

# CAN YOU TEACH AN OLD DOG NEW TRICKS? RE-ENVISIONING EVALUATION PRACTICE IN HIGHER EDUCATION

*Stephen Darwin*  
*The Australian National University*

## **Abstract**

Conventional student-led evaluation has been a characteristic feature of Australian higher education for at least two decades. Increasingly it is now being used to guide major institutional decisions around educational quality, academic promotion and most recently institutional funding by government. Yet significant research around student-led evaluation has demonstrated that this form of student opinion based evaluation remains highly fragile and susceptible to multiple forms of bias. This paper argues that student led evaluation is not sufficient robust to appropriately inform notions of educational quality or pedagogical capability of academics. An alternative approach to evaluation in higher education is proposed drawing broadly on the conception of *fourth generation evaluation* (Guba & Lincoln 1989). The model foregrounds academic collaboration in shaping a professional evaluative framework at an activity level (i.e. program or sub-discipline level) based on an ongoing dialogue between peers and with students, as well as qualitative evaluation of student learning. The paper will present some preliminary research on the effectiveness of this model in its the initial piloting in postgraduate programs in an Australian university.

## **Introduction**

Evaluation based around quantitative student opinion surveys of teaching are now universal across Australian higher education institutions (Chalmers 2007; Davies, Hirschberg, Lye & Johnston 2010). Indeed, so dominant is this type of evaluation that it has now become an acclaimed and largely unchallenged orthodoxy, performing as a means of measuring teaching quality at an individual, institutional and sectoral level. Reflecting this, student led evaluation is now the foundation of institutional assessment of teaching and curriculum quality, academic merit and increasingly, government performance based funding of higher education institutions. Although informal student evaluation of teaching has no doubt as ancient origins as universities themselves, the formal collection and consideration of this feedback is a relatively recent phenomenon. In Australian higher education, its gradual emergence can be detected in the mid-1980's and its dominance in broad institutional practice established a decade later. It is now performing powerful work as a highly regarded mediator of institutional judgement and a highly regarded reductive metric for comparative academic, faculty and university performance (Davies, Hirschberg, Lye, Johnston & McDonald, 2007).

Yet student based evaluation also remains largely a fringe dweller in contemporary academic life (Edstrom 2008; Darwin 2010). The reasons for this would appear to be both pragmatic and methodological in origin. The longstanding orthodoxies of higher education are under pressure like never before and in an environment of declining public investment, elevating social expectations of demonstrable 'outcomes' from higher education and emergence of unsettling learning technologies, there is

considerable scepticism around the value of student ratings as a means of negotiating through such complexity.

Increasingly, student based evaluation is the crude mechanism that is being used to negotiate the complex higher education ecology of quality assurance, performance management and economic accountability that exist within it. However, the paradox of student feedback in this ever more consumerist environment of higher education is that despite its considerable and influential institutional power, it is widely perceived by academics to be inherently narrow and potentially superficial in analysing and responding to the complex contemporary expectations on academics of generating high quality learning for growing, heterogeneous and increasing remote student populations (Johnson 2000; Kember, Leung & Kwan 2002; Arthur 2009). Moreover, there is at best a tenuous connection between evaluation and improved student performance (Zabaleta 2007). At a broader level, student feedback-based evaluation also necessarily vacillates between the conflicting discourses of consumerist quality assurance (*what students want to receive*) and academic quality enhancement (*what students need to effectively learn*) (Bowden & Marton 1998; Walker 2001).

### **Considering student-based evaluation**

It is notable that limited research has been undertaken into the value and work of student feedback based forms of evaluation in higher education. Although extensive research has been focussed on its technical improvement or its structural integration in institutions, less is apparent on the paradigms on which it is founded and its actual impact on the improvement of teaching and learning outcomes (Darwin 2010). However, clear evidence has emerged in meta research on student evaluation that indicates a range of inherent bias in student ratings, for instance in favour of small classes over large, in elective subjects over compulsory, in accessible content areas over the difficult, in discussion based subjects over lectures and in text based over laboratory subjects (Pounder 2007; Schunk, Gordon & Buchanan 2008; Gibbs n.d.).

Several other potentially distorting influences have been demonstrated to be levels of academic charisma, gender, culture, non-verbal behaviour and levels of personal level interaction (Seldin, 1989; Schunk et al. 2008). Other research demonstrates the outcomes of evaluation are sensitive to the timing of its completion (i.e. pre/post final assessment), the contextual frame provided for it in its administering and the level of student confidence in its usefulness. Finally, there is evidence of academic disengagement with the outcomes and usefulness of the retrospective data generated by such evaluation and its emergence at the conclusion of programs (Edstrom 2008). This evidence strongly suggests that student feedback based evaluation may be much more effective as a means of 'customer' feedback rather than as data to inform professional inquiry around enhancing the quality of teaching and learning.

There are also an increasing range of higher education researchers who are disrupting orthodoxies around student-based forms of evaluation, particularly in considering whether the object perceived to be being evaluated is either clear or able to be abstracted from a broad range of other social and individual contexts of meaning. In essence, there are unresolved questions as to whether students are indeed evaluating the object they claim to (or something else altogether). Schuck, Gordon & Buchannan (2008) argue that evaluation in higher education is increasingly sustained on powerful mythologies that engender both institutional oxygen and credibility as a powerful demarcator of pedagogical quality. Other researchers including Johnson (2000), Kulik

(2001), Kember, Leung & Kwan (2002), Zabaleta (2007), Schuck et al (2008), Edstrom (2008) and Furedi (2009) further identified range of other myths that illuminate significant limitations of the largely unchallenged orthodoxies of contemporary student opinion based evaluation models in higher education. On this basis, a series of prevailing mythologies around student feedback-based evaluation can be identified. These include:

- higher teacher satisfaction equates with improved student learning
- measuring teaching quality inevitably improves student learning outcomes
- students clearly identify the object of evaluation is teaching quality
- institutional accountability improves professionalism
- student evaluation encourages teacher performativity:
- professionalism can be effectively codified in defined knowledge, standards, behaviours *and practices*

### **Foundations of an alternative model of evaluation**

However, these serious reservations about the assumptions and value of orthodox forms of quantitative based student evaluation is not meant to suggest student evaluation in higher education environments is not significant or is dispensable. Instead, it provides the basis for arguing that this dominant paradigm may not be necessarily persuasive or productive in generating and sustaining change in teaching and learning in higher education. What is being proposed is an alternative conception of evaluation based on a broadened qualitative paradigm that may be able to respond more effectively to the increasingly complex, demanding and diversifying pedagogical contexts now emerging in contemporary higher education environments. In order to engage academics, it stresses collaborative dialogue, program level (as opposed to course level) analysis and an orientation to prospective course development at one level and more focussed academic development at another.

As Lincoln and Guba (1989) argue, evaluation is less a scientific or technical process and more one that is necessarily social, political and value-orientated. Indeed, the inspiration for the design of an alternative model was drawn from their conception of *fourth generation evaluation* that went beyond measurement, description and judgment orientations to a paradigm centred on evaluation as negotiation (Lincoln and Guba 1989, p.8). This is built on the perspective that contemporary evaluation needs to be understood as ‘sense-making’ and hence a co-construction between evaluators and evaluands. It also suggests evaluation is essentially socio-cultural in its design and intent, meaning it encounters the environments of social meaning, of power and of mediation, and is shaped as well as shaping by the context in which it developed.

Finally, contemporary evaluation needs to embody a bias for negotiated action, which engages participants seamlessly in evaluation and responsiveness, defining paths forward and similarly identifying tensions, conflicts and impediments to such progress. Table One (below) summarises the primary divergences identified as part of interpreting *fourth generation evaluation* into situated evaluative practice in a higher education environment.

Table One: Comparing characteristics of standard and learning evaluation models

Orientation	Standard evaluation model	Learning Evaluation model
Form of data	Primarily quantitative	Primarily qualitative
Data sources	Student opinion	Broad range of intelligence
Method	Deficit-incident	Developmental-continuous
Primary level	Atomised - subject level	Integrated - program level
Focus	Teachers and curriculum	Student learning outcomes
Teacher role	Largely peripheral as receiver	Essential as co-constructor
Use	Remedial action	Program and academic development
Visibility	Largely private and individual	Shared and transparent
Motive	Individual legitimacy	Enhancing student learning
Dynamic	Accumulated and compared	Enacted and re-evaluated
Evaluator role	Abstracted and objective	Embedded and inter-subjective

In designing a new potential model of evaluation for higher education, these key imperatives of Lincoln and Guba's *fourth generation evaluation* were embodied in its foundations. In addition, the model needed to respond to the array of limitations that are apparent in current student opinion-led forms of evaluation. Firstly, it needed to respond to the inherent limitations of student feedback based evaluation, most notably its preoccupation with inciting students to make pseudo-professional judgements about the quality of teaching, course design and assessment. It also needed to address the static, generic and summative dimensions that were essential to the current quantitative questionnaire form and were identified as key limiters to genuine academic engagement. There also needed to be responses the deficit assumptions of teacher ratings and a clear recognition of the variability of responses from individual student cohorts to often broadly similar learning programs. Another imperative was to ensure that a new model of evaluation was able to 'close the loop' and actually influence thinking about how course could be enhanced and individual academic capability broadened with targeted professional development initiatives. Finally, there needed to be a recognition of the strong and powerful role existing quantitative student opinion-led evaluation had for teachers, institutions and, to some extent at least, students themselves.

### The Learning Evaluation Model

This deliberation resulted in the design of a new learning evaluation model that foregrounded these broad constructivist and developmental motives. The key dimensions of this model are outlined below.

a) *Focus on a conception of student learning as the object of a broader and more holistic evaluation:* The learning evaluation model moves the evaluative focus to student learning outcomes, and in doing so seeks student input on their real area of expertise: what has facilitated and impeded their own learning. Hence, it seeks to reclaim their legitimate right of academics to be subject to appropriate professional

regard that may not always be popular but instead may provide what students need, rather than merely what they personally may want (Fuerdi 2006).

*b) Evaluation based on ongoing professional collaboration, reflection and construction of shared meaning at a Program level<sup>1</sup>:* Reflecting its broad socio-cultural foundations, the evaluation model recognises that higher education is enacted in a contested and fundamentally social environment which is strongly mediated by historical approaches and its primary artefacts (most notably curriculum, learning materials and assessment). It also acknowledging that evaluation must link to the ‘particular physical, psychological, social and cultural contexts within which (it is) formed and to which (it) refers.’ (Lincoln & Guba 1989, p.8).

*c) Use of formal and informal qualitative data to inform professional debates around specific actions and program development to enhance student learning:* Evaluation in higher education is a necessarily complex activity and hence needs to adopt an expansive orientation that productively draws on a diverse range of qualitative data, some collected formally in academic discussions and student surveys is a broad activity and informal sources such as academic reflection, peer interaction, student engagement and alternative pedagogical approaches found elsewhere. As such, evaluative data is both social and transparent in form.

*d) Designed reflexively to engage with the broad history, evolution, culture and pedagogical aspirations of the program itself:* Essential to conflating evaluation and action is the need to embed the specific design of the evaluation process in the unique character of programs. Programs develop over time, accumulating defined ways of working, signature pedagogies and critical artefacts. Evaluation that is developmental orientated needs to take regard of this strong historical foundation, not least of all because of fragile academic tolerance for ungrounded change in an environment of resource decline.

*e) Melding of the academic development and evaluator roles which are enmeshed in program rather than educational discourses:* Unsurprisingly, the role of the evaluator in this learning evaluation model is not abstracted or objectified, but integrated and connected. The challenges of designing and sustaining such an evaluation approach infer a strong facilitative role by an evaluator, but this role transforms into that of a learning designer and academic developer as data is frame, collected and collaboratively considered.

### **Experiences in piloting the *Learning Evaluation* model**

The *Learning Evaluation* model has been piloted in two sites within a large Australian university, both in significant postgraduate programs. Some preliminary outcomes of one of these pilots is presented below, however longitudinal research is also currently underway that will provide further data to determine longer term effectiveness of this alternative evaluative approach. The details of these pilots are outlined in the Table below.

Table Two: Details of the Learning Evaluation model pilots

<b>Case Study One - Evaluating a maturing postgraduate program (blended mode)</b>	<b>Case Study Two - Evaluating a new postgraduate simulated professional practice program</b>
Engaged 24 teachers and 42 students across four subjects in Graduate Certificate program	Engaged 16 teachers & approximately 90 students across six subjects in professional practice element of professional studies
Key research questions defined in preliminary workshop around learning enablers, impediments, activities, assessment, feedback and general views	As the program was new and involved radical redesign, key research questions were defined as part of the implementation phase. Centred on learning enablers, impediments, simulated activities, technology, assessment
Teachers formally contributed via blog (posts n=36) and students via learning questionnaires (n=112)	Teachers formally contributed via weekly forums and students via learning questionnaires and semi structured interviews (n=129)
Data thematically coded and evaluation report produced (around 4000 words) highlighting key outcomes and program/course development issues to consider in workshop	Data thematically coded evaluation report produced (around 5000 words) and related tag cloud highlighting key outcomes and program/course development issues to consider
Two day workshop considered this formal data and informal professional reflections and benchmarked alternatives	One day workshop considered this formal data and informal professional reflections and benchmarked alternatives
Twelve substantial changes identified to be enacted , four to be progressed in other forums	Eighteen substantial changes identified to be enacted , range of others to be progressed in other forums

### **Implications for the model from the pilot**

At a methodological level, the pilots of the *Learning Evaluation* model have proven largely effective in achieving its key objectives of broadening the sources of evaluative data, re-focussing the ‘object’ of its attention onto student learning outcomes at a program level and engaging academics in collaborative professional dialogue around improvement in collective pedagogical practices. Moreover, the facilitated collection and thematic analysis of richer qualitative data (than supplied by orthodox quantitative evaluation) provided a potent and authentic framework for spirited academic debate around the tensions, conflicts and development possibilities in the programs they contribute to. Indeed, in pilots the level of post workshop debate has been expansive and largely led by academics in spirit debate, with some issues resolved and others open for subsequent discursive analysis. Student contributions, although not as extensive as would be ideal, were thoughtful and considerable. Interestingly, although students were asked to comment at a subject level in separate questionnaires, the overwhelming majority of their commentary was cast at the program level.

However, a number of limitations were also identified in this methodological stage. Initial academic involvement was reluctant and uncertain, seemingly reflecting suspicion around the motive for more expansive forms of evaluation and reluctance borne of limited capability to respond. The ongoing dialogue during programs has been labourer and under-developed, despite considerable encouragement from the facilitator of the evaluation. Reflecting this, there was only limited evidence of any changes occurring during the life of the program as a result of this (limited) collective reflection. However, it is notable that in the second cycle of this evaluative model that

is currently underway with the same group of academics, these elements of engagement have improved noticeably.

## **Conclusion**

A key thesis that underpins this proposal of a more expansive form of evaluation centred more directly on student learning is founded on the belief that more complex learning environments demand more complex forms of evaluation than that offered by orthodox student opinion based evaluation. Although this orthodox form of evaluation enjoys a high level of institutional acceptance as a reductive quality assurance mechanism, its demonstrable impact on pedagogical practice and its credibility more generally as influential on academic thinking is not well justified. More complex learning environments are demanding much more of teaching academics, melding the emergent demands of more heterogeneous student demands, multiple sites and environments of learning, the integration of learning technologies and the elevating expectations of research and service obligations. It is difficult to see how a series of student generated ratings of teachers and courses is contributing to this eclectic mission. Indeed, it seems it may instead by only making it more complex by forcing teachers to restrain pedagogical change, appeal to student sentiment or most disastrously, reduce standards.

The time has now arrived to re-consider evaluative orthodoxies in higher education. To respond to the complex ecology of higher education, academics necessarily have to become much more autonomous and engaged learning professionals, who are able to self-monitor, reflect, collaborate on current and improved practice and be subject to peer-level review and scrutiny (Eruat, 1994; Walker 2001). The initial evidence of the *Learning Evaluation* model suggests it offers a potential alternative that would contribute to this developing professional identity of the teaching academic. However, further research needs to be undertaken to judge its ability to transform conventional evaluation assumptions in the collective academic mind and sustain significant and ongoing change in individual and collective pedagogical practices that are increasingly demanded in the new complex environments of higher education.

<sup>1</sup> *Program* here is defined as a broad study program students encounter, usually defined by stage (such as first year or Graduate Certificate level) or sub-discipline area (such as core elements of a Program that are taught by largely coherent teaching teams)

## References

- Arthur, L. (2009). From performativity to professionalism: lecturers' responses to student feedback. *Teaching in Higher Education, 14*(4), 441-454.
- Bowden, J., & Marton, F. (1998). *The University of Learning: Beyond quality and competence in higher education*. London: Kogan Page.
- Chalmers, D. (2007). *A review of Australian and international quality systems and indicators of learning and teaching*. Strawberry Hills: The Carrick Institute for Teaching and Learning in Higher Education.
- Darwin, S. (2010). *Exploring critical conceptions of student led evaluation in Australian higher education*. Paper presented at the Reshaping Higher Education, 33rd HERDSA Conference, July 6-9, Melbourne.
- Davies, M., Hirschberg, J., Lye, J., Johnston, C., & McDonald, I. (2007). Systematic Influences in Teaching Evaluations: The Case For Caution. *Australian Economic Papers, 46*(1), 18-38.
- Davies, M., Hirschberg, J., Lye, J., & Johnston, C. (2010). A systematic analysis of quality of teaching surveys. *Assessment and Evaluation in Higher Education, 35*(1), 83-96.
- Edstrom, K. (2008). Doing course evaluation as if learning matters most. *Higher Education Research and Development, 27*(2), 95-106.
- Eraut, M. (1994). *Developing Professional Knowledge and Competence*. London: The Falmer Press.
- Furedi, F. (2006). *Where have all the intellectuals gone? Confronting 21st century philistinism 2nd ed*. London: Continuum.
- Furedi, F. (2009). Now is the age of the discontented. *Times Higher Education Supplement, 4 June 2009*.
- Guba, E., & Lincoln, Y. (1989). *Fourth Generation Evaluation*. Newbury Park: Sage.
- Johnson, R. (2000). The Authority of the Student Evaluation Questionnaire. *Teaching in Higher Education, 5*(4), 419-434.
- Kember, D., Leung, Y. P., & Kwan, K. P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment and Evaluation in Higher Education, 27*(5), 411-425.
- Kulik, J. (2001). Student Ratings: Validity, Utility and Controversy. *New Directions for Institutional Research, 109*(Spring 2001).
- Pounder, J. (2007). Is student evaluation of teaching worthwhile? An analytical framework for answering the question. *Quality Assurance in Education 15*(2), 178-191.
- Schuck, S., Gordon, S., & Buchanan, J. (2008). What are we missing here? Problematising wisdoms on teaching quality and professionalism in higher education. *Teaching in Higher Education, 13*(5), 537-547.
- Seldin, P. (1989). Using student feedback to improve teaching. *New Directions for Teaching and Learning, 37*, 89-97.
- Walker, M. (2001). Mapping our Higher Education Project. In M. Walker (Ed.), *Reconstructing Professionalism in University Teaching*. Buckingham: SHRE/Open University Press.
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education, 12*(1), 55-76.