

Chapter 2

Research on the Design and Function of Student Evaluation

Abstract The exploration of the primary research focus of student evaluation—particularly the processes and integration of quantitative forms of survey-based student feedback—is foregrounded in this chapter. The critical foundations of this research have been largely framed around the validity and reliability of ratings-based metrics for teaching and course effectiveness. More contemporary research on student feedback has continued to focus on the improved survey design methods, as well as the development of frameworks for enhancing the capacity for comparative analysis of student evaluation data (such as against faculty, institutional or sectoral benchmarks). Other research has sought to investigate how the outcomes of student evaluation could be more effectively disseminated and integrated into the pedagogical actions of teachers. However, less research interest has been directed toward the foundational epistemological assumptions that underpin the use of quantitative student evaluation data for pedagogical and administrative decision-making in higher education. In this chapter, the broad range of research directed toward student evaluation is assessed, along with some of the issues that have received less scholarly interest. What this demonstrates is that student evaluation has attracted much stronger research interest from statisticians and academic developers than educational researchers.

Keywords Student evaluation research • Quantitative student evaluation • Epistemologies of student feedback • Limitations of student evaluation research

Introduction

In higher education environments, student evaluation is being increasingly employed to respond to multiple imperatives around quality improvement, quality assurance and performance management. In the complex teaching and learning environments of contemporary institutions, it is performing significant work as a usefully reductive and comparable signifier of teaching quality. Aside from student

evaluation being normalised within institutional life, over the last two decades student evaluation data has also moved beyond the academic and faculty level. Progressively it has become a highly regarded comparator for institutional assessment and a critical input in various university ranking scales.

As a consequence, it is also significantly shaping prospective student (and broader social) perceptions as to the value of higher education institutions and the programs they offer. For instance, the *Course Experience Questionnaire* has been used in Australia since 1993 to collect data from graduates about their experiences of university study. Similarly, since 2005 the *National Student Survey* has been offered to last year undergraduates in all universities and some colleges in the United Kingdom. In both instances, this institutional survey data is made publicly available via government-endorsed websites orientated to facilitating student choice.

In the case of the United States, where the higher education system is less centralized in its structure, institutional comparisons are more driven by the internal student survey data made public by individual institutions. In addition, a more sophisticated measure, the *National Survey of Student Engagement* is also offered to inform student decision making.

Therefore, there is clear evidence that student opinion data is continuing to evolve as an ever more powerful internal and external metric in higher education. Importantly, as will be detailed further in this chapter, extensive research over the last three decades has clearly demonstrated the robust validity and reliability of well designed quantitative instruments characteristically used in student rating systems (Benton and Cashin 2014; Marsh 2007). Increasingly, such research has tended to stress the need for multidimensionality in understanding the complex work of teachers in the interpretation of such outcomes.

Similarly, as the stakes surrounding improved student opinion continues to elevate, so research interest in how to more effectively address student expectations as reflected in such ratings data grows. This has characteristically centred on how student evaluation data can be most effectively assimilated into pedagogical practices of teachers, often based on situated case studies of practice investigating improved connections with professional development, mentoring or faculty engagement strategies.

However, what is notable is where the primary interest in student evaluation research has emanated: being dominated by statistical researchers (in the case of quantitative method of valid and reliable methods of student rating) and practitioners such as in-situ academic developers (in the case of enhancing responses to student opinion data). Interestingly, far less critical interest is apparent in the issue of student evaluation amongst higher education scholars, when compared to the weight of research energy dedicated to other dimensions of the higher education teaching and learning practices (such as curriculum design, pedagogical methods, learning technologies and student assessment).

Indeed—as will be argued in this chapter—although some significant reservations have been identified around the epistemic foundations of quantitative student feedback, it is polemicists rather than researchers have primarily mounted these

arguments. Research on student evaluation has tended to leave undisturbed the foundational epistemologies of quantitative ratings-based design of student feedback in its focus on statistical validity and strategies to enhance engagement with ratings data.

This lack of emphasis is somewhat more perplexing given the rising institutional and social weight of student opinion data. Specifically, contemporary student evaluation increasingly labours under the weight of several competing (and arguably conflicting) discourses. In broad terms, these discourses can be framed around two distinct and divergent motives for seeking student feedback, which themselves reflect the complex sociocultural formation of student evaluation in higher education. These distinct motives can be broadly characterised as:

- *quality enhancement of pedagogical and other practices*: reflecting the foundational professional and scholarly imperatives around student feedback to enhance the quality of higher education teaching. In this discourse, the inherent value of student feedback is toward pedagogical development (and related academic development), or other practices associated with enhancing student learning.
- *institutional quality assurance of teachers and teaching standards*: based on a largely deficit conception of teachers and teaching, student feedback is used to benchmark individual or collective teaching performance based, on internal and/or external comparators. This is primarily directed towards demonstrable shortfalls in performance requiring intervention or sanction. It also provides a metric for assessment of comparative academic performance for such things as appointment, promotion and awards.

As Walker (2001) observes, such motives are not only in inevitable tension, but also central to the formation of professional identities in the contemporary academy. The orthodox student feedback model has become normalised as a legitimate and ‘common sense’ arbiter of teaching quality. The consequence of this has been to challenge autonomous academic judgment on teaching effectiveness.

The originating quality improvement motive for student evaluation has become effectively subordinated to the more powerful demands of institutional quality assurance systems and individual performance management discourses. This has resulted in an ever more fragile settlement between these competing discourses (Kenway in Walker 2001).

Primary Research on Student Evaluation

This increasingly demanding context in which student ratings data is being used in contemporary institutional life would appear to provide fertile ground for more expansive educational research interest. Yet a review of research literature in this subject area reveals a continuing predominance of critical statistical investigations of the validity and reliability of the primary instruments employed for student rating systems internationally (Spencer and Pedhazur Schmelkin 2002).

Such research—which boasts an impressive three-decade history mirroring the rise of student evaluation systems—has most frequently centred on the confirmation or enhancement of the reliability and validity of the primary student ratings methods (Benton and Cashin 2014). This research stream has deep roots in significant formative research around student ratings and their utility. Examples of these seminal contributions include:

- Biggs and Collis' *Evaluating the Quality of Learning* (Biggs and Collis 1982) which introduced the SOLO evaluative taxonomy which introduced a measurement logic for assessing levels of student learning (and therefore teacher performativity).
- Marsh (1982a, 1987) whose research situated work pioneered the socialising of US quantitative student evaluation into Australian higher education settings.
- Ramsden (1991, 1992) who, building on the SOLO taxonomy, developed a quantitative student feedback model centred on levels of learning (which was later adapted to form the foundations for the iconic CEQ discussed earlier in this chapter).
- Centra (1993) who highlighted the significance of reflective evaluative enquiry based of quantitative student feedback.

Further, the considerations of quantitative student feedback strategies within broader academic development discourses are also relevant. Here the work of higher education researchers such as Prosser and Trigwell (1999), Toohey (1999), Laurillard (2002), Biggs and Tang (2007) are prominent.

Considerable research can be identified which is drawn from these foundational epistemologies of quantitative student feedback. Research with a focus on the usefulness or adaptation of prominent student feedback instruments (such as the widely regarded *CEQ*) is conspicuous in this research domain. Examples of that represent this research genre include Marsh (1982a), Cashin (1988), Miller (1988), Marsh and Roche (1994), Johnson (2000), Griffin et al. (2003), Dommeyer et al. (2004), Davies et al. (2009), Richardson (2005), Marsh (2007), Tucker et al. (2008), Nulty (2008) and Huxham et al. (2008). As this impressive range of research reflects, this orientation provides a substantial epistemological foundation for much of the contemporary research into approaches to deriving student feedback. A key focus of much of this research effort remains orientated toward assuring the multidimensionality of ratings scales—so as to generate a representative characterization of the broad teaching effort and to limit of the potential distorting effect of variables on outcomes (so as to limit the prospects of respondent bias). Inevitably, a matter of persistent interest is construct validity in questionnaire design (i.e. how to evaluate most effectively, instrument design and methodological adjustment). Without doubt this focus remains by far the most prominent and influential scholarship around issues related to student evaluation.

Hence, it is indisputable that extensive research has been undertaken since the originating work of important developers of the valid and reliable student rating scales (most notably Ramsden and Marsh). Indeed, it has been argued that this research interest in the technical design of student ratings—be it related to

instrument reliability or validity of ratings as a measure—has been the most researched dimension of higher education in recent history (Willits and Brennan 2015).

This strong empirical foundation has achieved much in the development of effective designs, enhanced technical precision and strategies to enhance the trustworthiness of quantitative, ratings-based forms of student evaluation. As a result, as noted in the first chapter, this research has provided the foundation for contemporary institutional (and increasingly social) confidence in the metrics generated by student ratings. This research work has been significant in securing widely held beliefs in the usefulness of student ratings as a valid form of evidence in assessing the effectiveness of teachers and their approaches to teaching. It has also been successful in both identifying and responding to some of the possible fragilities in the validity and reliability of student feedback (Richardson 2005). Such fragilities have largely centred on limiting the effects of potential subjectivities in student ratings, most notably through the introduction of elements of multi-dimensionality in ratings assessment. For instance, a variety of studies have sought to investigate and respond to potential subjectivities in student ratings outcomes where students:

- are participating in a relatively smaller class (as opposed to those in larger subjects)
- have been able to opt to participate, such as in an elective subject (as opposed to core or mandatory courses)
- are studying in relatively more accessible content areas (over those more difficult to apprehend)
- are engaged in more discussion-based subjects (compared to those involving primarily lectures or laboratory sessions)
- have a appealing academic figure, such as the experience of a favoured ‘type’ of individual (including gender, age or ethnicity) as their teacher.

Other research has sought to consider the sensitivity of student ratings to the timing of survey completion (i.e. pre/post final assessment), the context used to introduce surveys, student expectations of the eventual use of ratings data, student grade satisfaction and the level of student confidence the relevance of the instrument (Richardson 2005). The broad consensus developed from extensive field-based research is that the effect of such subjectivities in student ratings is frequently overstated, with less evidence of specific biases and more of variables necessitating specific responses to be effectively controlled (Benton and Cashin 2014; Gravestock and Gregor-Greenleaf 2008). Equally, there are a number of areas, particularly related to the effects of disciplines, teacher performativity and levels of subject difficulty where some more equivocal outcomes have emerged in research (Aleamoni 1999; Wachtel 1998). Further outcomes have suggested that such variables can be successfully mitigated, either through improved questionnaire design or enhanced methods of survey administration (Marsh 2007).

Unsurprisingly, an important emerging area of intensifying research interest within this domain is the transformation of student rating instruments from their traditional paper-based form to online completion. One critical consequence of the

move from the paper has been the spatial move from the relative control of the classroom to the more open domain of online completion. The effect of this has been to lower student response rates and potentially negatively affect the reliability of ratings outcomes (Avery et al. 2006).

Some recent research has begun to more systematically explore this issue, investigating the impact of higher levels of non-response on the validity of student ratings derived online and its potential to accentuate the influence of specific variables. For instance, research by Adams and Umbach (2012) explored the implications of these lower response rates across 22,000 undergraduates, discovering that online administration of student ratings tended to aggravate the potential for bias as poorer performing students were more likely not to participate. This is likely to be an area of intensified interest as concerns continue to mount about the potential implications of lower student response rates on the reliability—and consequently validity—of online student ratings systems.

A second important research focus in recent research around student ratings can be broadly cast as functional investigations into how to more effectively integrate or capitalise on the outcomes of ratings data. In particular, such research has tended to be situated within institutional contexts and explores strategies around how to more productively assimilate these student ratings outcomes into improved teaching practices.

Hence this research has most commonly investigated and reported on localised, experimental strategies and methods to more effectively respond to student rating outcomes. This often centres on how to strengthen student evaluation systems to ensure identified deficits are more effectively responded to, primarily through more tangible links to such things as academic or professional development strategies, provision of mentored support or via the use of specific pedagogical interventions. Another current of this research has developed around the vexed question of methods to strengthen the relationship between student feedback and faculty or institutional quality assurance mechanisms. More specifically, characteristic forms of this research has tended to broadly cluster around:

- extending the functional usefulness of student ratings by improving student evaluation systems, exemplified by the work of Martens (1999), Harvey (2003), Watson (2003), Fisher and Miller (2007), Kember and Leung (2008), Stein et al. (2012);
- better understanding of how students perceive ratings systems, evidenced in the work of Spencer and Pedhazur Schmelkin (2002);
- understanding how student rating outcomes are being used by teaching academics, such as Schmelkin et al. (1997), Moore and Koul (2005);
- strategies to more effectively link student ratings to improved teaching-teacher performance, for example Powney and Hall (1998), Ballantyne et al. (2000);
- how to harmonise student ratings outcomes with quality assurance strategies, such as Barrie et al. (2005); and
- enhancing links between student ratings outcomes and academic development, for example Arreola (2000), Smith (2008).

Hence, as this analysis demonstrates—as well as those also offered by researchers such as Richardson (2005)—the primary focus of student feedback research has remained firmly centred either on the instruments for deriving valid and reliable ratings (and particularly strategies to enhance quantitative validity) or on the effective use of the outcomes of student feedback to prospectively influence teaching practices. It remains conspicuous that there remains a relative paucity of research on the legitimacy of quantitative student feedback as a means of understanding and improving teaching pedagogies: that is *of itself*. Indeed, there appears to be almost an assumed inherent legitimacy in the student ratings method.

Similarly, there is also seemingly limited scholarly interest in how quantitative student feedback had evolved into a valid means of understanding and developing teaching. Reflecting this, it is difficult to identify significant research that analyses the sociocultural foundations of the quite specific form in which conventional student evaluation has evolved in higher education environments. Equally, there is little substantial work that critically reflects on how student ratings may actually work in practice to afford or constrain the enhancement of academic teaching. There is less interest still in the mediating effect of the student feedback on collective forms of pedagogical work in the changing realities of the contemporary university. This results in the unavoidable conclusion that the fundamental epistemological assumptions that underpin the design of quantitative student feedback models remain largely unchallenged.

This also suggests that student feedback is a matter of lesser critical interest when compared to other dimensions of the higher education teaching and learning process. It is apparent from this analysis that student feedback as a scholarly area of inquiry remains less disturbed by educational researchers than by the primarily technical drives of statisticians, institutional managers and occasional sectoral polemicists. Given the complex and multi-faceted character of the student voice in institutional life, it is difficult to understand why student evaluation has remained relatively anonymous in research when compared to other areas of scholarly inquiry in higher education such as curriculum design, pedagogical strategies, research supervision and assessment. This is all the more puzzling given its increasingly significant function in the professional assessment of academic teaching.

The Limits of Student Ratings-Based Models

So it can be reasonably asserted that the primary scholarship around student evaluation has been firmly focused on the design of student rating instruments and the more effective integration of quantitative outcomes in situated practice. Moreover, there is no doubt that considerable and impressive development work has occurred around the design and deployment of student ratings as a result. More recently, this research on student feedback has become accepted as the backbone for the comparative analysis of teaching quality within and between other related courses, and increasingly against faculty or institutional averages.

Similarly, it is certainly the case that demonstrable teaching improvements have occurred as a consequence of this determined research focus, which has improved the reliability and validity of student ratings and thereby the influence of the student voice on academic deliberations and the work of individual teachers. As this research focus continues to grow, so it will concern itself more with the troublesome issues of moving online and how to assure the continuing validity and reliability of ratings-based metrics in the face of declining levels of student response.

However, despite these realities, there remains a core unresolved epistemic question around quantitative forms of student ratings. It is a question that continues to limit its acceptance amongst many within higher education and even creates hostility toward the value and usefulness of the student voice. This question is whether the inherently reductive nature of student ratings—regardless of their demonstrable validity and reliability—can provide the critical insights necessary to drive improved teaching and student learning in the increasingly complex and multi-layered learning environments of contemporary tertiary-level learning.

As it frequently cautioned by student ratings researchers themselves, ratings can be only considered as one means of assessing the quality of teaching. As noted earlier in this book, in their emboldened contemporary manifestation student ratings are performing *of themselves* ever more powerful forms of work as a tool of individual and collective performance assessment, as well as a key indicator for institutional quality assurance. Put simply, increasingly student ratings are becoming a proxy for teaching quality.

Yet contemporary teaching—with its fragmenting certainties, heightening demands and, in most cases, diminishing resources—is less reducible than ever to inevitably one-dimensional rating scales and statistical reporting. As Barrie et al. (2008, p. 7) observe, the nature of how we chose to evaluate student learning inevitably reflects specific beliefs about ‘*what* is important to be measured, beliefs about *who* should do the measurement and what measurement might *mean*’ (original emphasis). Further, the form we adopt to understand the nature of student opinion and how it is then absorbed into the life of the academy inherently embodies a specific theory of learning and a conception of what is required (and what is not) of a teacher to afford student learning.

Therefore, it is increasingly necessary to transcend narrowing conceptions of student evaluation that are increasingly the centre of institutional conceptions of quality assurance, if we are to be serious about genuinely encouraging (and engaging with) the student voice to improve the quality of teaching, of curricula, of pedagogies, of assessment and of online technologies. As will be demonstrated through the prism of a series of case studies featured in this book, it is increasingly unlikely that conventional systems centred on student ratings will be able to achieve such an outcome. Instead, more sophisticated understandings of the student voice are necessary to both legitimately represent the depth of student opinion and to engage teaching academics in the professional dialogue necessary to genuinely lead to sustainable teaching improvement.

So this book proposes reconceptualising how we understand the student voice and the methods we use to undertake what is broadly characterized as student

evaluation. It asks us to reconsider student evaluation as a complex social activity that does considerable work in shaping teachers, teaching and courses, as well as the institutional and student sense of quality teachers and teaching. This suggests that perspectives on student evaluation must elevate beyond well-trodden debates about the nature of survey tools, statistical analysis or dissemination processes.

Moreover, as student evaluation performs ever more significant and influential functions at multiple levels in the contemporary university (and increasingly beyond), it is also time to reconsider the conflation of the discourses of quality improvement and quality enhancement within this spatial domain. The multiple purposes for which student feedback is now used mean it performs complex and heterogeneous work in the contemporary university. Some of these multiple dimensions of student feedback, and how they are manifested in contemporary higher education environments are outlined in Table 2.1.

Table 2.1 Dimensions and manifestations of student feedback in contemporary higher education

Dimensions	Manifestations
<i>Multi-leveled</i>	Student feedback is derived in both informal and formal means, as well as in formative and summative forms. This produces potential tensions between the quantified responses of students and the implicit intuitive sense of the teacher developed in the teaching environment. This creates potential tension between the relative validity and legitimacy of one form over the other
<i>Multi-charactered</i>	Student feedback is a somewhat unwelcome fringe dweller in teaching areas, being often poorly regarded, conceived of as largely ritualistic and of limited 'real' value (Anderson 2006; Edstrom 2008). Whilst at the same time, student feedback lives a regarded institutional life as a broadly reliable, robust and accountable indicator of comparative teacher quality and, by inference, student learning outcomes (Barrie et al. 2008)
<i>Multi-voiced</i>	Responses to student feedback are necessarily shaped by the differing experiences, expectations and anxieties of academics, faculties, disciplines and institutions. This means responses to student feedback cannot be considered homogenous in form and are necessarily multi-voiced. Responses are therefore a construction of differing meanings that are not necessarily shared at different levels of the institution
<i>Multi-focused</i>	The range of potential issues student feedback encounters includes such diverse objects as the teacher, pedagogical practices, student experiences, student engagement and curriculum suitability. In addition, its outcomes are also subject increasingly to broader inter and intra-comparability benchmarks of student opinion and courses. It therefore is also a measure of the relative value of individual and collective academic work
<i>Methodologically eclectic</i>	Approaches to deriving student feedback range along a continuum from highly subjective and interpretivist forms of situated judgment, to highly rationalist and abstracted quantitative surveys that rate teachers, teaching and courses.
<i>Locally mediated</i>	Forms of student feedback in higher education are locally mediated; being sociocultural constructions idiosyncratically shaped by the specific histories of student feedback models within institutions. Although this localism is in decline under the weight of standardised sectoral surveys, clear evidence of it remains (Barrie et al. 2008)

The Need to Broaden the Conventional Assumptions of Student Evaluation

Despite this burgeoning complexity, there is a relative paucity of research on alternative perspectives on the formation, use and contemporary functions of conventional quantitative student ratings. This tends to suggest that, despite its increasingly contested and disparate work in the contemporary university, student evaluation systems are generally regarded as a technical and benign (or even benevolent) in form. Schuck et al. (2008) contend this reflects the successful assimilation—and consequent heightened legitimacy—of standardised quantitative student ratings-based evaluation in contemporary higher education environments.

However, regardless of its origins, this limited breadth of research dialogue increasingly tends to limit the necessary scope of debate on this important area of higher education scholarship. Indeed, as argued earlier in this chapter, it could be reasonably asserted that student feedback (and its effect on pedagogical change) remains the least investigated element of higher education scholarship. Perhaps this is a consequence of its perceived peripheral assurance function or its low parity of esteem with other dimensions of the teaching and learning process (being consigned largely to being a ‘student’ issue). Perhaps it is the reality that the research space around evaluation has been largely and successfully occupied by statisticians and systems administrators investigating opportunities for ever-greater quantitative precision in the measurement and deployment of student opinion.

This reality is despite the rising challenges of increasingly complex environments of teaching and learning in the knowledge-technology era, where student feedback may usefully contribute greater insights to inform pedagogical decision-making. University teaching is under pressure as never before to respond effectively to the demands of more complex forms of knowledge, to abandon traditional pedagogies, to engage via multi-modal learning environments and to design relevant assessment to drive learning. All of these demands suggest an ever-greater need to understand more fully and completely the nature of student responses.

These imperatives also suggest the need to explore methods that go beyond refining traditional quantitative student feedback models to more sophisticated forms of engagement with the student voice. It is also all the more curious when considering that the outcomes of student feedback have recently become more contested within institutions, as its original quality improvement motive is challenged by the rising discourses of quality assurance, performance management and even institutional marketing. As a result, student evaluation is increasingly being called upon to do more complex work: some pedagogical, some individual, some institutional and some for the emerging student-consumer.

Moreover, in recent years, the outcomes of student evaluation are increasingly being made a public commodity, moving outside the individual professional domain of the teaching academic. This would seem to create both an imperative and a fertile space for critical research dialogue about the legitimacy of student

feedback as a measure of teacher performativity. Yet critical questions remain elusive in scholarly research, including how student feedback actually *functions* to:

- inform or debase academic judgment
- afford or hinder pedagogic change
- incite or dissuade professional development
- encourage or dissuade the development of curricula enhancement, learning activities or assessment.

Moreover, the role and function of student feedback also brings into sharper relief important tensions around teaching and learning practices. For instance, it necessarily encounters important contemporary tensions around:

- what constitutes valid knowledge about teaching and learning to frame prospective pedagogical development? (i.e. the relative rights and responsibilities of academics and/or institutions around student feedback outcomes)
- the rising uncertainties around the professional identity of teaching academics (i.e. what rights do teaching academics have to determine the suitability of ‘unpopular’ pedagogies, assessment and other practices, compared to institutions and students?)
- relative levels of autonomy of teaching academics (i.e. who interprets and initiates action on student feedback: the academic, the faculty or the institution?)
- the expected capability of the contemporary academic (i.e. how much can be reasonably expected of the teaching/research academic in response to student feedback at a time of reducing resources and elevating expectations?)

Therefore, this identified gap also became the critical foundation for framing the case study-based research reported later in this book. It motivated the specific focus on how the student voice could be further harnessed to more effectively influence and develop pedagogical practices and student learning in the ever more complex pedagogical environments of higher education.

Potential Limitations of Quantitative Student Evaluation

To consider these broader questions, it is useful to explore the arguments of those who have deviated from the dominant research discourses around student evaluation. These perspectives provide a preliminary context for the analysis that is undertaken later in this book.

There are a small but increasing number of higher education researchers and polemicists that are challenging the foundational assumptions of quantitative student feedback. This is particularly focused on whether students are able to reasonably discriminate what constitutes ‘good’ teaching, effective curriculum and approach assessment. That is, are students reasonably able to rate teachers, teaching and courses, and on what criteria is this based.

Similarly, can teaching be assumed to be 'good' if it is rated positively by students, or 'poor' if it is not rated highly? A related question is whether the 'object' perceived to be subject to evaluation (i.e. teachers and teaching approaches) is sufficiently distinct: are students evaluating the object they are assumed to be, or is it something else altogether (such as traits, environment or assessment outcomes)?

Researchers such as Schuck et al. (2008) argue that student evaluation is increasingly sustained on powerful mythologies that offer it considerable institutional credibility as a powerful demarcator of pedagogical quality. Some researchers have also raised questions about the inherently reductive nature of metric-based student feedback that is abstracted its social and individual contexts of meaning. Others have mounted research polemics to respond to what is perceived as the scepticism and disengagement by academics around quantitative student feedback. Such scepticism and disengagement is seen as arising from the inherently subjective, often inconsistent and retrospective nature of the data generated by quantitative student feedback (Edstrom 2008).

Researchers such as Johnson (2000), Kulik (2001), Kember et al. (2002), Zabaleta (2007), Schuck et al. (2008), Edstrom (2008) have identified and explored a series of other potential limitations in quantitative student evaluation models in higher education. Drawing from this collective research, a series of contestable assumptions around student evaluation can be quantified.

(a) *Higher satisfaction with teachers correlates with improved student learning*: It is axiomatic of conventional student evaluation that positive student feedback on teaching will correlate with improved student learning outcomes, yet the significance of this link is not clearly quantified in research outcomes. Although it has been more convincingly demonstrated that student-based evaluation may influence teachers to align self-perceptions with those of their students, it cannot be assumed this will actually lead to changed teaching behaviours or enhanced student learning outcomes (Richardson 2005).

Moreover, the relationship between higher evaluation ratings and higher student attainment remains contentious, with researchers such as Zabaleta (2007) failing to establish this in situated practice. Similarly precarious is the presumption that positive student feedback outcomes can be seen to equate to high quality teaching, and conversely that poor student satisfaction reflects low quality teaching. This is because the social practices of teaching are complex and often involve pedagogical or learning challenges that disrupt student certainties or confront misconceptions. Here the critical tension between the popular or charismatic teacher versus the effective or challenging educator can be complicated by student feedback. Students may understandably positively respond to the former or negatively respond to the latter. However, neither may actually provide a useful insight into pedagogical effectiveness.

(b) *Quantitatively measuring teaching quality improves student learning outcomes*: There is an implicit assumption in student evaluation models that in conducting student evaluations the quality of teaching (and therefore student learning outcomes) will demonstrably improve. This is based on the implicit belief that

individualised forms of metrics-based, deficit-focused student feedback are the most productive means of generating pedagogical engagement by teaching academics.

However, as Kember et al. (2002), Schuck et al. (2008) have demonstrated, the correlation between evaluation and improved student learning is highly dependent on the active intervention of academic development or supplementary evaluative strategies (both of which are increasingly novel in academic environments).

Although it can be reasonably argued that the assumed relationship between quality and outcomes is predicated on expectations of ancillary support—such as timely academic development support or the intrinsic motivation for promotion or other recognition—this is a difficult generalisation to sustain in the resourcing reality of the contemporary academy where evaluation data emerges largely of itself and undisturbed. Conversely, given socialised student expectations of teaching approaches, evaluation may also paradoxically act as a conservatising brake on pedagogical change for academics cautious to avoid ‘disrupted’ (and therefore dissatisfied) students evaluating teaching (Gibbs n.d.).

(c) *Quantitative student feedback is essentially an objective, benign and valid measure of teaching effectiveness:* It is anticipated in the design of evaluation systems that students can clearly and consistently rate their experiences of teaching and learning and that this in turn can provide useful and legitimate insights towards pedagogical improvement. Moreover, it is conventionally assumed that students are able to adopt a consistent and comparable rating schema in assessing things like teaching quality in their range of evaluative response. Yet these assumptions are fragile at best. A range of subjective influences have been demonstrated to manifest themselves in student ratings, which may render student evaluation in particular contexts less a barometer pedagogical quality and more an essentially environmental barometer (Schuck et al. 2008).

Moreover, Likert-type scales inherently remain essentially interpretive and intersubjective, based on students’ own definition of ‘good’ teaching, curricula and the further abstracted relationship to this imposed rating scale (Knapper 2001). There is the related assumption that no viable alternative methods of harnessing the student voice are able to improve the quality of teaching and learning as do metric-based evaluations. This is not to suggest that such ratings are simply dispensable or that they may not provide some forms of insight into sound or poor student approaches to learning. However the question is how much diagnostic value can be assumed from ratings-based, quantitative data and can this then be legitimately used to justify their literal use as a form of teaching performance indicator?

(d) *Institutional accountability improves professionalism:* Although it is inferred that accountability driven by student evaluation enhances teacher professionalism, such professional knowledge is predicated on autonomy, independence and expertise rather than compliance to an aligned to a prescribed notion of student arbitrated ‘good teaching’ (Eraut 1994).

Hence, the discourses of professionalism and accountability would appear to be in conflict where enhanced professional practice is automatically correlated with student accountability (Schuck et al. 2008).

(e) *Student feedback encourages teacher performativity*: There is a belief in the student evaluation model that correlates evaluative outcomes to improved teacher performance. Yet student feedback simultaneously responds to various institutional demands, such as those centred on quality improvement, quality assurance, performance management and institutional marketing.

When student evaluation goes beyond this first quality improvement drive, rather than achieving improved teacher performance, it may instead cultivate fear and self doubt, especially when aligned to punitive performance management or performance processes (Johnson 2000). Moreover, given the reality that university learning cannot be defined as a 'product', this approach may actually incite a perspective that the institution should provide students what they *want* as opposed to what they may actually educationally *need* (Furedi 2006).

(f) *Professionalism can be effectively codified*: Although it is generally assumed that 'good' teachers get 'good' ratings, this is based on the foundational conception that such 'good' knowledge, standards, behaviours and practices can be clearly defined, agreed and understood by respondents and readily compared. Given the contested nature of this conception, this represents a complex faith-based construction that may not be realistic, appropriate or dynamic in its form (Kulik 2001; Schuck et al. 2008).

These range of collective understandings serve to legitimise and sustain student feedback models that are now so predominant in contemporary higher education contexts. Given the extent of contestable logic, it is highly desirable that these actual foundational paradigms of student evaluation are subject to further critical scholarly enquiry. This task is all the more pressing in the transforming teaching environments, where potentially fragile and reductive data may not best serve the elevating needs for pedagogical development. This is all the more essential as student feedback is increasingly employed to inform academic performance management and frame institutional reputation.

Conclusion

This chapter provided an exploration of the dominant research literature around student evaluation, suggesting this was both substantial and at the same time clustered primarily around statistical, technical and dissemination issues. It demonstrated that the foundational epistemological assumptions that guide the conventional quantitative form of student feedback lie largely undisturbed in this research landscape. Yet there are a definable range of contestable assumptions around quantitative student feedback that deserve greater attention beyond the limited number of researchers and polemicists who are engaged in this questioning.

This imperative has grown as student feedback is increasingly used for differing (and arguably contesting) functions in the contemporary university. Student evaluation has been increasingly called upon to perform more extensive functions within and beyond institutions. As a result the actual role and function role it should performs has been subject to contestation within higher education institutions.

These rising tensions around student evaluation suggest the need for a heightened research focus on the validity of the core assumptions that sustain this orthodox quantitative approach, as well as the potential of other more expansive forms of harnessing the student voice to restore its function as an important influence in pedagogical deliberation. As the evidence in this chapter has illustrated, the contemporary work of student evaluation (and the key assumptions that underpin) can only be fully understood in the context of how it has been historically and culturally shaped. This provides a means of more effectively understanding how student feedback has contributed (or otherwise) to the formation of contemporary teaching practices. This allows the consideration of possible alternative conceptions of the use of the student voice that may resonate with the elevating pedagogical demands of increasingly complex contexts of university teaching.



<http://www.springer.com/978-3-319-41892-6>

Student Evaluation in Higher Education
Reconceptualising the Student Voice

Darwin, S.

2016, XXI, 185 p. 3 illus., Hardcover

ISBN: 978-3-319-41892-6