Undergraduate Education in 2017, namely "to develop an updated and rigorous comprehensive teaching evaluation process". To date, the student evaluation survey has been redesigned and has gone through several iterations after feedback from a great many stakeholders. Right now, it is in the senate awaiting approval. Our real challenge is to establish a quality process for peer review of teaching at AUC. I hope that most departments/schools will join me in adopting and institutionalizing one such program that would be worthy of our faculty and our profession.

In the next few weeks the Center for Learning and Teaching will be offering one-hour workshops for faculty interested in serving as Peer Observers for their departments and for others. It is my hope that Chairs will encourage a couple of their faculty to take advantage of this opportunity.

## References

- Benton, S.L. & Cashin, W.E. (2011). IDEA Paper No. 50: Student ratings of teaching: A summary of research and literature. *The IDEA Center*  
Retrieved from  
[https://www.researchgate.net/publication/308904846\\_IDEA\\_Paper\\_No\\_50\\_Student\\_ratings\\_of\\_teaching\\_A\\_summary\\_of\\_research\\_and\\_literature](https://www.researchgate.net/publication/308904846_IDEA_Paper_No_50_Student_ratings_of_teaching_A_summary_of_research_and_literature)
- Horstein, H.H. (2017). Student evaluation of teaching are inadequate assessment tools for evaluating faculty performance. *Cogent Education*, 4(1).  
DOI: <https://doi.org/10.1080/2331186X.2017.1304016>
- Rowan, S., Newness, E.J., Tetradis, S., et al. (2017). Should student evaluation of teaching play a significant role in the formal assessment of dental faculty? Two viewpoints. *Journal of Dental Education* 81(11): pp. 1362–72.
- Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. DOI: 10.14293/S2199-1006.1.SQR-EDU.AOFRQA.v1Stark.
- Uttl, B., White, C. A., & Gonzalez, D. W. (2016). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, pp. 54, 22–42.

# Teaching Effectiveness Framework at a Glance



The Institute for Learning Teaching (ILT), Version 1.0

(c) 2019 Colorado State University 2019 CC BY-NC-ND 4.0