A Systems Approach to Re-envision Course Evaluations at LUMS

Janette Barrington, Ph.D., August 2019

Summary

LUMS has a proud history of developing indigenous and innovative case-based teaching. It has established itself nationally as a quality teaching institution. To build an international footprint, LUMS embarked on a trajectory of research intensity which continues to be an institutional priority. To rebalance and strengthen a research-teaching nexus, LUMS is providing career progression for faculty to achieve excellence in both core functions of research and teaching. More broadly, LUMS is a learning institution that seeks to accomplish this dual goal in part through a renewed commitment to teaching excellence and student development.

I was invited to LUMS in May 2019 to discuss with administrators and a sample of faculty and students, the course evaluation system in place and, more broadly, to contextualize its role in the evaluation of teaching. One overarching observation from these interviews is that perceptions at LUMS mirror findings that have been documented across institutions around the world. These findings confirm that course evaluations play an important role but are also a source of controversy.

This report offers a framework for re-envisioning course evaluations by referring to the literature on course evaluations and the current discourse on the evaluation of teaching[1]. The report also recommends how the framework can be operationalized through best practices and examples from other universities. The following five recommendations provide a systems view for decision makers to adapt operational aspects that would best serve the multiple stakeholders who interface with course and teaching evaluations (Table 1 below).

These ambitious recommendations build on a widespread view that current practices offer new windows of opportunity for a system-wide transformation and the time for change is now.

Table 1

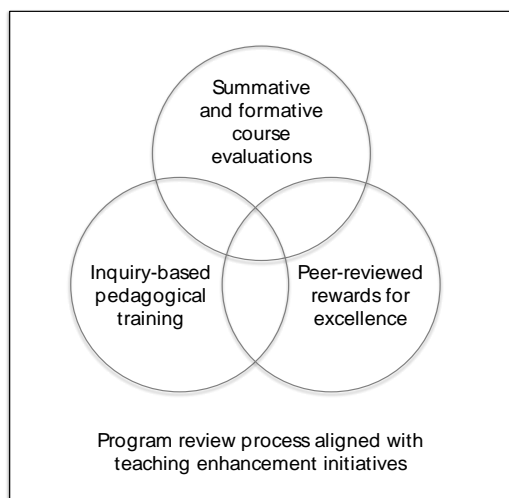| |
|---|
| **1.** **Develop a new policy for course evaluations as a vital part of an overall learning and quality enhancement system.** |
| **2.** **Adopt a cascading evaluation framework to create ownership and shared responsibility in course evaluations.**
2.1. Create a LUMS-specific course evaluation tool modeled after the University of Toronto's instrument.
2.2. Explore alternative strategies to promote student response rates.
2.3. Address what statistics reveal and their limitations. |
| **3.** **Strengthen support for teaching development and effective pedagogical practices to help faculty operationalize the data collected.**
3.1. Establish feedback mechanisms such that students see tangible results of their input.
3.2. Formalize support mechanisms that nurture reflective practice through critical inquiry, coaching and dialogue.
3.3. Link enhancement initiatives to the new Student Partnership Initiative. |
| **4.** **Monitor and analyze course evaluation data for a range of institutional and School level enhancement activities.** |
| **5.** **Develop new policy for achieving tenure on the basis of teaching excellence as well as guidelines for annual and milestone reviews**. |

Implementation Plan

Recommendation 1: Develop a new policy for course evaluations as a vital part of an overall learning and quality enhancement system.

Course evaluations—typically in the form of an end-of-course survey—are but a single element embedded in multiple interconnected processes aimed at continuous improvement of teaching. Where the data come from, where they go, how they are interpreted and used, how they are narrated and discussed, and with whom, to what end, and what decisions or actions do they influence—all give indications of other parts of the system that can be informal and/or tacit. Discussions about course evaluations benefit from keeping this systems perspective in mind to avoid creating unintentional cross-purposes within these systems[2,3,4].

Like most universities, the course evaluation system at LUMS operates in the background with little or no shared understanding of what effective teaching means. Committees mandated over the years to review and update the system immediately debate the tool rather than the bigger picture. Scholars, however, remind us that course evaluations are part of the larger system of evaluating teaching so, practically, re-envisioning course evaluations will require taking stock of existing evaluation structures and processes.

Figure 1: Lund's systems approach to enhancement

Summative and formative course evaluations

Inquiry-based pedagogical training

Peer-reviewed rewards for excellence

Program review process aligned with teaching enhancement initiatives

Building a shared understanding of what a larger systems approach to course evaluations can potentially offer is key to sustaining it. The Faculty of Engineering at Lund University in Sweden[5], for example, views the enhancement of teaching as a system of multiple interconnected initiatives that includes course evaluations, training support, rewards for excellence and program review (Figure 1). Central to this approach is the belief that 'more and better conversations about student learning lead to better quality education.' In other words, changes in practice come from an evolving conversation on how to improve learning among faculty in partnership with students. And this is far more effective than isolated trial and error interventions based on student ratings.

Another example from the University of Toronto[6] lays out a course evaluation framework structured around principles which incorporate institutional priorities, inclusive approaches to address diversity, equity and transparency, input from students about their learning experience in the form of formative and summative feedback, anonymity and confidentiality of student respondents and assurance of reliable and meaningful data to instructors, administrators and students.

A new policy is needed at LUMS that similarly outlines the principles guiding course evaluations and, at the same, situates them as a vital part of an overall learning and quality enhancement system. A culture of experimenting with evidence-based and effective teaching methods that are documented in a trajectory of improvement will likely attract a large segment of faculty who can focus on learning and teaching as a way to gain permanence and tenure.

Recommendation 2. Adopt a cascading evaluation framework to create ownership and shared responsibility in course evaluations.

Faculty want course evaluations for self-authored, professional development purposes—to get feedback on their teaching for the purpose of helping guide their course design decisions and classroom practice[7,8,9]. Administrators think of course evaluations as performance indicators for merit and remediation purposes[10,11]. Students see them as a means of having input in their education[12,13,14]. How the student voice is integrated more productively into course evaluation processes receives increasing attention among higher education researchers[14,15,16,17].

The LUMS community has raised a number of concerns about course evaluations (Appendix A). Students feel disengaged from the process and believe their feedback is neither considered important, nor taken seriously. Faculty are generally dissatisfied and perceive the evaluation process to be unreliable and disconnected from promotion and tenure decisions. And administrators struggle to balance these views in light of institutional priorities to measure teaching effectiveness. Together, these conditions point to a course evaluation system challenged by concerns of fairness, credibility and trust.

The University of Toronto faced similar challenges in developing their new Cascading Student Evaluation Framework[18]. The framework is designed to address teaching and learning priorities at different stakeholder levels. At the institutional level, senior administrators agree on priorities (of student engagement, deep understanding, an atmosphere conducive to learning, and opportunities to improve and demonstrate understanding) that translate into respective course evaluation questions to be asked of all courses at the University, regardless of the discipline. These 4 questions plus an overall question reflect institutional priorities (Appendix B). The School and Department (or Program) each agree on 3 additional questions (for a total of 6) for all of their courses to give contextualized meaning to the results regardless of the program students are enrolled in. And, finally, individual instructors select up to 4 questions from a validated and reliable pool of questions that can be used for formative purposes to enhance their teaching. The survey should not exceed 15 questions plus two open-ended questions for written comments.

2.1 Create a LUMS-specific course evaluation tool modeled after the University of Toronto's instrument.

When deciding on what course evaluation questions to include in a LUMS specific tool, it is important to remember that valid questions ask students to report on what they are qualified to report on—their individual learning experiences. The Course Experience Questionnaire (CEQ)[19] is designed for this purpose and could be used to jump-start LUMS' own bank of validated questions. Grounded in an evidence-based theoretical framework, the CEQ (Appendix C) asks students to distinguish features in a course known to encourage deep over surface approaches to learning. High scores indicate a course that encourages conceptual understanding rather than rote learning[20,21,22]. Lund University has successfully implemented the CEQ over the past 20 years with the explicit goal of informing changes in practice.

Developing a LUMS-specific questionnaire provides an opportunity to communicate institutional priorities laid out in the Vice-Chancellor's White Paper on New Directions, to respect the disciplinary differences of each School and department, and to allow faculty to select questions at the course level. One further recommendation is to include students as a stakeholder group not only in validating questions (the practice at the University of Toronto) but also in selecting questions. Students are ideally situated to contribute questions on what helps or hinders their learning.

It is anticipated that the new program enhancement office responsible for overseeing the course evaluation system would finalize the new tool with the anticipated goal of having it in place for the 2020-2021 academic year. The IST team will need to be consulted to ensure they can accommodate dynamic uploading of questions as well as update and validate questions on an ongoing basis (more on this in recommendation 4).

2.2   *Explore alternative strategies to promote student response rates.*

The biggest issue raised with respect to the online course evaluation system at LUMS is response rates. When first administered online, response rates dropped dramatically so an incentive scheme was put into place resulting in a university-wide average of 75%-95%. However, the practice of penalizing students for not responding by not allowing them to access their grades means they do it hurriedly or out of fear and this is perceived as unfair. Achieving target response rates often requires specific strategies to inspire student participation, but these strategies should not include offering participation incentives[23]. Acceptable alternatives include: allowing time in class to complete the online evaluations, promoting the importance of student feedback, and making evaluation data public.

Part of validating the new course evaluation tool will require considering the proportion of students who must respond to have confidence in the ratings. A high percentage of non-responders introduces uncertainty into the evaluation data and can make it impossible to generalize feedback to a whole class. Table 2 establishes the target percentage of responses to be meaningful for different class sizes as reported by the University of Toronto[24].

Table 2: Interpretation of response rates at given course size ranges

Margin of error interval	Interpretation	Course size				
		1-25	26-50	51-100	101-200	200+
<±0.1	Very precise estimate	>90%	>80%	>80%	>60%	>50%
<±0.2	Precise estimate	>80%	>70%	>70%	>50%	>40%
<±0.5	Somewhat precise estimate	>70%	>50%	>40%	>20%	>10%
<±1.0	General estimate	>60%	>20%	>10%	>10%	>10%
<1.0	Very general estimate	<30%	<10%	<5%	<3%	<1%

2.3   *Address what statistics reveal and their limitations.*

Teaching is more likely to improve on a continuous basis when there is a sense of trust in how it is evaluated and documented. Statistics experts argue that the responses obtained from surveys (Likert-type items) are not numeric in nature, rendering arithmetic averages and comparisons based on these statistics misleading[25,26,27]. If arithmetic averages are useful, as some argue[28], factors such as course level, course size and discipline all have effects on results[12, 33], with any comparisons across these contextual factors likely containing biases.

Instead of including mean scores on all question items (the current practice at LUMS), it is considered more reliable and valid to report only the composite mean of the scores on the 5 institutional questions[24] with frequency distributions of student ratings using bar charts on all other questions. Comparisons can be provided for deeper interpretation of results but only with similar courses (same type, level, size and discipline). In short, this fundamental issue with respect to comparisons needs to be addressed and appropriate changes made to policies and guidelines on how and when course evaluation results are meaningful.

Recommendation 3: Strengthen support for teaching development and effective pedagogical practices to help faculty operationalize the data collected.

Course evaluations are problematic primarily in the way they are used at universities. Without appropriate systems in place, they typically remain privately interpreted, summative judgments of teaching that drive a deficit discourse[51,29,30]. Critical discourse of course evaluations among peers has been identified as a key process for promoting quality enhancement and pedagogical learning[8, 9,51,55,56]. Students are also critical stakeholders in the quality of teaching and learning. How student input is integrated more productively into evaluation processes receives increasing attention among higher education researchers[14,31,32,33]. This approach is validated by case studies of universities where students participate in making sense of course evaluation surveys with instructors and program directors[34,35,36].

3.1 Establish feedback mechanisms such that students see tangible results of their input.

Combining end-of-semester evaluations with mid-course evaluations can significantly enhance the utility of course evaluation ratings[37]. For example, some faculty at LUMS already collect anonymous feedback during the semester—what is helping or hindering student learning—to enable problems to be addressed during a course and to help build a relationship with students. This practice is especially helpful when teaching a course for the first time or experimenting with modified or innovative practices and should be actively encouraged.

Another exemplary practice at Lund University involves post-evaluation meetings and formative course evaluation reports on every course. The course instructor, program coordinator/director and two student volunteers from the course meet (usually over lunch) to discuss the statistical results including any written comments (with any disparaging comments removed) as well as data collected through formative measures. A brief report is written from all three perspectives on possible improvements to be made to the course. These reports (that also include summary statistics but no written comments) are emailed to all students registered in the course and uploaded to the Faculty intranet. This practice not only builds trust in the course evaluation system but also helps to incentivize student response rates.

3.2 Formalize support mechanisms that nurture reflective practice through critical inquiry, coaching and dialogue.

Informal mechanisms for improving teaching already operate in parallel to course evaluations. For example, in some Schools new faculty are matched appropriately with experienced faculty for development purposes. In other Schools, faculty are encouraged to observe each other's classes as part of milestone reviews and to include a brief peer review report in their teaching dossier. Some administrators also have an open-door policy or regular sessions with the student body to seek feedback on teaching. This is a good start to a diverse feedback loop.

Formal support for teaching has begun with piloting several modules of a teaching certificate program including the implementation of the 3-day Instructional Skills Workshop, facilitator training and the blended teaching experimentation course (see Dr. Launa Gauthier's report). In addition, the new Learning Institute will provide a range of additional supports for collecting formative feedback on teaching. An inquiry-based training program modeled after Lund University would further help to improve student learning by engaging faculty in practice-based research on their own courses and teaching. Scholarly reports produced can be shared with colleagues and accumulate into evidence of continuous development as a teacher—one of Lund's criteria for excellence in teaching.

3.3 Link enhancement initiatives to the new Student Partnership Initiative.

Partnering students with faculty engaged in teaching enhancement initiatives, for example, in co-developing course outlines and seeking feedback throughout a course, would provide further opportunities for students to be engaged in discussions on quality. Some faculty in SOE and HSS are already doing this, which should be encouraged. There is growing global interest in partnerships with students, particularly when related to pedagogical inquiry[13,14,48]. Furthermore, when student partners are paid, they can include the work experience on their curriculum vitae and the institution allocate resources towards an initiative that is genuinely transformative for both students and faculty[38].

Recommendation 4: Monitor and analyze course evaluation data for a range of institutional and School level enhancement activities.

For course evaluations to be functional in quality enhancement, faculty and administrators need formal and informal support in interpreting, critically analyzing, processing and engaging in productive, critical discourse surrounding course evaluations[3,13,39,40,41,42,43]. Different types of systems have been created to address these issues using academic epistemologies and scholarship around teaching and learning to guide this discourse[9,17,44,45,46,47,48].

There is a perception among administrators that student ratings at LUMS are skewed towards the higher levels with most faculty rated 4.5-4.6. Analysis of cumulative data, however, shows faculty can be differentiated on the category of "instructor" items across all five Schools (Appendix D). Contextual factors (academic discipline, course level and class size) affect ratings but these effects are very small. Faculty tend to fall into 3 groups: one third (sometimes less) are outstanding teachers, two thirds have room for improvement (sometimes a lot) and a small percentage require significant improvement. This is an example of how analysis of cumulative course evaluation data can (and should) inform enhancement efforts.

The new LUMS-specific course evaluation tool will need to be similarly validated for robustness. Response rates after different interventions will need to be monitored, different ways to manage bias explored, trends over time identified for individual faculty members and programs, trends in student comments analyzed within and across semesters as part of multifactor evaluations, and impact determined of student feedback and faculty reflections on course and program design and delivery. The LUMS community needs to learn more about how course evaluations are used in different ways and not only for tenure decisions. A vital step in this direction is to involve faculty, students and administrators in ongoing research. It is important for the institution to exhibit scholarly interest in enhancing learning and teaching at every opportunity.

Recommendation 5. Develop new policy for achieving tenure on the basis of teaching excellence as well as guidelines for annual and milestone reviews

Course evaluations are problematic as the only measure of teaching quality. They can be influenced by many irrelevant factors, for example, the instructor's age[49,50], gender[51,52,53], ethnicity[21,54], grading leniency[3], attractiveness[55], body language and expressiveness[56,57], and whether they speak with an accent[58]. Student factors similarly influence them, such as grade expectations[59], fatigue[11], and their attitudes on the usefulness of course evaluations[60]. Contextual factors also influence study ratings, such as the timing of administration[61], academic discipline[62,63], course type[64], course size[65], and the physical conditions of the classroom[66]. Moreover, in their typical format, course evaluations are not well related (even negatively related) to indicators of quality teaching such as student performance on standardized exams (multi-section, common exams) or in subsequent 'post-requisite' courses[12, 22,67,68,69,70].

The University of Toronto sets out three principles to guide effective interpretation of course evaluation data[24]: (1) course evaluations are key but only provide a partial portrait so any assessment of teaching should make use of multiple sources of evidence; (2) the institutional composite mean (described in 2.3) is a reasonably valid and reliable indicator of student experiences in a course but not sufficient to support comparisons and rankings; and (3) conclusions from course evaluation data should be drawn only from clear trends and patterns after considering all available data, and in consideration of context (e.g., course type, size).

In 2011, the University of Sydney established a weighting system for providing financial rewards based on contributions to the scholarship of teaching that were thought to be particularly effective at supporting the enhancement of teaching and learning. The index represented scholarly activities that delivered the greatest benefit to the University. For example, teaching awards (5 points), grant applications (5 points), qualification in university teaching (10 points), foundations of research supervision (2 points), and formal mentoring of teaching (2 points). The program shows how teaching can be allocated a significant weighted element of institutionalized tenure and promotion criteria.

An alternative approach to assessing teaching excellence at Lund University is also worth considering[5]. Faculty apply for a reward for excellence in teaching by submitting a teaching dossier that is rigorously peer-reviewed based on three criteria: (1) a learning-centred approach to teaching; (2) signs of continuous development as a teacher, and (3) evidence of being engaged with colleagues about discipline-specific educational matters. Successful applicants receive the official job title of *Excellent Teaching Practitioner* (ETP) as well as a permanent raise in salary. The applicant's department receives an equivalent amount in funding ensuring shared responsibility in the pursuit of excellence. To date, over 100 instructors (out of a total of 800) have the title of ETP.

The teaching dossier has proven to be useful in documenting different types of evidence and reliable as an alternative approach to the evaluation of effective teaching. Dossiers are also an excellent tool for developing and improving teaching through a process of documenting goals and achievements and reflecting on teaching activities and accomplishments. Dossiers should include a teaching philosophy statement with examples of how instructors are making changes in teaching methods or experimenting with technology or discipline-specific strategies or applying theory. Faculty also need to think about how learners are different, how they accommodate learners with different needs, how they add to students' learning experiences.

Connecting promotion and tenure to good teaching was a recurring theme during my interviews. Existing guidelines for promotion and tenure (APT Policy 7.5) and the criteria for a teaching award (Faculty Handbook, p. 128) will need to be revised for this to become a reality. Guidelines for preparing the teaching dossier (APT Policy 6.5) must also be revised to move beyond using course evaluation scores as the only measure of teaching quality to requiring evidence-based dossiers that track educational interventions and their impact.

Conclusion

Teaching is a complex, multi-faceted and dynamic activity. Its outcomes are notoriously difficult to quantify and equally open to bias and misinterpretation. Developing a systems approach to course evaluation is a challenge. However, taking on that challenge is absolutely worth the effort. Universities who have implemented changes recommended in this report offer a way forward and provide LUMS a framework to further differentiate its legacy of an excellent learning organization. I am confident that the above framework and the suggested operational parameters will contribute to the development of a campus culture that looks forward to feedback, and the creation of a system that adapts itself

effectively to the various individuals, groups and programs involved in providing LUMS' students with a quality learning experience. I thank the Vice-Chancellor for giving me an opportunity to contribute.

References

[1] Cassidy, R. A preliminary report on student evaluations of teaching in higher education: Principles and exemplars from the research literature. Concordia University: Internal Report. (2018). [Used with permission.]

[2] Morgan, D. A., Sneed, J. & Swinney, L. Are student evaluations a valid measure of teaching effectiveness: Perceptions of accounting faculty members and administrators. *Manag. Res. News* **26**, 17–32 (2003).

[3] Stroebe, W. Why good teaching evaluations may reward bad teaching: On grade inflation and other unintended consequences of student evaluations. *Perspect. Psychol. Sci.* **11**, 800–816 (2016).

[4] Darwin, S. What contemporary work are student ratings actually doing in higher education? *Stud. Educ. Eval.* **54**, 13–21 (2017).

[5] Roxå, T., Ahmad, A., Barrington, J., Van Maaren, J. and Cassidy, R. (in preparation). Reconceptualising student ratings of teaching towards a coherent quality enhancement discourse: A systems approach.

[6] University of Toronto Policy on the Student Evaluation of Teaching in Courses (2011). Retrieved from the UofT website: https://www.utm.utoronto.ca/dean/sites/files/dean/public/shared/pdfs/Policy%20on%20the%20Student%20Evaluation%20of%20Teaching%20in%20Courses.pdf

[7] Olsson, T. & Roxå, T. A model promoting conceptual change in higher education: An integrated approach. in *Research and Development in Higher Education* **35**, 213–223 (HERD, 2012).

[8] Mårtensson, K., Roxå, T. & Olsson, T. Developing a quality culture through the Scholarship of Teaching and Learning. *High. Educ. Res. Dev.* **30**, 51–62 (2011).

[9] Edström, K. Doing course evaluation as if learning matters most. *High. Educ. Res. Dev.* **27**, 95–106 (2008).

[10] Spooren, P., Brockx, B. & Mortelmans, D. On the validity of student evaluation of teaching: The state of the art. *Rev. Educ. Res.* **83**, 598–642 (2013).

[11] Uttl, B., White, C. A. & Gonzalez, D. A. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Stud. Educ. Eval.* **54**, 22–42 (2017).x

[12] Blair, E. & Noel, K. V. Improving higher education practice through student evaluation systems: Is the student voice being heard? *Assess. Eval. High. Educ.* **39**, 879–894 (2014).

[13] Cook-Sather, A., Bovill, C. & Felten, P. *Engaging students as partners in learning and teaching: A guide for faculty*. (Jossey Bass, 2014).

[14] Cook-Sather, A. Catalyzing multiple forms of engagement: student-staff partnerships exploring teaching and learning. In *The student engagement handbook: Practice in Higher Education* 549–566 (Emerald Publishing Groups, 2013).

[15] Cook-Sather, A. Student-faculty partnership in explorations of pedagogical practice: A threshold concept in academic development. *Int. J. Acad. Dev.* **19**, 186–198 (2013).

[16] Healey, M., Flint, A. & Harrington, K. Students as partners: Reflections on a conceptual model. *Teach. Learn. Inq.* **4**, 1–13 (2016).

[17] Ahmad, A. *et al. Investigating impact in higher education*. (Higher Education Research and Development Society of Australasia, 2018).

[18] Centre for Teaching Support & Innovation. (2018). *University of Toronto's Cascaded Course Evaluation Framework: Validation Study of the Institutional Composite Mean (ICM)*. Toronto, ON: Centre for

Teaching. Support & Innovation, University of Toronto (2019). Retrieved from the UT website: https://teaching.utoronto.ca/wp-content/uploads/2018/09/Validation-Study_CTSI-September-2018.pdf

[19] Ramsden, P. "A Performance Indicator of Teaching in Higher Education: The course experience questionnaire. *Studies in Higher Education, 16*(2) (1991).

[20] Marton, F., and S. A. Booth. *Learning and Awareness*. Mahwah, NJ: Lawrence Erlbaum Associates. (1997).

[21] Ramsden, P., ed. *Learning to Teach in Higher Education*. London: Routledge. (2005).

[22] Ramsden, P., and N. J. Entwistle. "Effects of Academic Departments on Students' Approaches to Studying." *British Journal of Educational Psychology*, 51: 368–383 (1981).

[23] Gravestock, P. & Gregor-Greenleaf, E. Student Course Evaluations: Research, Models & Trends. HEQCO Report (2008). Retrieved from the HEQCO website: http://www.heqco.ca/SiteCollectionDocuments/Student%20Course%20Evaluations.pdf

[24] Centre for Teaching Support & Innovation. University of Toronto Course Evaluation Interpretation Guidelines for Academic Administrators (2018). Retrieved from UofT website: https://teaching.utoronto.ca/wp-content/uploads/2018/10/Interpretation-Guidelines_Final_Oct.1.2018.pdf

[25] Stark, P. & Freishtat, R. An evaluation of course evaluations. *Sci. Res.* (2014). doi:10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

[26] Jamieson, S. Likert scales: How to (ab)use them. *Med. Educ.* **38**, 1217–1218 (2004).

[27] Hornstein, H. A. Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Educ.* **4**, 1304016 (2017).

[28] Norman, G. Likert scales, levels of measurement and the 'laws' of statistics. *Adv. Health Sci. Educ. Theory Pract.* **15**, 625–632 (2010).

[29] Kember, D., Leung, D. Y. P. & Kwan, K. P. Does the use of student feedback questionnaires improve the overall quality of teaching? *Assess. Eval. High. Educ.* **27**, 411–425 (2002).

[30] Lutovac, S., Kaasila, R., Komulainen, J. & Maikkola, M. University lecturers' emotional responses to and coping with student feedback: A Finnish case study. *Eur. J. Psychol. Educ.* **32**, 235–250 (2017).

[31] Cook-Sather, A. Student-faculty partnership in explorations of pedagogical practice: A threshold concept in academic development. *Int. J. Acad. Dev.* **19**, 186–198 (2013).

[32] Healey, M., Flint, A. & Harrington, K. Students as partners: Reflections on a conceptual model. *Teach. Learn. Inq.* **4**, 1–13 (2016).

[33] Ahmad, A. *et al. Investigating impact in higher education*. (Higher Education Research and Development Society of Australasia, 2018).

[34] Roxå, T. & Mårtensson, K. Developing higher education organisations and evaluating effects. in *Teaching skills assessments - Qualitätsmanagement und Personalentwicklung in der Hochschullehre* (eds. Egger, R. & Merkt, M.) **17**, 91–115 (Springer, 2016).

[35] Kiger, J. Evaluating evaluations: An ouroboros for medical education. *Med. Educ.* **51**, 131–133 (2017).

[36] Lubicz-Nawrocka, T. M. Students as partners in learning and teaching: The benefits of co-creation of the curriculum. *Int. J. Stud. Partn.* **2**, 47–63 (2018).

[37] Wilson, J. H., & Ryan, R. G. Formative teaching evaluations: Is student input useful? In M. E. Kite (Ed.), Effective Evaluation of Teaching: A Guide for Faculty and Administrators. 22-29 (2012). Retrieved from the Society for the Teaching of Psychology website: http://teachpsych.org/ebooks/evals2012/index.php

[39] Roxå, T. & Bergström, M. Kursvärderingar i system – akademiska lärares upplevelser och organisationens förmåga till utveckling [Course evaluation systems: Academic teachers' experiences and the institution's capacity for development]. *Högre Utbild. High. Educ.* **3**, 225–236 (2013).

40 Mårtensson, K., Roxå, T. & Stensaker, B. From quality assurance to quality practices: An investigation of strong microcultures in teaching and learning. *Stud. High. Educ.* **39**, 534–545 (2014).

41 Yao, Y. & Grady, M. L. How do faculty make formative use of student evaluation feedback? A multiple case study. *J. Pers. Eval. Educ.* **18**, 107–126 (2005).

42 Arthur, L. From performativity to professionalism: Lecturers' responses to student feedback. *Teach. High. Educ.* **14**, 441–454 (2009).

43 Shakurnia, A. H., Mozaffari, A. R. & Karami, M. A. Comparison of teachers' attitude with high and low scores evaluation on the importance and use of student assessments. *Educ. Res. Med. Sci.* **1**, 38–46 (2012).

44 Freeman, R. & Dobbins, K. Are we serious about enhancing courses? Using the principles of assessment for learning to enhance course evaluation. *Assess. Eval. High. Educ.* **38**, 142–151 (2013).

45 Payne, S. L., Whitfield, J. M. & Flynn, J. A. Assessing the business capstone course through a method based on the SOTL and the stakeholder process. *J. Educ. Bus.* **78**, 69–74 (2002).

46 Kember, D. & Ginns, P. *Evaluating teaching and learning*. (Routledge, 2012).

47 Walker, J. D., Baepler, P. & Cohen, B. The scholarship of teaching and learning paradox: Results without rewards. *Coll. Teach.* **56**, 183–190 (2008).

48 Hamilton, D. Towards a scholarship of teaching, learning and leadership: A case study on the use of course evaluation data to promote collaborative faculty engagement. *Transform. Dialogues* **2**, 1–20 (2008).

49 Arbuckle, J. & Williams, B. D. Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles* **49**, 507–516 (2003).

50 Bianchini, S., Lissoni, F. & Pezzoni, M. Instructor characteristics and students' evaluation of teaching effectiveness: Evidence from an Italian engineering school. *Eur. J. Eng. Educ.* **38**, 38–57 (2013).

51 Basow, S., Codos, S. & Martin, J. The effects of professors' race and gender on student evaluations and performance. *Coll. Stud. J.* **47**, 352–363 (2013).

52 MacNell, L., Driscoll, A. & Hunt, A. N. What's in a name: Exposing gender bias in student ratings of teaching. *Innov. High. Educ.* **40**, 291–303 (2015).

53 Boring, A. Gender biases in student evaluations of teaching. *J. Public Econ.* **145**, 27–41 (2017).

54 Huston, T. A. Race and gender bias in higher education: Could faculty course evaluations impede further progress toward parity. *Seattle J. Soc. Justice* **4**, 591–611 (2006).

55 Wolbring, T. & Riordan, P. How beauty works: Theoretical mechanisms and two empirical applications on students' evaluation of teaching. *Soc. Sci. Res.* **57**, 253–272 (2016).

56 Ambady, N. & Rosenthal, R. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *J. Pers. Soc. Psychol.* **64**, 431–441 (1993).

57 Williams, W. M. & Ceci, S. J. "How'm I doing?" Problems with student ratings of instructors and courses. *Change Mag. High. Learn.* **29**, 12–23 (1997).

58 Subtirelu, N. C. "She does have an accent but…": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. *Lang. Soc.* **44**, 35–62 (2015).

59 Worthington, A. C. The impact of student perceptions and characteristics on teaching evaluations: A case study in finance education. *Assess. Eval. High. Educ.* **27**, 49–64 (2002).

60 Spooren, P. & Christiaens, W. I liked your course because I believe in (the power of) student evaluations of teaching (SET): Students' perceptions of a teaching evaluation process and their relationship with SET scores. *Stud. Educ. Eval.* **54**, 43–49 (2017).

61 Grimes, A., Medway, D., Foos, A. & Goatman, A. Impact bias in student evaluations of higher education. *Stud. High. Educ.* **42**, 945–962 (2017).

62 Bini, M. & Masserini, L. Students' satisfaction and teaching efficiency of university offer. *Soc. Indic. Res.* **129**, 847–862 (2016).

[63] Uttl, B. & Smibert, D. Student evaluations of teaching: Teaching quantitative courses can be hazardous to one's career. *PeerJ* **5**, e3299 (2017).

[64] Alhija, F. N.-A. Teaching in higher education: Good teaching through students' lens. *Stud. Educ. Eval.* **54**, 4–12 (2017).

[65] Bedard, K. & Kuhn, P. Where class size really matters: Class size and student ratings of instructor effectiveness. *Econ. Educ. Rev.* **27**, 253–265 (2008).

[66] Feldman, K. A. Class size and college students' evaluations of teachers and courses: A closer look. *Res. High. Educ.* **21**, 45–116 (1984).

[67] Boring, A., Ottoboni, K. & Stark, P. Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Sci. Res.* (2016). doi:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

[68] Braga, M., Paccagnella, M. & Pellizzari, M. Evaluating students' evaluations of professors. *Econ. Educ. Rev.* **41**, 71–88 (2014).

[69] Carrell, S. E. & West, J. E. Does professor quality matter? Evidence from random assignment of students to professors. *J. Polit. Econ.* **118**, 409–432 (2010).

[70] Hassan, S. & Wium, W. Quality lies in the eyes of the beholder: A mismatch between student evaluation and peer observation of teaching. *Afr. Educ. Rev.* **11**, 491–511 (2014).

Appendix A: Highlights from Interviews at LUMS with Administrators, Faculty and Students (May, 2019)

- The LUMS CE system is limited.
  - Students have no faith because they see no change in poor instructors.
  - Managers perceive CE results as skewed—most faculty get high ratings (over 4 mostly 4.5-4.6) and we know this is unlikely.
  - Faculty find it challenging to use CE feedback to improve teaching because the questionnaire focuses on what the teacher did not do right and their subject knowledge rather than how they helped students with problems and anxiety—the human aspect.

- The biggest issue is incentivizing students to respond in an online system.
  - The current practice of penalizing students for not responding by not allowing them to access their grades means they do it hurriedly or out of fear.
  - Students do not recognize the importance of CE; they need assurance that the CE number is not a random number, it is the pulse of the course and helps to improve courses based on feedback about what challenges learning.
  - Students also need to know who sees what to alleviate their concern over anonymity.

- Informal mechanisms for improving teaching operate in parallel to CE.
  - Students can be very candid when they are really unhappy.
  - Managers have an open door policy or regular sessions with the student body to seek informal feedback on teaching
  - When the word spreads that an instructor needs improvement there is follow-up— and news travels fast.
  - Some department chairs attend classes of new faculty who have no prior teaching experience.
  - Some schools arrange for mentoring/peer review of teaching with new faculty matched appropriately with experienced faculty—needs to be seen as development not evaluation.
  - Some faculty collect anonymous mid-term feedback—what students are happy about and unhappy about—this enables problems to be addressed during a course and helps to build a relationship with students.

- Promotion and tenure needs to be connected to good teaching.
  - A culture of caring about teaching needs to be communicated—effective methods, evidence-based, trajectory of improvement.
  - Historically, LUMS recognized the fight was in the trenches—you had to work at teaching, add value in class and students were compelled to come to class—as the university expanded, commitment to teaching and student development slowly eroded towards a campus culture that is more research focused; teaching is still important but not on an equal pedestal to research.
  - Faculty need to know teaching is as valuable as research.
  - Teaching needs to be a significant weighted element of institutionalized tenure and promotion criteria.

- Changes in student body impact good teaching.
  - An explosion in student enrolment at the same time as cuts in course sections and electives has resulted in a significant increase in class size.
  - Students are increasingly fixated on grades, on how grading works, resulting in lots of grading discussions between faculty and students; they are more concerned with getting a degree than engaging with learning challenges.
  - Socio-economic diversity among students is problematic—15-20% of scholarships go to students from the villages who have language and cultural barriers; some faculty spend a lot of time in pastoral care.

- Recognition for teaching excellence is important.
  - Different kinds of incentives are needed in support of teaching excellence, e.g., prestigious awards, substantial rewards, and school level endowment for distinguished teachers.
  - Great teaching needs to be rewarded the way research is rewarded—with upward mobility.

- Opportunities are needed for knowledge of teaching to evolve.
  - Faculty need to learn about innovative classroom environments, to come together, to talk and share ideas.
  - Training needs to be sensitive to different kinds of pedagogical style in the Humanities, for example, compared to Science and Engineering or Business.
  - The Post Graduate Certificate in Higher Education (PGCHE) in the UK is an online pedagogical model that may be worth emulating. https://www.nottingham.ac.uk/education/study/pgche/index.aspx
  - There needs to be formal orientation on teaching for new faculty and teaching assistants.

- Suggested criteria for the evaluation of teaching:
    - Pay attention to CE results but also to how students evaluate over the years as well as their written comments, especially if scores are low
    - Teaching core courses well (very different than teaching electives and we are very short of faculty serving core courses)
    - Mentoring students after class (service used to be a criteria, some faculty spend a lot of time on pastoral care, need both sides of learning with 60+ students frequently go beyond the call of duty)
    - Generosity of time, graduate school counselling, Fulbright applications; if students get a good placement abroad, e.g., MIT, then letters can be provided
    - Student supervision (30 projects per year is like teaching one course) and distribution of who students go to is not even, helps faculty stand out
    - New course development or minimum 20% of course changed or upgraded
    - Innovation in teaching methods, e.g., developing simulations, writing case studies, entrepreneurial inspiration, faculty looking at what other faculty are doing across campus, co-teaching (we are presently in disciplinary silos not the real world, we need more interdisciplinary courses, joint approach, bridges/spaces)
    - Research on pedagogy, e.g., the impact of teaching on student learning, setting up new courses collaboratively and then reflecting upon the impact on content and teaching, as well as published case studies
    - Self-reflection, e.g., a teaching statement that includes examples of making changes in teaching methods or experimenting with technology or discipline-specific strategies or applying theory; teachers need to think about how learners are different, how they cater to learners with different needs, how they add to students' learning experiences (teaching philosophy is not talked about at LUMS, people are just interested in whether a course is very popular based on having a waiting list)
    - Not letters from colleagues but 5 people from abroad (selected from a pool)
    - Not letters from current students but 3-4 from graduating students (random selection)
    - Caring about teaching – effective methods, evidence-based, trajectory of improvement; presently more focus on research than learning

- Students' dream of having more input on teaching:
    - More networking between faculty and students, e.g., lunches, sports activities, out of class activities and meeting with new faculty.
    - Include student voice when developing course outlines, e.g., students who have taken the course before, develop course based on incoming student diversity / demographics otherwise half the class may not have prior knowledge to succeed.
    - Seek student feedback throughout the course, e.g., TA solicits feedback not simply attendance (problem is they are not paid enough to do more work).
    - Students would interact more with TAs if they are undergraduate students
    - Presently peers influence what courses to take via social media, how can this be formalized.

List of people interviewed

Dr. Arshad Ahmad, Vice-Chancellor
Dr. Adnan Khan, Dean of Students
Dr. Syed Mubashir Ali, Registrar

SASOL
Dr. Zubair Abbasi
Dr. Uzair Kayani

SDSB
Dr. Alnoor Bhimani, Dean
Dr. Muhammad Naiman Jalil, Associate Dean
Dr. Bushra Naqvi, Associate Dean
Dr. Muhammad Adeel Zaffar, Associate Dean
Dr. Asad Alam, EMBA

SHSS
Dr. Kamran Asdar Ali, Dean and Acting Dean of Law
Dr. Aamna Khalid
Dr. Adnan Zahid
Dr. Ali Khan
Dr. Yasser Hashmi
Dr. Maryam Wasif Khan

SOE
Dr. Tahir Raza Shah Andrabi, Dean Dr. Faisal Bari
Dr. Tayyaba Tamin
Ayesha Shahid

SSE
Dr. Shahid Masud, Dean
Dr. Tariq Mahmood Jadoon, Chair EE, and President Faculty Council
Dr. Ihsan Ayyub Qazi, Chair CS
Dr. Ammar Ahmed Khan

Faisal Kheiri, Head IST and team
Hiba Zakai, Campus Climate Manager

Representatives from the Student Council

Hammad ullah Joyia
Wardah Noor
Noor Fatima
Muaaz Ahmed Noor
Amman Nadeem
Momina Rasool
Maida Tahir
Talha Ejaz
Hattaf Ayub

Appendix B: Course Evaluation Question Items

<u>Student choices</u>: Not at all ->> Somewhat ->> Moderately ->> Mostly ->> A great deal

Institutional Questions[1]
1. I found the course intellectually stimulating (engagement)
2. The course provided me with a deeper understanding of the subject matter (knowledge)
3. The instructor created a course atmosphere that was conducive to my learning (learning environment)
4. Course projects, assignments, tests, and/or exams allowed me to demonstrate my understanding of the course material (understanding)
5. Overall, the quality of my learning in this course improved

<u>Note 1</u>: These questions are common across Schools, Departments and Programs.

School Questions[2]
6. The course inspired me to think further about the subject matter outside of class (transfer)
7. Compared to other courses in the School, the workload for this course was… (workload)
8. I would recommend this course to other students (satisfaction)

Department/Program Questions[2]
9. The evaluation methods used in this course were fair (assessment)
10. The course provided opportunities to receive feedback (feedback)
11. The course expectations were clearly communicated (objectives)

<u>Note 2</u>: School and Department questions can vary across LUMS. They should focus on learning instead of instructor questions (covered below)

Individual Questions[3]
12. The instructor motivated me to do my best work (motivation)
13. The instructor made the subject interesting (useful)
14. The instructor clear explanations (clarity)
15. The instructor related to students in ways that promoted mutual respect (respect)

<u>Note 3</u>: The responses to these individual questions are only seen by the instructor. Instructors may choose to replace these questions with other questions they'd like feedback on. These questions also present opportunities to get additional qualitative feedback, (for eg. "How did the instructor help or hinder your learning?" "What aspects of the course were most valuable to your learning?", etc.)

Open Ended Questions
1. Please provide feedback on how this course can be improved
2. Any other comments

Appendix C: Course Experience Questionnaire

Ramsden, P. (1991). "A Performance Indicator of Teaching in Higher Education: The course experience questionnaire. *Studies in Higher Education, 16*(2).

Wilson, K.L, Lizzio, A. & Ramsden, P. (1997) The development, validation and application of the Course Experience Questionnaire, *Studies in Higher Education, 22*:1, 33-53.

CEQ Scales

Source: http://www2.rmit.edu.au/departments/planning/ircu/docs/ceqexp.pdf

Appendix D: LUMS Cumulative CE Data Analysis
(R. Cassidy and J. Barrington)

Can instructors be differentiated on their scores on the 'instructor' items[1]?

The plots in Figure 1 show the 95% confidence interval (red line) for the 'instructor' scores for each of 499 instructors. Any two instructors with vertically non-overlapping red lines can be statistically differentiated.

In all five Schools (SASOL, SDSB, SHSS, SOE, SSE), the 'instructor' items on the LUMS CE questionnaire were able to differentiate instructors.

Figure 1: Analysis of instructor scores for academic years 2016-17 & 2017-18
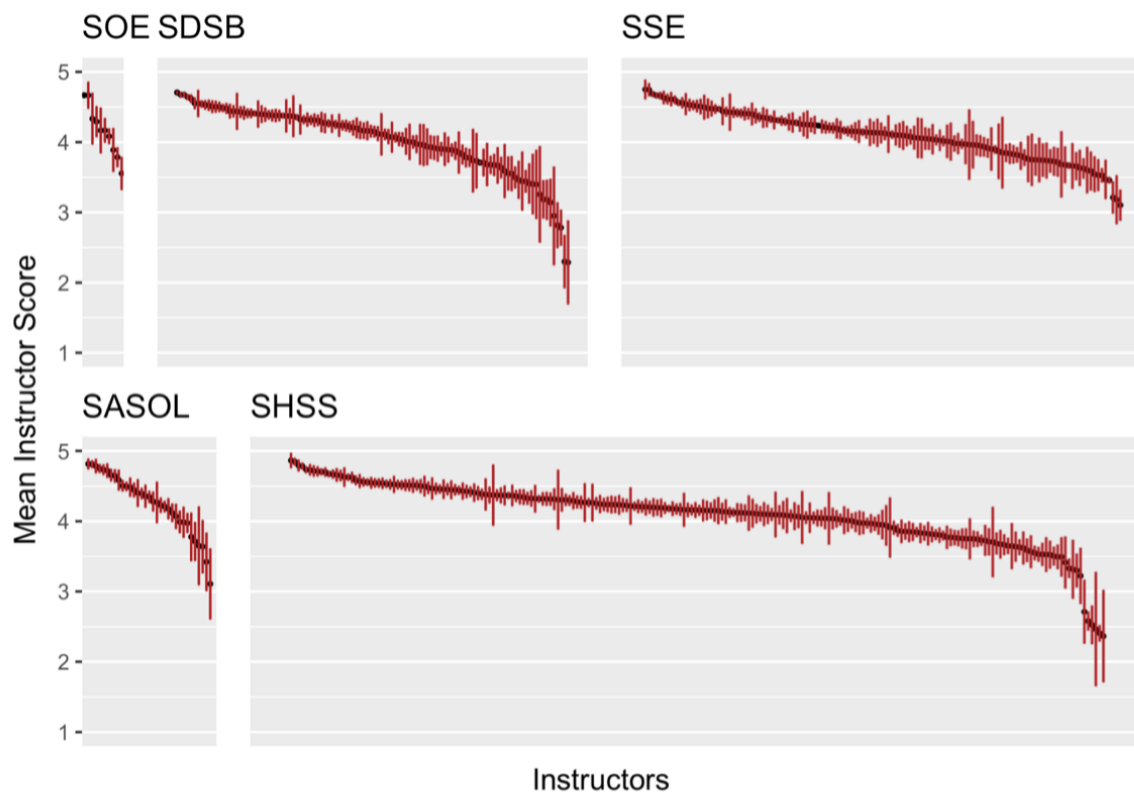
[1]Instructor items: 1) The instructor was well prepared for class meetings. 2) I would recommend this instructor to other students interested in this subject. 3) I would take another course with this instructor. 4) The instructor was fair and impartial with me. 5) In my opinion, the instructor was punctual and adhered to scheduled class hours. 6) The instructor is available for feedback or assistance during scheduled office hours.

Do contextual variables (academic discipline, course level, class size) influence 'instructor' scale ratings?

Table 1 shows the number of instructors per School, the % of variance explained by the instructor (INSTR) versus contextual factors (CONTX).

Also included are the % of 'instructor' scores in each range of the CE questionnaire Likert scale.

Contextual factors did affect ratings on the 'instructor' items but the effects were very small.

Table 1: Comparison of Schools for academic years 2016-17 & 2017-18

School	N=	INSTR	CONTX	4.5-5	4-4.5	3.5-4	3-3.5	2.5-3	2-2.5
SASOL	33	43%	2%	30.3%	42.4%	21.2%	6.1%		
SDSB	112	42%	2%	10.7%	50%	25%	9.8%	2.7%	1.8%
SHSS	214	40%	2%	15.9%	53.3%	24.8%	3.3%	1.4%	1.4%
SOE	10	29%	0%	20%	50%	30%			
SSE	130	29%	4%	12.3%	52.3%	31.5%	3.8%		