



# Reconceptualizing student ratings of teaching to support quality discourse on student learning: a systems perspective

Torgny Roxå<sup>1</sup> · Arshad Ahmad<sup>2</sup> · Janette Barrington<sup>3</sup> · John Van Maaren<sup>4</sup> · Robert Cassidy<sup>5</sup> 

© Springer Nature B.V. 2021

## Abstract

This paper summarizes the discourse on student ratings of teaching in higher education. It reconceptualizes student ratings within a larger process of promoting quality in teaching and student learning. As students engage in productive dialogue with teachers and administrators, metrics drawn from decontextualized surveys are admittedly a vital resource. Our paper contends, however, that student ratings can only become a tool for enhancement when they feed reflective conversations about improving the learning process and when these conversations are informed by the scholarship of teaching and learning. We illustrate this view with a case study of an Engineering Faculty that uses three interconnected initiatives to support an evolving conversation on quality among teachers in partnership with students. The role of student engagement in enhancement efforts is discussed as well as potential challenges to implementation. Our purpose is to spark wider discussion of a systems perspective on student ratings that supports a coherent discourse towards positive change.

**Keywords** Course evaluations · Higher education · Quality enhancement · Practice-based research · Approaches to learning · Students as partners · Systems perspective

## Introduction

Student ratings of teaching are often considered fundamental for measuring the quality of teaching in higher education, for educational development and for the enhancement of student learning. Different stakeholder interests—national, institutional, teacher and

---

✉ Robert Cassidy  
rob.cassidy@concordia.ca

<sup>1</sup> Faculty of Engineering, Lund University, Lund, Sweden

<sup>2</sup> Lahore University of Management Sciences, Lahore, Pakistan

<sup>3</sup> Centre for Teaching and Learning, Concordia University, Montreal, Canada

<sup>4</sup> MacPherson Institute, Concordia University, Hamilton, Canada

<sup>5</sup> Centre for Teaching and Learning, Department of Education, Concordia University, De Maisonneuve Blvd. W, Montreal H3G 1M8, Canada

student—create tensions for its multiple purposes. In national and international discourses where higher education is increasingly seen as a vital pathway for human development, student ratings are widely used as an instrument to measure the quality of teaching, teachers, courses and programs and to influence rankings and funding to institutions.

One recent example is the teaching excellence framework (TEF) in the UK, a government-initiated regulatory framework for the higher education sector administered independently through the Department for Education. The primary goal of the TEF is to create a competitive market by providing improved information on teaching. Another more altruistic goal is to promote equal opportunities for students and improved learning outcomes regardless of a student's entry qualifications, characteristics and subject of study. The TEF reminds all stakeholders that 'variability is not simply a statistic', because underperformance of an institution can represent 'thousands of life opportunities wasted, of young dreams unfulfilled, all because of teaching that was not as good as it should have been, or because students were recruited who were not capable of benefitting from higher education' (Department for Business, Innovation, and Skills 2016: 46). As one metric, the TEF uses the independently administered National Student Survey to gather feedback from students registered in participating colleges and universities.

Clearly, the TEF uses student ratings to measure quality in teaching and learning at the institutional level. It also pursues and integrates many other complementary metrics. Yet, student ratings for summative purposes are largely ineffective as a quality-enhancing role. Indeed, in the case of the TEF, it can prompt institutional managers to respond in ways that increase revenue, with some reporting changes to their admission routines rather than changes in their teaching (Copeland, 2017). Thus, while allowing for student ratings to be used for summative purposes, we also need to recognize that this will not necessarily lead to the enhancement of teaching and learning.

Dissatisfaction with the quality assurance role of student ratings, especially given the biases in instruments used for measuring quality (discussed below), has served to alienate teachers and to weaken the student voice and has prompted some scholars to consider abandoning them entirely (Uttl et al. 2017). It is this tension between measurement and enhancement that we address by considering how student ratings can more effectively enhance the quality of teaching and learning.

We begin with a review of the literature evaluating whether student ratings can be trusted for measuring teaching quality and then discuss complementary literature supporting the usefulness of student ratings. The second half of the article provides a case study of how a research-intensive institution achieved positive change in teaching by using course evaluations and other enhancement initiatives to fuel an evolving conversation on teaching quality among teachers and institutional managers in partnership with students.

## Can student ratings of teaching be trusted?

Nearly 40 years ago, Marsh (1980) analyzed responses from 511 undergraduate courses taught by 221 instructors in search of systematic bias in how students rated their teachers. The analysis included twenty potential biases and concluded that: 'a variety of variables that might have affected student evaluations of teaching effectiveness do not seem to make much difference' (p. 236). Marsh and others examining the validity of student ratings conclude that they are valid and reliable mechanisms to measure teaching quality (Abrami et al. 1990; Kulik 2001; Marsh and Cheng 2008; Marsh and Roche 1999). In other words, decades of research have laid a foundation of confidence in student ratings.

However, long before Marsh's original study and even predating the formal collection of student ratings data, Borg and Hamilton (1956) described a different reality. In a study at Lackland Air Force Base, 89 instructors rated themselves and then were rated by their students (60 per instructor), their peers and their supervisors on six elaborated criteria for good teaching. None of the four sets of ratings correlated with outcomes on a performance test. Student ratings in this study appeared to be unrelated to the intended outcome of improved teaching and learning.

The debate continues along similar lines today. A second-order meta-analysis by (Uttl et al. 2017) re-analysed previously published meta-analyses on the correlation between student ratings and the quality of student learning. They found systematic biases in methods and analyses and concluded that student ratings and student learning are unrelated. They further suggest 'universities and colleges focused on student learning may need to give minimal or no weight to [student] ratings' (p. 40). Numerous studies support Uttl et al.'s conclusion and emphasize a number of biases in student ratings.

Studies have shown that student ratings are influenced by teacher body language (Ambady and Rosenthal 1993), gender (Andersen and Miller 1997; Boring 2017; Sprague and Massoni 2005), ethnicity (Basow et al. 2013; Boatright-Horowitz and Soeung 2009; Huston 2006), leniency (Stroebe 2016) and attractiveness (Wolbring and Riordan 2016). Ratings of teaching are also influenced by student fatigue (Spooren et al. 2013), weather conditions (Braga et al. 2014), timing (Grimes et al. 2017), physical space (Kwan 1999), academic discipline (Bini and Masserini 2016), course type (Alhija 2017) and students' beliefs in their usefulness (Spooren and Christiaens 2017). One study found that student emotional response to classroom experience is twice as strong an indicator as cognitive reaction (Ang et al. 2018). In addition, the consistently lower response rate from online surveys may privilege anomalous responses and further skew the data (Morrison 2013).

Thus, student and contextual factors may have as much influence on student ratings of teaching as teaching behaviour (Bedgood and Donovan 2012). The volume of such compelling research findings is a cogent reminder that student ratings do not provide a consistent measure of good teaching, good teachers or good education. Even though Marsh and others arrived at a different conclusion, it is clear that student ratings are problematic as a measure of teaching quality.

Furthermore, problems with student ratings are not restricted to what they measure and do not measure but also to how they are used. Edström (2008) interviewed teachers in a technical university and found that groups of teachers sometimes use selected aspects of results from student ratings to confirm pre-existing and disparaging views of students. Various results were interpreted as signs of student laziness, poor work ethic or a superficial understanding of university education. Kember et al. (2002) reached a similar conclusion after evaluating a mandatory policy on student ratings of teaching in a Hong Kong university. The results show that student ratings do not improve quality, which they attribute to the ineffective use of the results and lack of incentives to improve teaching.

Many teachers also report difficulty interpreting results (Wong and Moni 2014). Stein et al. (2013) found that while most teachers reported reviewing the data (87%), looking for teaching feedback (77%) and comparing with previous data (77%), few sought help interpreting data (12%) and slightly less than half reported discussing data with colleagues (47%). While teachers often report mildly positive views of student feedback, evidence of impact on improved teaching practice is limited (Beran and Rokosh 2009; Hendry et al. 2007; Nasser and Fresko 2002; Stein et al. 2013).

The issue of teacher perception raises important questions about how higher education institutions implement and use student ratings. Roxå and Bergström (2013) found that

teachers were emotionally affected by student ratings, especially if the results were poor (Arthur 2009; Yao and Grady 2005). In the same institutional context, teachers were left on their own to process the results, regardless of whether they were positive or negative. Student ratings can likely hamper innovation in teaching since it is better to meet student expectations than to try something new that may lead to poor results (Arthur 2009; Darwin 2017).

As institutions evolve toward increasingly managerial structures (e.g. Gunn 1995; Meyer 2002), trust issues also arise that can be divisive and increasingly unproductive. The variety of uses of student ratings of teaching betray not only the complexities of the university (Darwin 2017) but also the variety of implicit notions of how the overall system works, who has authority and how change and improvement occur (Trowler and Cooper 2002). Administrators can easily understand the university through a scientific management lens (Kezar 2018) where student surveys become a blunt instrument of a managerial agenda and a 'good-enough' metric of work performance to be used for rewards and remediation (as envisioned by the TEF in the UK). Such use risks fuelling mistrust, especially in contrast to teachers who wish to use student feedback for more self-authored, professional development purposes.

The Canadian context offers a glimpse into an alternative national response to these tensions. A legal precedent (Ryerson University v. Ryerson Faculty Association, 2018) has ruled that student ratings of teaching are not to be used in decisions on tenure and promotion and further prompted the Ontario Confederation of University Faculty Associations (2019) to publish the following statement: 'To return student questionnaires to their original purpose—as sources of formative feedback—the working group recommends discontinuing their use for summative evaluation of teaching performance' (10). This raises an important question for all stakeholders: how can evaluation data considered inadequate for summative purposes be productively used for formative purposes? We return to this question in the case study below.

## Teacher perceptions of student ratings

Although the research portrays student ratings as a flawed measure of teaching quality, many teachers retain a moderately positive perception of student ratings. For example, Murray's (1997) synthesis of fourteen studies found that 73.4% of teachers found student feedback useful while 68.8% thought it improved teaching. Data from other studies confirm the moderately positive teacher perception of student ratings (e.g. Arthur 2009; Beran and Rokosh 2009; Darwin 2017; Flodén 2017; Moore and Kuol 2005; Nasser and Fresko 2002; Schmelkin et al. 1997; Stein et al. 2013; Wong and Moni 2014; Yao and Grady 2005). A small but compelling Finnish study (Lutovac et al. 2017) suggests that academic teachers with high ratings may find themselves in a positive spiral of development fuelled by these results (cf. Shakurnia et al. 2012). However, teachers also experience a negative spiral where low ratings create a sense of failure and poor confidence contributing to teaching behaviours that make matters worse (cf. Blair and Noel 2014).

It appears that while administrators and managers might put more trust in student ratings (Linse 2017; Morgan et al. 2003), some teachers and students still remain sceptical. A recent study (Hammer et al. 2018) surveyed 2241 university teachers of all ranks and from varied institutions about their positive attitudes and trust in student ratings of teaching. Findings suggest that, despite believing in negative myths (student vindictiveness, lack of maturity, and negative evaluations of low-achieving students), more than half of the

teachers trusted student ratings and found them useful. One-third mistrusted them and felt underestimated by their students, and a small subgroup (9%) felt extremely underestimated making student feedback counter-productive to their self-image.

Darwin (2017) further highlights how a teacher's rank can influence the value they assign to student ratings. Experienced and tenured teachers balance the results as *one* out of many aspects when reflecting on quality in their teaching. Early career teachers tend to assign student ratings greater value leading to speculations about their importance for future career opportunities (cf. Flodén 2017; Yao and Grady 2005). These tensions make teachers more vulnerable in times when higher education is seen as becoming part of the knowledge economy and thereby influenced by neoliberal principles of production, accountability and governance. The UK's TEF is an example of how this view has become operationalized. Teacher vulnerability is compounded as a larger proportion of academic teachers are placed under precarious working conditions in combination with institutions demanding more teaching to more diverse groups through more innovative methods (Field 2015).

Institutionalized student surveys result in decontextualized statistics that do not match the experience of teachers, leading to further alienation. The numbers ironically create uncertainty, fuelling an accountability process that alienates teachers from their disciplinary norms and ways of thinking and practice (McCune and Hounsell 2005). Student ratings also become a tempting means for managers to reach into the heart of knowledge production, student learning and teaching without sufficient contextual knowledge of either content or process.

Student ratings have been described as a driver of teaching quality and innovation. In reality, their current use often impedes development and sometimes makes matters worse. By ignoring some students' views and by some teachers ignoring results, it would appear that student ratings of teaching are not adequately capturing the student voice neither for institutional decision-making processes nor for professors' improvement efforts.

### **Should student ratings continue to play an integral role in higher education?**

In reality, student ratings of teaching are not used simply as surveys of student satisfaction; they have become institutionalized for the purpose of informing promotion and tenure decisions. To understand what this implies, Darwin (2017) discussed with a sample of teachers their experience with student ratings and the implications of policy. All respondents acknowledged the need for student input in any discussion on quality in teaching, and many expressed concerns that current policies discourage innovative teaching methods. Pedagogical innovations are often complex and can require several iterations before achieving their potential. Meanwhile, there is a risk of low student ratings leaving a trace of poor teaching behind. Innovations may also imply a break in traditional teaching and thereby create unease among students. Even if this unease is part of the pedagogical intention, it can result in lowering scores.

Darwin's respondents also expressed concern that students entering university often hold simplistic conceptions about teaching and learning that are inconsistent with those favoured by evidence-based pedagogy. For example, while at university, students often follow an epistemological development trajectory from absolute knowing to contextual knowing (Baxter Magolda 1992; Perry 1970). Provoking students in ways that lead to proficiency in dealing with knowledge uncertainty and evidence-based arguments—both necessary competencies in the twenty-first century—may violate their expectations and lead to

negative feedback (e.g. O'Donovan 2017). With greater emphasis on student ratings, especially in hierarchical institutions, it may become harder for teachers to confront student-held assumptions and expectations. Should we then abandon student ratings of teaching altogether, as some institutions are already considering (Flaherty 2018)?

Darwin's (2017) response is an emphatic 'no'. He advocates returning to the original intent of student ratings of teaching, namely, to invite student input in the process of improving student learning. 'By transforming conventional student evaluation models from individual-deficit discourses into one centred on collaborative professional dialogue, the student voice can potentially become a more valued (rather than derided) input into collective and individual forms of academic deliberation' (20). The way to do this is to let students report on their individual learning experiences. This, in combination with more diverse material, should then be subjected to collegial interpretation among teachers and students. 'Such an engagement offers the real potential to generate more substantial professional dialogue centred on improvement and innovation' (21).

Darwin (2017) further calls for reconceptualising not only student ratings but also the entire process of determining quality—from metrics founded on decontextualized surveys and key performance indicators to a critical dialogue among all stakeholders. We agree and go further. Higher education is flooded with data on various aspects of quality; what is missing is an evolved, critical discourse around what these data mean in terms of the principles underpinning much of the meaning making, interpretation, rationale and judgement with respect to quality.

For example, discipline-based professional judgement abounds in higher education, but it is often based on an information transfer ideology that represents a conflation between teaching and learning. It has been shown that the limiting factor for an evolved discussion about quality is not the biases that are built into student ratings but rather naïve conceptions on what constitutes quality and quality enhancement (Edström 2008; Kember et al. 2002). The way to counteract this reality is to include student ratings as the basis for critical discourse. More importantly, this discourse must involve those who are directly involved in the educational process: students, teachers and their closest managers.

In the second half of this paper, we bring Darwin's ideas on using student ratings as a mechanism for driving teaching development into practice and put a fresh perspective on a divided issue. The change in terminology from 'student evaluations of teaching' commonly used in the literature, to the more accurate term of 'student ratings of teaching' used by Darwin (2017) and others, to the term 'course evaluations' used in our case study is purposeful. On the one hand, the shift in language may represent a devaluation of these data as a yardstick for quality, instead of seeing them as student opinions of their course/teacher/experience. On the other hand, it represents the change in thinking and practice we advocate—from a preoccupation with measuring teaching quality towards a rejuvenated focus on improving student learning. The need for such change is urgent given the ubiquity of student ratings of teaching as an indicator for quality despite their many flaws.

## The case of Lund University

The following case study illustrates an approach to reconceptualising student ratings of teaching within a larger process of promoting quality in student learning. The case is situated in the Faculty of Engineering (LTH) at Lund University, a traditional research-intensive university in Sweden. The LTH comprises approximately 9000 undergraduate and

master's students, 600 Ph.D. students and 800 teaching staff. Two-thirds of the LTH budget comes from competitive research grants contributing to the university's recognition among the Times Higher Education Rankings top 100 institutions globally.

The case is described within its national and institutional context. Three initiatives—course evaluations, pedagogical training and a reward for excellence in teaching—are elaborated in sufficient detail to facilitate understanding their evolution over time. The role of student engagement in enhancement efforts is discussed as well as potential challenges to implementation. Each section begins with a guiding principle extracted from our review of the research literature to provide an overall framework for understanding the LTH as an illustrative case for reconceptualising student ratings of teaching from a systems perspective.

## A systems perspective

*To avoid creating unintentional cross-purposes within different enhancement systems, discussions about student ratings benefit from keeping a systems perspective in mind (Darwin 2017).*

In the late 1990s, the LTH established an educational development unit staffed by academics with scholarly interests in developing and studying a holistic approach to enhancement (Mårtensson et al. 2011; Roxå 2013; Roxå et al. 2008). A key concept of this approach involves attending to relationships between what, at first, might seem unrelated processes (Senge 2006). Similar initiatives related to course evaluation, pedagogical training and recognition for excellence in teaching operate throughout the world of higher education. At the LTH, these initiatives work together as one system with policies and practices supporting a planned and coordinated systems perspective on enhancement—at the level of the student, the teacher and the institution (Biggs 1993).

The overall goal for enhancement—what Bamber et al. (2009) call the 'theory of change' guiding implementation—is grounded in the belief that more and better conversations about student learning (especially among teachers and their students) lead to better quality education. The meaning of more conversations is perhaps self-evident, whereas 'better conversations' needs an explanation. One aspect of better is that conversations are deliberate, build on systematic observations of student learning and relate to relevant research on learning. Another aspect, articulated in LTH policy, encourages variation in teaching methods with all teachers expected to support their instructional choices with rational arguments. This cycle of discourse shaping practices and practices shaping discourse is identified with socio-cultural theory; an understanding that the social world of universities 'focuses on the way "practice" itself yields knowledge and learning' (Bamber et al. 2009: 9).

The pattern of discourse and practice in the present case is likened to growth toward the scholarship of teaching—from improving one's own teaching, to dialogue with colleagues about teaching and learning in the discipline, to carrying out scholarship to improve student learning (Weston and McAlpine 2001). An essential ingredient for success lies in supporting teaching as a mode of critical inquiry. The vision for a culture of scholarship in teaching and learning that mirrors what exists in research is, we believe, key to successful implementation. It is the application of the academic epistemology of critical inquiry, the type that can advance understanding and knowledge of a phenomenon—and the quality of that inquiry—that is improving over the years. Student ratings (with all their imperfections) play an essential role in this process of positive change.

## Course evaluations

*Valid course evaluation questionnaires let students report on what only they can report—their own learning experiences (Darwin 2017).*

Two decades ago, the Swedish government mandated universities to administer a course evaluation (CE) system with a view to enhancing the quality of teaching and giving more weight to the student experience. Students lobbied for this mandate during the 1990s and at the same time requested that CE results be made available to them. Policy within the LTH further differentiated between formative and summative evaluation (Warfvinge, 2003). Largely inspired by classroom assessment techniques (Angelo and Cross 1993), formative evaluation aims to encourage teachers to evaluate continuously during a course, to focus on student progress and to use the results in a productive dialogue with students about their expectations. Formative evaluation is strongly encouraged at the LTH as a signifying part of good teaching. Summative evaluation—using an end-of-course survey—generates statistics intended to inform changes in teaching practice. Policy explicitly states that statistical results should not influence career and promotion decisions, although, in accordance with the overall goal of teacher development, any individual putting forward for promotion is expected to demonstrate critical reflection on their course evaluation results.

The student survey administered at the LTH for course evaluation purposes is the Course Experience Questionnaire (CEQ). Conceived by British and Australian researchers, the CEQ is designed to measure the quality of students' learning experiences (Ramsden, 2005, 1991; Wilson et al. 1996). The researchers' original intent was to change the level of analysis, from the individual teacher to the academic unit responsible for teaching quality. The educational theory underpinning the CEQ distinguishes deep over surface approaches to student learning (Marton and Booth 1997; Ramsden 2005). Students are asked whether they experience features in a course that are known to encourage either a deep or surface approach to learning. Studies have shown that students are more likely to attempt to understand if they perceive teaching to be clearly structured and helpful and more likely to engage in rote learning under conditions of high workload and restricted choice over methods and content of learning (e.g. Ramsden and Entwistle 1981).

Question items relate to five scales: good teaching (clarity of explanation, level at which material is pitched, enthusiasm and help with study problems), clear goals and standards, appropriate workload, appropriate assessment and generic skills (i.e. problem solving, communications, planning and team work). One overall question prompts for students' satisfaction with the course. Although rarely used, teachers are encouraged to add up to three of their own questions to the questionnaire before it is distributed to students.

Two reports on the CEQ results are produced. A working report<sup>1</sup> includes only the statistical results—the means and distribution of scores, the percentage of students satisfied across all 26 items and contextual data (response rate, student gender, the extent to which they participated in various course activities). Comparison statistics among teachers in the same department and Faculty are not provided. Written comments on what students appreciated about the course and what is most in need of improvement are included but only after being vetted independently by students appointed by the student union.

The working report is distributed to the teacher, the two student representatives and the program coordinator for them to discuss at a meeting together. The LTH student union routinely

<sup>1</sup> An example can be found at [https://www.ceq.lth.se/info/dokument/filer/exempel\\_arbetsrapport\\_en.pdf](https://www.ceq.lth.se/info/dokument/filer/exempel_arbetsrapport_en.pdf)



recruits student representatives for every course and trains them as advocates for better quality education. This meeting is an opportunity for different stakeholders to engage in meaningful dialogue on their interpretations of factors that can potentially influence student ratings as well as their personal experiences with the course and data collected through formative measures. Generally, teachers at LTH are interested in students' reflections on courses and teaching and, with very few exceptions, students engage seriously during these CEQ meetings.

The program coordinator invites the teacher and the students to a 30–45 min meeting that normally takes place during a lunch break where participants bring something to eat. The students begin by offering their view of the course, usually reflecting upon not only the CEQ results and their personal experiences but also what they have heard from their classmates. Sometimes, they refer to the CEQ results from previous years. The teacher then offers his or her experience, and the meeting becomes a conversation with the program coordinator asking clarifying questions. For the program coordinator, this meeting is a valuable glimpse into the course. There are times when the atmosphere becomes tense, but, for the most part, these CEQ meetings are viewed as positive and meaningful for everyone concerned.

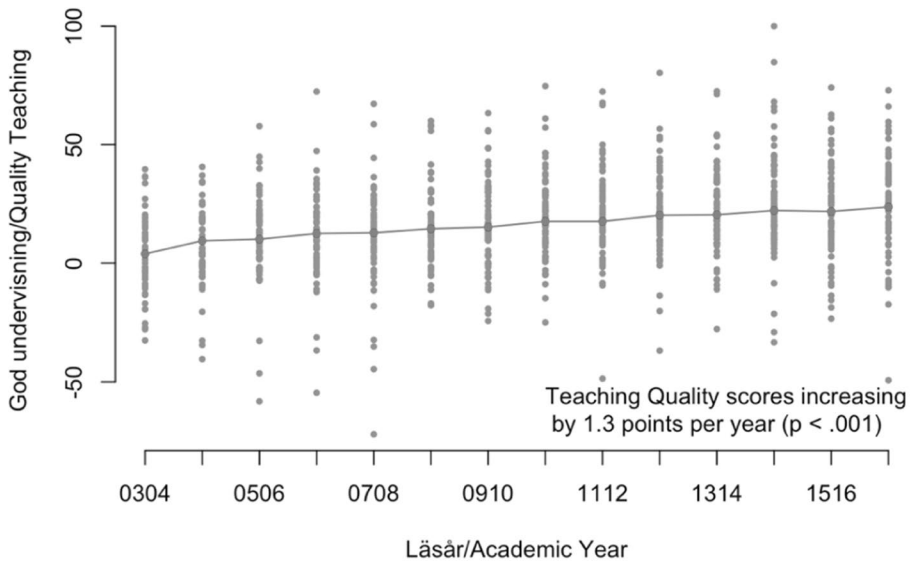
After the meeting, the three parties independently write up a brief response to the student ratings including changes to be made to course design and teaching, when appropriate. These responses form part of the final CE report together with the statistical results and exclude the written comments. Since the national regulations require that students should be informed about the outcome of the evaluation, the final report is emailed to all students registered in the course and uploaded to the Faculty's intranet for access by the head of department and senior managers. The fact that responses are in written form creates a 'trace of learning' (Wenger 1999) for each course; it also challenges participants to think hard on the matter because what they write will be shared with others.

What evidence is there that the LTH course evaluation system is having the desired effect? The analysis of aggregate CEQ results related to teaching quality over the past 13 years shows an increase in mean scores from 3.9 to 23.7 (Fig. 1). The most interesting finding in Fig. 1 is the linearity in development, not the numbers. We do not claim this development to be exceptional, even in a system designed for continuous input of novice teachers. Clearly, however, teaching at the LTH has steadily improved as perceived by a large number of students. If the LTH is worthy of inspiring others, it is because it has evolved in a positive direction over a long period of time.

The CEQ with its students approaches to learning theory is elegant in its simplicity. Questions asked of students about their course experience aid teachers' interpretation of how they might adjust their teaching to improve student learning. The same theoretical framework underpins all LTH enhancement initiatives promoting a common language of pedagogy throughout the Faculty. What is valued at the LTH is the teacher's ability to analyse and make use of the CEQ results, not the numbers produced by the system. It is the ability to acknowledge the need for interpretation and discussion of student ratings, with the process of negotiation almost becoming an end in itself that contributes to enhancement. A teacher's understanding of how students learn in their subject area begins during the CE process and deepens with participation in professional development opportunities.

## Pedagogical training

*'To respond to the complex ecology of higher education, academics necessarily have to become much more engaged learning professionals, who are able to self-monitor, reflect,*



**Fig. 1** Development in teaching at the LTH. Based on analysis of 247,224 surveys (on average 17,500 per academic year from 2003/2004 to 2016/2017). Questions rated using five-point Likert scale: (1) fully disagree coded  $-100$ , (3) neutral coded  $0$ , (5) fully agree coded  $+100$

*collaborate on current and improved practice and be subject to peer-level review and scrutiny*' (Darwin 2017: 21).

Professional training in university teaching was voluntary in Sweden during the 1990s and became mandatory in 2003. This external mandate has since been lifted; however, most institutions including the LTH decided to keep training mandatory as they feel it contributes to a coherent discourse on teaching quality. Pedagogical courses are designed in different ways depending on the institution but all adhere to one national framework (Association of Swedish Higher Education Institutions 2016). Presently, employment policy requires all teachers to engage in 200 h of training before employment or within 2 years of employment. Ph.D. students are also required to spend 20% of their time teaching as part of their doctoral program. In addition, promotion from lecturer to associate professor depends on evidence of critical inquiry and a capacity to reflect effectively on teaching and student learning within the discipline. Thus, all teachers are required to engage in a trajectory of growth in teaching and significant learning as they move from novice to expert teacher (Lave and Wenger 1991).

As mentioned, the vehicle for improvement at the LTH is the scholarship of teaching where teachers use observations and evidence of student learning in their arguments for why they teach the way they do (Felder 2000; Mårtensson et al. 2011). Teachers are required to make their arguments public in a scholarly report that is shared with colleagues. The aim is to motivate teachers to become deliberate and explicit about their teaching, developing what Shulman (1986) calls pedagogical content knowledge "(PCK). PCK is the knowledge that teachers develop over time, and through experience, about how to teach particular content in particular ways in order to lead to enhanced student understanding" (Loughran et al. 2012: 7). The pedagogical training program has a positive reputation mainly because it focuses on discipline-specific pedagogical realities experienced by

teachers rather than educational theory that has limited transfer value (e.g. Broekkamp and van Hout-Wolters 2007; Nuthall 2004).

Training is offered in modules of 80 to 120 h. Participants engage in dialogue with their colleagues about their own courses and teaching. The idea is to improve student learning within courses offered at the LTH through critical discourse and inquiry. An educational expert ensures the right conditions for learning: an appreciation of differing personal beliefs, colleagues talking with colleagues of equal status and facilitated dialogue (Bohm 1996; Senge 2006). The outcome of inquiry is a scholarly report shared with colleagues via the Faculty intranet and newsletter or campus-wide teaching and learning conference.

How do we know the LTH pedagogical training system is having the desired effect? A searchable database accessible within the LTH currently contains more than 600 scholarly reports where teachers in engineering have written to other teachers in engineering about teaching and student learning. This represents a percentage of more or less 50% of faculty who have accomplished training. Many of the reports result from participants taking the training multiple times as well as group projects with the main benefit coming from teachers co-writing a short text with their colleagues about student learning in their discipline. Research on these reports shows improvement in the focus on student learning, the use of systematic observations with reference to relevant literature and in terms of coherent writing (Larsson et al. 2015). Making scholarly reports public contributes to an increased awareness of engineering education throughout the Faculty; it also motivates teachers to continue engaging meaningfully in professional development even after training. (For further description and observations, including summary of five external national and international evaluations, see Warfving, 2018.)

## Reward for excellence in teaching

*It is the notion of quality negotiated over time among all stakeholders in the system—students, teachers and administrators—that has the greatest potential for educational change (Trowler 2002).*

Introduced in 2001, the LTH reward for excellence provides administrators an opportunity to contribute to the quality discourse by clarifying the criteria for excellence against which teaching can be judged. The reward aligns with other parts of the enhancement system around three criteria: (1) a learning-centred approach to teaching (initiated during the CE process), (2) signs of continuous development as a teacher (initiated during pedagogical training) and (3) evidence of being engaged with colleagues about discipline-specific educational matters (continued after training). Reference to literature and research on teaching and learning is also expected.

The reward confers the title of Excellent Teaching Practitioner (ETP) on successful applicants together with a permanent raise in salary (Olsson and Roxå 2013). The recipient's department also receives a permanent increase in funding (taken from the budgets of all departments) equivalent to the amount awarded when a researcher is promoted to Docent (a senior research title). The ETP application process involves submitting a teaching dossier that consists of two parts. The first part is a statement on learning (rather than the traditional teaching philosophy statement) in which applicants explain their beliefs about how students learn and specific challenges to learning in their subject area. The second part comprises a number of cases written up as scholarly reflections where applicants demonstrate through the use of observations of student learning the link between their intentions in teaching and students' actual learning outcomes. The

assessment then focuses on the coherence between these two parts, especially the quality of the arguments for student learning put forward by the applicant. CE results are referenced only as auxiliary material.

Cases are mostly descriptions of a specific teaching problem, such as identified learning thresholds among students or not enough independence demonstrated by students during project work. An example of a case in mechanical engineering focused on how students fail to link theory (presented to them in lectures) with practical design assignments (performed during laboratory work). The teacher presented several observations on how she perceived the problem and outlined a few interventions, such as deliberate instruction to link theory and practice in lab reports. In a second intervention, she linked theory and practice problems more explicitly in the final exam. When during a third iteration she introduced a mid-course diagnostic test with a similar type of exam, she observed effects that mirrored her expectations. She realized the need for practice and feedback related to a major learning outcome; the traditional lecture and lab were not sufficient.

Previous reward recipients assess the ETP teaching dossiers against the three criteria specified above and conduct in-depth interviews with applicants. If the applicant meets the criteria, he or she receives the reward. If not, the process itself is seen as an opportunity for constructive feedback and mentorship from excellent teachers. The main tenet of the ETP reward is peer review that leads to possibilities for learning not only in individuals but also in the institution (Olsson and Roxå 2013). The reward scheme has been in place for almost 20 years, and, to date, over 100 teachers have earned the ETP title.

How do we know the reward scheme is having the desired effect? The proportion of ETPs among senior managers (56%), heads of departments (39%) and program coordinators (25%) is higher than the proportion among regular teaching staff (18–20%). Most of these recipients earned their ETP status and acted on the ETP review board early in their careers, an experience they have carried with them as they shoulder more responsibilities. They are individuals recognized for engaging in what is productive for the LTH (Olsson and Roxå 2013). Unlike university awards for teaching excellence that rely on nominations of a small number of teachers across a Faculty, any teacher can apply for the ETP reward and be subject to rigorous peer review, in the same way as they can apply for the title of Docent (senior researcher). The number of people who have been granted a reward provides evidence of faculty engagement in the scholarship of teaching and far-reaching conversations on student learning.

Perhaps more impressively, the LTH reward scheme has spread to the Faculties of Science and Social Sciences at Lund University as well as to 27 other institutions in the country (Winka 2017) and a number of institutions internationally (Olsson et al. 2018). Evidence to support the ETP initiative has been published by Lund scholars (Mårtensson et al. 2011; Roxå and Mårtensson 2016). Additional research suggests that expertise developed from being an assessor on an ETP panel spreads to processes of hiring and promotion (Mårtensson 2010), and the quality of these conversations has improved over time (Kottmann et al. 2016).

The ETP funding model further communicates a disciplinary community that values teaching excellence. Shulman (1993) calls this “teaching as community property” that means the disciplinary community takes responsibility for teaching quality, and this requires written artefacts to capture the complexity and richness of teaching as well as peer review applied to teaching (and not simply colleagues in the office next door). Teaching in solitude was never ideal; now in a culture that gives more weight to the student experience, it is not even viable.

## Student engagement in enhancement

*'The challenge is for higher education institutions committed to learning to consider moving beyond their familiar and conventional approaches centred on student ratings alone and to investigate deeper and more qualitative engagement with the student voice' (Darwin 2017: 21).*

Like most institutions, the CE system at the LTH relies on students responding to an end-of-course survey that is by definition a passive activity. In fact, one of the biggest issues in the transition to online surveys is incentivising students to respond. The average response rate at the LTH is between 30 and 35%. Research in Lund and Stockholm shows that this percentage is sufficient to get an overview of students' experiences (Havtun and Hjelm, 2019). The researchers have also shown that a response rate of 80% does not produce alternative statistical patterns. Furthermore, students are engaged in two ways: firstly, they are represented in the follow-up meeting where the working report is discussed; secondly, they all receive the results by email shortly after completing the course. Lastly, response rate is not an issue at the LTH because it is clear in policy as well as in practice that student ratings are *not* a measurement of quality, but data to be interpreted by those directly involved in the course.

CE policy at the LTH further prompts teachers to elicit formative feedback from students to provide an opportunity to discuss course improvements early enough in the course to make a difference to their learning. This also allows for more targeted course-specific feedback. Students are ideally situated to contribute ideas on what helps or hinders their learning rather than simply to respond to questions generated by others at different stakeholder levels. One criticism of a survey like the CEQ is that it identifies generic features of a course that do not always match factors specific to a particular course context or the uniqueness of the discipline. Formative evaluation then becomes an opportunity to engage students more deeply in discussions on learning to learn in their discipline.

In higher education, graduate students are often involved in course improvement in their role as teaching assistants and, sometimes, undergraduate students in their role as tutors and peer mentors. Rarely are students systematically involved as subjects of inquiry on improving teaching or as participants in research on improving student learning, as they are at the LTH. During the process described, interventions and evaluation measures are designed to address specific issues related to student learning and performance. Students have an opportunity to speak on behalf of their peers and to see what it takes behind the scenes to make the system work.

The benefits of students working in partnership with teachers on pedagogical innovation and reflection are well documented (Cook-Sather 2013; Healey et al. 2016). There is also evidence that such partnerships result in changing student and faculty approaches to teaching and learning (Cook-Sather et al. 2014). Thus, when students are paid as part of a formal partnership program, resources can be allocated towards institutional change that is genuinely transformative (Ahmad et al. 2017).

## Potential challenges to implementation

Top-down policies are known to be ineffective in promoting change in higher education (Henderson et al. 2011). Successful strategies 'are aligned with or seek to change the beliefs of the individuals involved; involve long-term interventions, lasting at least one semester;

and require understanding a college or university as a complex system and designing a strategy that is compatible with this system' (952). The LTH approach to enhancement (summarized in Table 1) represents a series of successful strategies that have evolved over a period of 25 years. This principled approach is intended to advocate for synergy between student ratings of teaching and the development and recognition of quality teaching. It is not intended as a master plan for others to implement without serious consideration.

The first potential barrier to implementation is that Lund University affords academic freedom to the LTH to negotiate its own quality indicators in engineering education—captured in their criteria for excellence in teaching. The first criterion is for a learning-oriented approach to teaching. Most teachers would support moving students away from evaluating their teaching (for which they are not qualified) to letting them report on their own learning experiences. Would this require adopting the CEQ and its theoretical framework or a similarly grounded and validated questionnaire? Potentially, yes. Importantly, however, the frameworks and tools used should be judged primarily by how well they feed discourse about teaching and learning and can guide improvements in quality. We are not advocating for the CEQ or the perfect analytical tool but rather getting to the heart of the issue that is supporting teachers in making best use of CE results for improving student learning. Another important aspect concerns the role of the student union in recruiting student representatives for every course and training them as advocates for enhancement. This practice of electing student delegates has long been established in Swedish universities but may be challenging to replicate in another university context.

The second criterion for excellence in teaching at the LTH is continuous development as a teacher. The LTH exists in a national and institutional context that values professionalism in teaching with mandatory pedagogical training required for hiring and promotion purposes (a culture prevalent in Europe and Australia but not in North America). The LTH further requires a significant portion of a Ph.D. curriculum devoted to teaching. Both policies may transfer more easily to a smaller institution or Faculty in the same discipline. Successful implementation would also depend upon an inquiry-based training programme facilitated by highly skilled educational experts. The funding model for the ETP reward for excellence operating in parallel to an existing tenure and promotion process may be equally challenging to negotiate; although, there are examples of the reward being successfully implemented elsewhere.

It is important to note that training was first piloted at the LTH following the student union's national inquiry into higher education in 1988. Funding for training doctoral students and experienced teachers was negotiated between the Swedish government and the faculty union during the financial crisis in 1991 as a replacement for raising salaries. There has been almost no resistance to this training. It began with people who were interested, and positive momentum was built department by department until fully scaling up in 2000. The key is to work with problems that teachers are experiencing; then, they will come back for more. Pedagogical courses have the additional benefit of raising awareness before making structural changes. An Erasmus comparative case study provides more details on this initiative from its inception (Simon et al., 2019).

The third criterion for excellence involves teachers engaging with colleagues in discipline-specific educational matters, and this includes conducting scholarship on teaching and learning. In terms of epistemology, the LTH approach relies upon growth in the scholarship of teaching to make the overall system work. Teachers can initially be sceptical about this concept, feeling that it is not their responsibility to do research on teaching but the purview of educational researchers. We would argue that educational research is directed towards generalized knowledge of teaching and learning, not the transfer of this

**Table 1** The LTH approach to enhancement: advocating for synergy between student ratings of teaching and the development and recognition of quality teaching

	Course evaluation system	Pedagogical support	Reward for excellence
Policies related to educational change	Policy prompts for both formative and summative evaluation measures	Policy requires mandatory training (200 h) and Ph.D. time on teaching (20%)	Recipients receive a title and pay increase and their department equal funding
Principles related to excellence in teaching	Learner-centered approach to teaching (course experience questionnaire)	Continuous development as a teacher (facilitated dialogue with colleagues)	Engagement with colleagues on engineering educational matters
Practices related to dialogue on teaching and learning	Teachers reflect on results with two student reps and programme coordinator	Teachers engage in self-inquiry into own teaching and student learning	Teachers engage in scholarly inquiry on teaching and learning
Practices related to scholarship on teaching and learning	Final reports include stakeholder responses and planned improvements	Scholarly reports include evidence of inquiry into improved teaching practice	Teaching dossiers include evidence of research into improved student learning

knowledge to a specific teaching context. At the LTH, teachers generally experience scholarship in teaching and learning positively as a way to acquire better information about students and courses allowing them to make better decisions about their teaching. Teachers need motivation to do what only they can do—discipline-based research on student learning. Since conducting research can jeopardize an already heavy workload, teachers at the LTH have the option to pursue a reward for excellence or not. This personal enhancement agenda may not transfer easily to an institutional context that does not appeal to similar values and beliefs (Kezar 2018).

To sum up, the LTH approach to enhancement represents an internal accountability framework that values student input and improved learning outcomes and links funding to quality teaching, similar in its altruistic goal to the UK's TEF framework. The approach illustrates how student ratings can be used effectively for developmental purposes to spark rich conversations about student learning with conversations over time among individuals driving growth. It is hard to say retrospectively where these conversations began, and they certainly have not ended. Conversations continue today with development of new LTH policy on a cyclical program review process. It is fair to say, reflecting over the LTH case, that it is the conversations between stakeholders especially teachers and students that drive positive change; the course evaluation, pedagogical training and reward systems are rather necessary scaffolding for these conversations.

## Conclusion

In this paper, we argue that student ratings of teaching have the potential to function as a tool for the development of teaching and learning. We also argue that because of the issues associated with student ratings, too much emphasis on control and measurement diminishes their potential to fuel development. When results are used instrumentally in management processes without consideration of the biases and contextual differences built into them, they are less meaningful to those directly involved—teachers and students.

We acknowledge that student ratings are ubiquitous in higher education. We also argue that when too much weight is placed on using student ratings to measure quality, they are likely to impede educational innovation, negatively affect the working conditions of early career teachers and further alienate the student voice from discussions on educational quality. We nevertheless argue that institutions should not abandon collecting student perceptions through surveys; yet to improve practice, these surveys must also become the starting point for critical discourse over effective teaching practices. It is not rational to teach thousands of students in ways that will develop them epistemologically and ethically without also engaging them in discussions about their learning experiences.

Higher education is not a laboratory experiment where success is determined by measuring what goes in and what comes out. Our most important argument to reduce the drift that continues to occur due to conventional student evaluation systems is one that re-engages students in conversations about improving the quality of their learning. We further contend that only when student ratings feed reflective conversations about improving the learning process, and these conversations are informed by local practice-based research and evidence-based pedagogical principles, can they fulfil their original purpose of driving teaching development.

We illustrate our thesis with a case example that supports critical dialogue among all stakeholders about educational quality. We anticipate that additional cases will offer further



insights to spark a critical analysis on how institutions can further reconceptualize the role of student ratings. After all, critical dialogue is the most promising strategy for change in academia and has always been a powerful way to move our collective forward.

Furthermore, our case example illustrates a response to a universal problem. Simply providing data from evaluative processes and expecting positive change is like using a trial-and-error process. The same argument can be used against introducing market-driven change processes, since these encourage institutional managers to use similar strategies for change forcing higher education institutions to become more and more similar in attempts to compete with each other based on a limited number of quality aspects. In the private sector, this type of organisational behaviour has been observed for a long time—organisations mimic each other in their effort to compete (DiMaggio and Powell 1983). To avoid this tendency, policies and initiatives aimed at improving teaching and learning need a strategy for change anchored in a robust theory of change (Kezar 2018).

We conclude that the core issue is not the quality of student ratings but the quality of the discourse they generate. Given the many different contexts in which institutions find themselves, is it possible to embed student ratings of their learning experiences in a wider discussion, as argued here? What would this look like in your context? What are the barriers? Where are the voices of students and teachers heard?

## References

- Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). The validity of student ratings of instruction: what we know and what we don't. *Journal of Educational Psychology*, 82, 219–232.
- Ahmad, A., Ali, A., VanMaaren, J., Barrington, J., Merritt, O., & Ansilio, K. (2017). Partnership in practice: implementing Healey's conceptual model. *International Journal for Students as Partners*, 1(2), 1–10. <https://doi.org/10.15173/ijasp.v1i2.3197>
- Alhija, F.N.-A. (2017). Teaching in higher education: good teaching through students' lens. *Studies in Educational Evaluation*, 54, 4–12. <https://doi.org/10.1016/j.stueduc.2016.10.006>
- Ambady, N., & Rosenthal, R. (1993). Half a minute: predicting teacher evaluations from thin slices of non-verbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64(3), 431–441.
- Andersen, K., & Miller, E. D. (1997). Gender and student evaluations of teaching. *PS: Political Science and Politics*, 30(2), 216–219.
- Ang, L., Breyer, Y. A., & Pitt, J. (2018). Course recommendation as a construct in student evaluations: will students recommend your course? *Studies in Higher Education*, 43(6), 944–959. <https://doi.org/10.1080/03075079.2016.1199543>
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: a handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass.
- Arthur, L. (2009). From performativity to professionalism: lecturers' responses to student feedback. *Teaching in Higher Education*, 14(4), 441–454. <https://doi.org/10.1080/13562510903050228>
- Association of Swedish Higher Education Institutions. (2016). *Recommendations on general learning outcomes for the teaching qualifications required for employment as academic teacher and on mutual recognition* (No. REX 2016:1). Stockholm, Sweden. [https://subf.se/app/uploads/2019/03/REK-2016-1-On-general-learning-outcomes-for-teaching-qualifications\\_Dnr-024-16.pdf](https://subf.se/app/uploads/2019/03/REK-2016-1-On-general-learning-outcomes-for-teaching-qualifications_Dnr-024-16.pdf)
- Bamber, V., Trowler, P. R., Saunders, M., & Knight, P. (Eds.). (2009). *Enhancing learning, teaching, assessment and curriculum in higher education: theory, cases, practices*. New York: Open University Press.
- Basow, S., Codos, S., & Martin, J. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal*, 47(2), 352–363.
- Baxter Magolda, M. B. (1992). *Knowing and reasoning in college: gender-related patterns in students' intellectual development*. San Francisco: Jossey-Bass.
- Bedggood, R. E., & Donovan, J. D. (2012). University performance evaluations: what are we really measuring? *Studies in Higher Education*, 37(7), 825–842. <https://doi.org/10.1080/03075079.2010.549221>
- Beran, T. N., & Rokosh, J. L. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science*, 37(2), 171–184. <https://doi.org/10.1007/s11251-007-9045-2>

- Biggs, J. B. (1993). From theory to practice: a cognitive systems approach. *Higher Education Research and Development*, 12(1), 73–85.
- Bini, M., & Masserini, L. (2016). Students' satisfaction and teaching efficiency of university offer. *Social Indicators Research*, 129(2), 847–862. <https://doi.org/10.1007/s11205-015-1141-0>
- Blair, E., & Noel, K. V. (2014). Improving higher education practice through student evaluation systems: is the student voice being heard? *Assessment & Evaluation in Higher Education*, 39(7), 879–894. <https://doi.org/10.1080/02602938.2013.875984>
- Boatright-Horowitz, S. L., & Soeung, S. (2009). Teaching white privilege to white students can mean saying good-bye to positive student evaluations. *American Psychologist*, 64(6), 574–575. <https://doi.org/10.1037/a0016593>
- Bohm, D. (1996). On dialogue. New York: Routledge.
- Borg, W. R., & Hamilton, E. R. (1956). Comparison between a performance test and criteria of instructor effectiveness. *Psychological Reports*, 2, 111–116. <https://doi.org/10.2466/pr0.1956.2.3.111>
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27–41. <https://doi.org/10.1016/j.jpubeco.2016.11.006>
- Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41, 71–88. <https://doi.org/10.1016/j.econedurev.2014.04.002>
- Broekkamp, H., & van Hout-Wolters, B. (2007). The gap between educational research and practice: a literature review, symposium, and questionnaire. *Educational Research and Evaluation*, 13(3), 203–220. <https://doi.org/10.1080/13803610701626127>
- Cook-Sather, A. (2013). Catalyzing multiple forms of engagement: student-staff partnerships exploring teaching and learning. *The student engagement handbook: practice in higher education* (pp. 549–566). Emerald Publishing Groups: Bingley.
- Cook-Sather, A., Bovill, C., & Felten, P. (2014). *Engaging students as partners in learning and teaching: a guide for faculty*. San Francisco: Jossey Bass.
- Copeland, R. (2017). Understanding the United Kingdom's Teaching Excellence Framework and its implications. *Academic Matters: OCUFA's Journal of Higher Education*. Winter. <https://academicmatters.ca/understanding-the-united-kingdoms-teaching-excellence-framework-and-its-implications/>
- Darwin, S. (2017). What contemporary work are student ratings actually doing in higher education? *Studies in Educational Evaluation*, 54, 13–21. <https://doi.org/10.1016/j.stueduc.2016.08.002>
- Department for Business, Innovation, and Skills. (2016). *Success as knowledge economy: teaching excellence, social mobility & student voice* (White Paper). London. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/523396/bis-16-265-success-as-a-knowledge-economy.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/523396/bis-16-265-success-as-a-knowledge-economy.pdf). Accessed 2 February 2019
- DiMaggio, P., & Powell, W. (1983). The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160.
- Edström, K. (2008). Doing course evaluation as if learning matters most. *Higher Education Research & Development*, 27(2), 95–106. <https://doi.org/10.1080/07294360701805234>
- Felder, R. M. (2000). Random thoughts: the scholarship of teaching. *Chemical Engineering Education*, 34(2), 144.
- Field, L. (2015). Whither teaching? Academics' informal learning about teaching in the “tiger mother” university. *International Journal for Academic Development*, 20(2), 113–125. <https://doi.org/10.1080/1360144X.2015.1028067>
- Flaherty, C. (2018). Teaching eval shake-up. *Inside Higher Ed*. <https://www.insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story>
- Flodén, J. (2017). The impact of student feedback on teaching in higher education. *Assessment & Evaluation in Higher Education*, 42(7), 1054–1068.
- Grimes, A., Medway, D., Foos, A., & Goatman, A. (2017). Impact bias in student evaluations of higher education. *Studies in Higher Education*, 42(6), 945–962. <https://doi.org/10.1080/03075079.2015.1071345>
- Gunn, B. (1995). The paradigm shift in university management. *International Journal of Educational Management*, 9(1), 28–40. <https://doi.org/10.1108/09513549510146584>
- Hammer, R., Peer, E., & Babad, E. (2018). Faculty attitudes about student evaluations and their relations to self-image as teacher. *Social Psychology of Education*, 21, 517–537. <https://doi.org/10.1007/s11218-018-9426-1>
- Havtun, H., & Hjeltn, N. (2019). En undersökning av vad de studenter som inte fyller i kursenkäten tycker om kursen [An investigation of what students who didn't complete the course evaluation think of the course]. Presented at the 7th Development Conference for Swedish Engineering Education, Lulea Technical University.

- Healey, M., Flint, A., & Harrington, K. (2016). Students as partners: reflections on a conceptual model. *Teaching & Learning Inquiry*, 4(2), 1–13.
- Henderson, C., Beach, A., & Finkelstein, H. (2011). Facilitating change in undergraduate stem instructional practices: an analytic review of the literature. *Journal of Research in Science Teaching*, 48(8), 952–984. <https://doi.org/10.1002/tea.20439>
- Hendry, G. D., Lyon, P. M., & Henderson-Smart, C. (2007). Teachers' approaches to teaching and responses to student evaluation in a problem-based medical program. *Assessment & Evaluation in Higher Education*, 32(2), 143–157. <https://doi.org/10.1080/02602930600801894>
- Huston, T. A. (2006). Race and gender bias in higher education: could faculty course evaluations impede further progress toward parity. *Seattle Journal of Social Justice*, 4(2), 591–611.
- Kember, D., Leung, D. Y. P., & Kwan, K. P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education*, 27(5), 411–425. <https://doi.org/10.1080/0260293022000009294>
- Kezar, A. (2018). *How colleges change: understanding, leading, and enacting change* (2nd ed.). London: Routledge.
- Kottmann, A., Huisman, J., Brockerhoff, L., Cremonini, L., & Mampaey, J. (2016). *How can one create a culture for quality enhancement? Final report*. Enschede: Center for Higher Education Policies.
- Kulik, J. A. (2001). Student ratings: validity, utility, and controversy. *New Directions for Institutional Research*, 109, 9–25. <https://doi.org/10.1002/ir.1>
- Kwan, K. P. (1999). How fair are student ratings in assessing the teaching performance of university teachers? *Assessment & Evaluation in Higher Education*, 24(2), 181–195. <https://doi.org/10.1080/0260293990240207>
- Larsson, M., Anderberg, E., & Olsson, T. (2015). Researching the transformation in the scholarship of teaching and learning through teaching portfolios and conference papers: a reflection at the institutional level. In S. A. Booth & L. Woollacott (Eds.), *The scholarship of teaching and learning in higher education: On its constitution and transformative potential* (pp. 113–136). Stellenbosch: Sun Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Linse, A. R. (2017). Interpreting and using student ratings data: guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94–106. <https://doi.org/10.1016/j.stueduc.2016.12.004>
- Loughran, J., Berry, A., & Mulhall, P. (2012). Pedagogical content knowledge. In J. Loughran, A. Berry, & P. Mulhall (Eds.), *Understanding and developing science teachers' pedagogical content knowledge* (Vol. 12, pp. 7–14). Rotterdam: SensePublishers.
- Lutovac, S., Kaasila, R., Komulainen, J., & Maikkola, M. (2017). University lecturers' emotional responses to and coping with student feedback: a Finnish case study. *European Journal of Psychology of Education*, 32(2), 235–250. <https://doi.org/10.1007/s10212-016-0301-1>
- Marsh, H. W. (1980). The influence of student, course, and instructor characteristics in evaluations of university teaching. *American Educational Research Journal*, 17(2), 219–237.
- Marsh, H. W., & Cheng, J. (2008). *National student survey of teaching in UK universities: dimensionality, multilevel structure, and differentiation at the level of university and discipline (preliminary results)*. New York: Higher Education Academy.
- Marsh, H. W., & Roche, L. A. (1999). Reply upon SET research. *American Psychologist*, 54(7), 517–518. <https://doi.org/10.1037/0003-066X.54.7.517>
- Mårtensson, K. (2010, December 15). *Bedömning av pedagogisk kompetens på LTH [Evaluation of pedagogical competence at Lund University Faculty of Engineering]*. Presented at the Pedagogiska Inspirationskonferens [Pedagogical Inspiration Conference], Lund University.
- Mårtensson, K., Roxå, T., & Olsson, T. (2011). Developing a quality culture through the Scholarship of Teaching and Learning. *Higher Education Research & Development*, 30(1), 51–62. <https://doi.org/10.1080/07294360.2011.536972>
- Marton, F., & Booth, S. A. (1997). *Learning and awareness*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McCune, V., & Hounsell, D. (2005). The development of students' ways of thinking and practising in three final-year biology courses. *Higher Education*, 49(3), 255–289. <https://doi.org/10.1007/s10734-004-6666-0>
- Meyer, H. (2002). The new managerialism in education management: corporatization or organizational learning? *Journal of Educational Administration*, 40(6), 534–551. <https://doi.org/10.1108/09578230210446027>

- Moore, S., & Kuol, N. (2005). Students evaluating teachers: exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education*, 10(1), 57–73. <https://doi.org/10.1080/1356251052000305534>
- Morgan, D. A., Sneed, J., & Swinney, L. (2003). Are student evaluations a valid measure of teaching effectiveness: perceptions of accounting faculty members and administrators. *Management Research News*, 26, 17–32. <https://doi.org/10.1108/01409170310783565>
- Morrison, K. (2013). Online and paper evaluations of courses: a literature review and case study. *Educational Research and Evaluation*, 19(7), 585–604. <https://doi.org/10.1080/13803611.2013.834608>
- Murray, H. G. (1997). Does evaluation of teaching lead to improvement of teaching? *International Journal for Academic Development*, 2(1), 8–23. <https://doi.org/10.1080/1360144970020102>
- Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, 27(2), 187–198. <https://doi.org/10.1080/02602930220128751>
- Nuthall, G. (2004). Relating classroom teaching to student learning: a critical analysis of why research has failed to bridge the theory-practice gap. *Harvard Educational Review*, 74(3), 273–306. <https://doi.org/10.17763/haer.63.1.uh00236162314763>
- O'Donovan, B. (2017). How student beliefs about knowledge and knowing influence their satisfaction with assessment and feedback. *Higher Education*, 74(4), 617–633. <https://doi.org/10.1007/s10734-016-0068-y>
- Olsson, T., & Roxå, T. (2013). Assessing and rewarding excellent academic teachers for the benefit of an organization. *European Journal of Higher Education*, 3(1), 40–61. <https://doi.org/10.1080/21568235.2013.778041>
- Olsson, T., Winka, K., Førlund, O., & Roxå, T. (2018). *Implementation of reward systems for excellence in university teaching: critical aspects from a nordic perspective*. Higher Education Close-Up. Cape Town: RSA.
- Ontario Confederation of University Faculty Associations. (2019). *Report of the OCUFA student questionnaires on courses and teaching working group*. <https://ocufa.on.ca/assets/OCUFA-SQCT-Report.pdf>. Accessed 17 March 2019
- Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years*. Oxford, England: Holt, Rinehart & Winston.
- Ramsden, P. (Ed.). (2005). *Learning to teach in higher education*. London: Routledge.
- Ramsden, P., & Entwistle, N. J. (1981). Effects of academic departments on students' approaches to studying. *British Journal of Educational Psychology*, 51, 368–383.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: the Course Experience Questionnaire. *Studies in Higher Education*, 16(2), 129–150.
- Roxå, T., & Bergström, M. (2013). Kursvärderingar i system – akademiska lärares upplevelser och organisations förmåga till utveckling [Course evaluation systems: academic teachers' experiences and the institution's capacity for development]. *Högere Utbildning [Higher Education]*, 3(3), 225–236.
- Roxå, T., & Mårtensson, K. (2016). Developing higher education organisations and evaluating effects. In R. Egger & M. Merkt (Eds.), *Teaching skills assessments - Qualitätsmanagement und Personalentwicklung in der Hochschullehre* (Vol. 17, pp. 91–115). Wiesbaden: Springer.
- Roxå, T. (2013). Slow and steady - educational development at Lund University. In R. Macdonald (Ed.), *SEDA Special 33 - supporting educational change* (pp. 29–30). London: Staff and Educational Development Association.
- Roxå, T., Olsson, T., & Martensson, K. (2008). Appropriate use of theory in the Scholarship of Teaching and Learning as a strategy for institutional development. *Arts and Humanities in Higher Education* 7(3), 276–294. <https://doi.org/10.1177/1474022208094412>
- Ryerson University v. Ryerson Faculty Association. Ryerson University v Ryerson Faculty Association. , Pub. L. No. CanLII 58446 (ON LA) (2018). Toronto. <https://canlii.ca/t/hsqkz>. Accessed 24 October 2019
- Schmelkin, L. P., Spencer, K. J., & Gellman, E. S. (1997). Faculty perspectives on course and teacher evaluations. *Research in Higher Education*, 38(5), 575–592. <https://doi.org/10.1023/A:1024996413417>
- Senge, P. M. (2006). *The art & practice of the learning organization. The fifth discipline*. New York: Doubleday.
- Shakurnia, A. H., Mozaffari, A. R., & Karami, M. A. (2012). Comparison of teachers' attitude with high and low scores evaluation on the importance and use of student assessments. *Educational Research in Medical Sciences*, 1(1), 38–46.
- Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1993). Teaching as community property: putting an end to pedagogical solitude. *Change*, 25(6), 6–7.

- Simon, A., Karm, M., Soomere, T., Roxå, T., Löfgreen, J., Roxå, T., Löfgreen, J., Simon, E., & Pleschová, G. (2019). *A comparative study of three educational development courses for doctoral students and teaching and learning initiatives in higher education*. Erasmus+ [https://qualityteaching.euba.sk/www\\_write/files/project-outcomes/o4\\_study\\_with\\_supplements.pdf](https://qualityteaching.euba.sk/www_write/files/project-outcomes/o4_study_with_supplements.pdf)
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: the state of the art. *Review of Educational Research*, 83(4), 598–642. <https://doi.org/10.3102/0034654313496870>
- Spooren, P., & Christiaens, W. (2017). 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. *Studies in Educational Evaluation*, 54, 43–49. <https://doi.org/10.1016/j.stueduc.2016.12.003>
- Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: what we can't count can hurt us. *Sex Roles*, 53(11–12), 779–793. <https://doi.org/10.1007/s11199-005-8292-4>
- Stein, S. J., Spiller, D., Terry, S., Harris, T., Deaker, L., & Kennedy, J. (2013). Tertiary teachers and student evaluations: never the twain shall meet? *Assessment & Evaluation in Higher Education*, 38(7), 892–904. <https://doi.org/10.1080/02602938.2013.767876>
- Stroebe, W. (2016). Why good teaching evaluations may reward bad teaching: on grade inflation and other unintended consequences of student evaluations. *Perspectives on Psychological Science*, 11(6), 800–816. <https://doi.org/10.1177/1745691616650284>
- Trowler, P., & Cooper, A. (2002). Teaching and learning regimes: implicit theories and recurrent practices in the enhancement of teaching and learning through educational development programmes. *Higher Education Research & Development*, 21(3), 221–240. <https://doi.org/10.1080/0729436022000020742>
- Trowler, P. R. (2002). *Higher education policy and institutional change*. Buckingham: SRHE & Open University Press.
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22–42. <https://doi.org/10.1016/j.stueduc.2016.08.007>
- Warfvinge, P. (2003, July 1). Policy for evaluation of first and second cycle studies. Lunds Tekniska Högskola. [https://www.ceq.lth.se/info/dokument/filer/utvarderingspolicy\\_eng.pdf](https://www.ceq.lth.se/info/dokument/filer/utvarderingspolicy_eng.pdf). Accessed 16 October 2018
- Warfvinge, P. (2018). *Centre for Engineering Education at Lund University - self-assessment*. Lund, Sweden: Lund University. [https://www.lth.se/fileadmin/cee/Documents/CEE\\_Self\\_Assessment\\_Final.pdf](https://www.lth.se/fileadmin/cee/Documents/CEE_Self_Assessment_Final.pdf). Accessed 26 June 2020
- Wenger, E. (1999). *Communities of practice: learning, meaning, and identity*. Cambridge: Cambridge University Press.
- Weston, C. B., & McAlpine, L. (2001). Making explicit the development toward the scholarship of teaching. *New Directions for Teaching and Learning*, 86, 89–97. <https://doi.org/10.1002/tl.19>
- Wilson, K. L., Lizzio, A., & Ramsden, P. (1996). *The use and validation of the Course Experience Questionnaire*. Brisbane: Griffith Institute for Higher Education, Griffith University.
- Winka, K. (2017). *Kartläggning av pedagogiska meriteringsmodeller vid Sveriges högskolor och universitet [Mapping models for evaluating pedagogical merit in Swedish higher education]*. Göteborg: Göteborgs University.
- Wolbring, T., & Riordan, P. (2016). How beauty works: theoretical mechanisms and two empirical applications on students' evaluation of teaching. *Social Science Research*, 57, 253–272. <https://doi.org/10.1016/j.ssresearch.2015.12.009>
- Wong, W. Y., & Moni, K. (2014). Teachers' perceptions of and responses to student evaluation of teaching: purposes and uses in clinical education. *Assessment & Evaluation in Higher Education*, 39(4), 397–411. <https://doi.org/10.1080/02602938.2013.844222>
- Yao, Y., & Grady, M. L. (2005). How do faculty make formative use of student evaluation feedback? A multiple case study. *Journal of Personnel Evaluation in Education*, 18(2), 107–126. <https://doi.org/10.1007/s11092-006-9000-9>