

Evaluation of teaching

Denise Chalmers* & Lynne Hunt

University of Western Australia, Australia

This review focuses on evaluating university teaching from the perspective of teachers and the context of their teaching, arguing that evaluation of university teaching should be part of a reflective cycle, informed by evidence that leads to: improved opportunities for students' learning; enhanced curricula; and career development. It reviews four broad sources of evidence that provide robust and valid information that enables teachers to develop their skills and practices to support their students' learning and to demonstrate their professional practice to the wider academic community. The underlying premise of this review is that all sources of evidence should be used holistically to provide the richest possible picture to capture the different aspects of university teaching including the context, the processes and the outcomes. It raises questions about: what should be evaluated; the reliability of tools and information; and the potential for teachers to be de-professionalised if they do not take ownership of their evaluation. It concludes that we cannot keep locking our doors saying that teaching is all an alchemy or mystery. We know what good teaching looks like and, in the interests of our students and our own careers and professionalism as teachers, we can reasonably ask if we are doing it.

Keywords: *University teaching evaluation; Sources of evidence; professionalising university teaching*

“Shall we evaluate teachers and teaching?” Simpson asked in 1967 (p. 286). It’s a question that wouldn’t be asked today because, as Simpson observed at the time, “Whether we like it or not, teacher evaluation is [already] an existing subjective phenomenon in every college and university.” In retrospect, his paper makes instructive reading because it begins from the assumption that “the overriding purpose of teacher evaluation is to improve the judgments and decisions of those concerned with teaching.” (p. 286).

* Email: denise.chalmers@uwa.edu.au

This review of the evaluation of university teaching and learning starts from the same position. It focuses on evaluating teaching from the perspective of teachers and their teaching, arguing that evidence arising from the evaluation of university teaching and learning should be part of a reflective cycle that leads to: improved opportunities for students' learning; enhanced curricula; and career development.

At about the same time as Simpson asked his question, Hilderbrand, Wilson & Dienst, (1971, p. 5) argued that reliable and valid forms of evaluation of teaching that inform teachers of their effectiveness must be used in promotion decisions. They contended that this is the single, most important requirement for the "improvement of university teaching" (Hilderbrand et al., 1971, p. 5). While the early focus of evaluation of teaching was on enhancing teacher effectiveness and career progression, its scope has now expanded beyond individual teachers to encompass the broader teaching context of courses, departments, institutions and beyond.

This paper reviews four broad sources of evidence that provide useful and valid information that enables teachers to develop their skills and practices better to support the development of their students' learning and to demonstrate their professional practice to the wider academic community. The underlying premise of this review is that all sources of evidence should be used holistically to provide the richest possible picture to capture the many different aspects of university teaching including the context, the processes and the outcomes. This multifaceted and multi-sourced view of teaching evaluation is critical, for there is no one way to carry out valid and reliable evaluation (Theall, 2010). Likewise, there is no one source of evidence that should be weighted or privileged over others.

I. Background

Simpson identified three main purposes for the evaluation of university teachers and teaching. These were to:

- 1) redirect the future teaching behaviour of faculty members, using evaluative data to modify what is taught or how it is taught;
- 2) assess past performance to help in judgments relating to promotion in rank and pay; and
- 3) assess the overall success of the total institutional program.
(Simpson, 1967, 286-7)

These purposes of evaluation remain current today. The major change is that they have been extended to the evaluation of teaching beyond

institutions to sector-wide, national and international levels. Between the 1960s and 1990s, institutions were considered responsible for the evaluation of teaching. However, in the 1990s, governments and their agencies started to take an interest in the quality of teaching in universities (Chalmers, 2007; Krause, 2012). Government quality assurance agencies emerged to review teaching and student outcomes in universities. For example, in 1992, the UK instituted the Higher Education Quality Council, which became the Quality Assurance Agency for Higher Education (QAA) in 1997. Similarly the Australian Government established the Committee for Quality Assurance in Higher Education in 1992. This was followed by the Australian Universities Quality Agency (AUQA) and later the Tertiary Education Quality and Standards Agency (TEQSA). This move towards external evaluation, within and between universities, included the introduction of standards for higher education, and public reporting of teaching and student outcomes (Chalmers, 2008).

The trend to increased governmental interest in university teaching, and the associated emergence of external agencies to carry out evaluation arose from greater student numbers, known as the ‘massification’ of higher education and the consequential increased government investment in tertiary education. Further, higher education became a significant export earner. In Australia, for example, higher education earned a record \$17.6 billion in 2014, making it Australia’s fourth largest export (Department of Education and Training, 2015). Governments want to ensure the quality of the education provided and protect both the international and domestic markets by demonstrating that institutions meet the highest international standards. In addition, all students, whether domestic or international, want value for their investment of time and money. In brief, they all want to know that universities teach well and provide high quality, relevant courses.

The management and processes of university administration changed in response to the greater external scrutiny and accountability required of them, and to the need to teach increasing numbers of students. However, while university administrations and governments forged ahead in response to ever greater calls for accountability and evaluation, academic staff largely resisted. There was a clash of cultures between traditional notions of the university and corporate managerial processes of which the requirements for evaluation were seen to be a part. For example, Barnett (2013, p. 41) observed that the culture of universities lives partly “in the imagination, in the ideas, sentiments, values and beliefs that individuals hold in relation to the university ... For individuals do not just think the university but they *embody* the university”.

The administrative requirement that different forms of evaluation be undertaken in such fluid, values-based and culturally diverse university settings inevitably gave rise to concerns among academics and contributed to a loss of trust in university management. These required forms of evaluation continue to be seen as surveillance and a threat to academic freedom. In this context, the multiple purposes of the evaluation of teaching became mired. This jeopardises the reflective cycle of quality improvement, which informs this review paper, because reflective quality improvement requires the active engagement of university teachers in developing students' learning, enhancing curricula, and achieving career advancement based on teaching excellence.

Reflective and meaningful cycles of teaching improvements have also been disrupted by changes in the academic workforce since the 1990s surge in the establishment of university teaching evaluation processes. Today, there are decreasing numbers of tenured academic staff and increasing numbers in teaching-only and part-time or casual teaching positions (Kezar & Holcombe, 2015; Pricewaterhouse Coopers, 2016) to the extent that up to 80% of undergraduate teaching may be carried out by non-traditional academics in many Australian institutions. The quality of teaching in universities is now largely in the hands of teachers who may have little or no access to professional development or continuity of employment – let alone career progression. Casual teachers typically have little engagement in decisions related to curriculum design or teaching practices, and have limited, ongoing engagement with students (Chalmers et al., 2003; Harvey et al., 2014; Kezar & Holcombe, 2015). This can result in tenured staff being called on to respond to students with whom they have had little contact, and with students coping with multiple teachers and, potentially, an overall fractured learning experience.

Some universities have begun to re-define their academic workforce into research-intensive and teaching-intensive positions in addition to traditional teaching-research positions (Probert, 2013). Simultaneously, they are establishing processes that rely on the systematic evaluation of all who are engaged in teaching, regardless of their role as professional or academic staff. Increasingly, a solution is seen to be in establishing more teaching-intensive positions and ensuring there is an enhanced provision of professional development for all who contribute to the teaching endeavour. This has been described as the professionalization of teaching. The Office for Learning and Teaching (OLT), in Australia, commissioned two projects to develop this concept (Chalmers et al., 2014, 2015; James et al., 2015). These trends point to one of the original purposes of evaluation, namely: to demonstrate professionalism in teaching. Hence there is a window of

opportunity for university administrations to work in partnership with teachers (whatever the terms of their employment) to establish evaluation processes that will support teachers’ development and career progression.

2. Sources of information for evaluation

In the 1960s, Simpson (1967) argued in favour of self-evaluation to inform teacher development. This included written self-assessments, reviewing students’ achievements in and out of college, as well as input from students and work colleagues. Later, Hildebrandt et al. (1971) argued for the widespread use of surveys completed by students and peers to provide comparative information to inform judgements associated with career advancement. Berk (2005, 2014) has long advocated for the use of multiple strategies, drawing on multiple sources, to inform on teaching effectiveness. He lamented that there has been little uptake of many of them. Smith (2008) subsequently synthesised much of this earlier work in a 4-Quadrant (4Q) approach to evaluation, to inform teachers and professional development programs.

Braskamp (2000) argues for a holistic approach noting that we need to consider the “teacher and teaching, and the learner (student) and learning” (p. 20) as the building blocks for evaluation. These four building blocks were adopted as the key elements of the *Dimensions of Teaching Quality Framework* (Chalmers, 2007, p. 99) because they ensure that university teachers are understood in the context in which they work. They are part of an enterprise with colleagues, and a core component of an institution in assuring quality. In similar vein, students are learners engaged in the enterprise of learning that takes place in a complex and multifaceted context. These same building blocks frame the following discussion of the sources of information for evaluation (Figure 1).

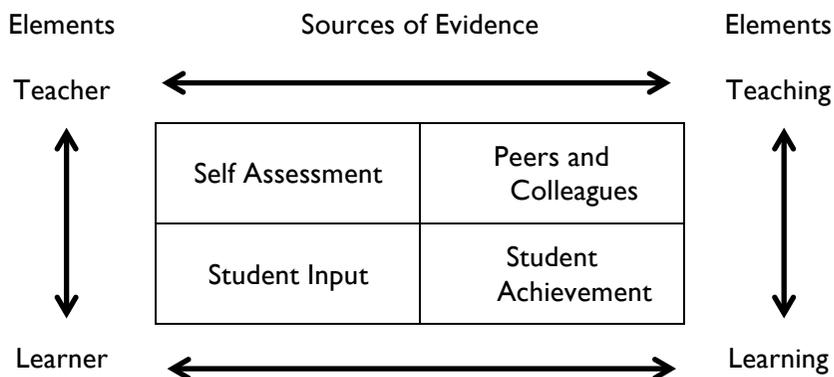


Figure 1: Critical elements and sources of evidence for evaluating teaching

The four sources of evidence—student input, student achievement, peers and colleagues and self-assessment—have been identified as appropriate for the evaluation of teaching for more than half a century, and considerable work has been devoted to identifying tools and processes to collect information. However, there remains ongoing debate among teachers and administrators about their validity, reliability and suitability in providing useful information to inform the evaluation of university teachers and their teaching.

We argue that all four sources of evidence should be used holistically to provide the richest possible picture to capture the many different aspects of university teaching including the context, the processes and the outcomes. This multifaceted and multi-sourced view of evaluation of teaching is critical if we are to carry out valid and reliable evaluation (Theall, 2010).

Input from students

Input from students refers to matters such as student feedback, perception surveys, and a range of formal and informal information that is collected from students about their experiences of learning and teaching.

Student feedback surveys

Historically, student feedback surveys in Australia have been used to provide information to individual teachers for developmental, appraisal and accountability purposes (Moses, 1988). Prior to the mid-1980s, individual teachers and departments sought student input through administering their own surveys. However, these were not particularly trusted, nor were they systematically administered (Moses, 1988). In the late 1980s, Australian universities began to establish whole-of-university approaches to student feedback, normally administered through central university departments. The University of Sydney's 'Students' Evaluation of Educational Quality' (SEEQ), developed by Marsh in 1984, and the University of Queensland's student feedback survey, were highly influential across Australia (Davies et al., 2009) in terms of the questions, structure, and process of administration¹.

Surprisingly little has changed. Current student feedback surveys of teaching retain a similar structure and focus across most universities. Two reviews of Australian university student feedback surveys (Barrie et al., 2008; Davies et al., 2009) found that all Australian universities had an established student feedback survey about teaching, with the data used



primarily to inform the individual teacher's improvement and development. In addition, at the teachers' discretion, it could be used as a source of evidence for performance review and promotion. Initially, many universities restricted access to teaching data, reserving it solely for the personal use of individual teachers. This was largely in response to the distrust of teachers regarding the use of this information for management purposes. The strategy of limited access also responded to doubts about students' capacity to judge teaching. Restricted access to student feedback about teaching is now lessening because of the greater centralised management of data and the growth of teaching and learning analytics (Alderman et al., 2012; Siemens et al., 2013).

Today, student feedback surveys are routinely administered centrally and through learning management systems. Utilisation of standard communication processes is particularly important with the increase in casual and part-time teachers (Kezar & Holcombe, 2015; Pricewaterhouse Coopers, 2016). Although this increases opportunities for students to provide feedback on all subjects in which they have enrolled, it can potentially disconnect the immediacy of teachers seeking feedback and responding to it. This does not have to be the case—as demonstrated by Hunt and Sankey (2013)—with the thoughtful use of learning management systems and appropriate processes that close the loop on evaluation by providing feedback to all students.

In the 90 years since the publication of the first report on student feedback of teaching by Remmers and Brandenburg in 1927, there have been several thousands of research studies about them. A common conclusion is that student feedback is most effective when it is used for: reflection and improvement by the teachers involved in teaching the subject; for improving the program of study by the course coordinators, and for reporting back to the students in a timely manner (Nair et al., 2008). Comprehensive and rigorous research and meta-reviews of these thousands of studies (See Abrami, et al., 2007; Marsh, 2007; Spooren et al., 2013) conclude that student evaluation of teaching is:

- ▶ multi-dimensional
- ▶ reliable and stable
- ▶ primarily a function of the instructor who teaches a course rather than the course that is taught
- ▶ relatively valid against a variety of indicators of effective teaching
- ▶ relatively unaffected by a variety of variables hypothesised as potential biases

- ▶ seen to be useful by faculty as feedback about their teaching, by students for its use in course selection and by administrators for use in personnel decisions (Marsh, 2007, p. 319)

Abrami's et al. (2007) review of student feedback concluded that, while student ratings should not be used indiscriminately for summative decisions about teaching effectiveness, particularly when drawing on specific teaching dimensions, global items have demonstrated high validity and can be confidently used for summative decisions (p. 432). It is, therefore, perplexing that a significant number of teachers and administrators remain sceptical of the validity of student evaluations of teaching, particularly as universities are communities of scholars educated to rely on evidence-led conclusions. This is especially the case when it is used to inform decisions about teaching performance and career progression (Surgenor, 2013). With such a significant body of research, why does commentary about the biases and unreliability of student feedback persist? It would appear that the majority of university administrators and academics remain unaware of the massive body of research and substantiating evidence (Surgenor, 2013).

Outside of individual institutions, the usefulness of information gathered from student feedback surveys is limited. As Barrie et al. (2008, p.103) noted, such surveys "have remained idiosyncratic institutional practices, developed within universities and operating independently of any national system and usually without reference to each other". In a subsequent review, Alderman et al. (2012, p. 271) agree, "Australian universities continue to develop individual approaches, which demonstrate considerable variation in question topics, wording and rating scales and ways the information is gathered, interpreted and acted on". As a consequence, the use of surveys outside of the institution is limited if the intention is to benchmark and compare with other institutions. Alderman et al. (2012) argue for the need for universities to work cooperatively together to develop a framework for evaluation in which "a valid, reliable, multidimensional and useful student feedback survey constitutes just one part of an overarching approach to the evaluation of teaching. Such an approach can inform not only the teacher but disciplines, institutions and allow for benchmarking across institutions" (Alderman et al., 2012, p. 271). National student feedback surveys have been developed to fill this void, for example, the Course Experience Questionnaire (CEQ) in Australia has been administered nationally to all graduating students since 1993, and the University Experience Survey (UES) administered to enrolled university students since 2012. International examples are the National Student Survey (NSS) in the United Kingdom and the National Survey of Student

Engagement Survey (NSSE) in the United States of America (Chalmers, 2007; Wilson et al., 1997)

Informal feedback from students

There are many informal sources of student feedback in addition to standard university surveys including: teacher-administered student surveys on student engagement (standardised and/or teacher generated); responses to in-class polling or anonymous feedback on a particular teaching approach undertaken through the semester; and unsolicited feedback from students (both positive and negative). More formally, the teacher can seek input from students through group interviews, or by encouraging students to nominate a class representative and regularly seeking their comments. Alternatively, student opinion can be gleaned from focus groups on a particular aspect of teaching or the curriculum, or by setting up a system of anonymous student feedback throughout the subject, and scanning student discussion, logs and journals. These provide different opportunities for students to reflect on their experiences of learning and for teachers to reflect on their teaching practices and inform their development as teachers.

Some institutions restrict input from students to institutional surveys only, or they place greater weight on formal surveys in the mistaken belief that these are the only valid and reliable source of student input. This does a disservice to teachers and their students. Formal and informal inputs from students are all valid sources of information that add richness to standard survey data and should rightfully be included in multi-source, multi-dimensional evaluations of university teaching. However, student feedback cannot provide an assessment of teaching content, pedagogical practice or ethical standards of practice—these dimensions of teaching are best assessed by peers and colleagues.

Peers and colleagues

Half a century ago, Simpson (1967) and Hildebrand (1971), identified peers and colleagues as legitimate and valuable sources of information about teachers and teaching. Despite this long period of advocacy, peer assessment for both formative and summative purposes was largely neglected until the mid-1990s. Shulman (1993) noted that, unlike communities of research, there is no comparable community of teachers within which ideas and experiences of university teaching could be exchanged:



We close the classroom door and experience pedagogical solitude, whereas in our life as scholars, we are members of active communities: communities of conversation, communities of evaluation, communities in which we gather with others in our invisible colleges to exchange our findings, and methods and our excuses. (Shulman, 1999, p. 24)

He argued that we: “need scholarship that makes our work public and susceptible to critique. It then becomes community property, available for others to build upon” (Shulman 1999, p. 16). Atwood et al. (2000) agree, noting that teaching should be like any other scholarly activity—a shared community property. Their argument is that the driver of peer review is not associated with teaching deficiencies. Rather, it is associated with the invisibility of teaching. They conclude that the responsibility to document, share, and seek critique and feedback that is taken for granted in research work, is not evident in university teaching.

The reluctance to accept the value and legitimacy of peer review of teaching remains: “It is a remarkable feature of higher education—until recently—that the processes relating to teaching and learning have not traditionally been subject to formal processes of peer review” (Gosling, 2014, p. 14).

That peers should be considered as a source for the evaluation of teaching has been as contentious as using input from students to evaluate university teaching. A common objection is that peer review is a form of surveillance that risks undermining academic freedom. This goes to the heart of the different purposes of peer review. For example, Chism (2007, p. 5) distinguished between peer review that provides formative evaluations through which teachers are provided with information used to improve teaching. This can be offered confidentially and may be “informal, ongoing, and wide-ranging”. In contrast, summative evaluations are used to make personnel decisions associated with hiring, promotion, and merit pay. Informal evaluations by peers, while not necessarily welcomed, are generally considered to be more acceptable than summative peer evaluations. Despite academics’ reservations about peer review, Braskamp (2000) argues that peers must be a part of both formative and summative evaluation of teaching because

only faculty collectively have the experience and standards of scholarship that are both credible and useful to individual faculty. Peer evaluation does not violate academic freedom, but instead provides one of its best

defences. Judgment by peers is essential if faculty desire to enjoy their relative autonomy and self-governance. (p. 27)

Braskamp makes a key point in regard to the professionalization of teaching because peer review is a core academic process. Failure to open up the academy of teaching risks the very autonomy and self-governance to which academics appeal as a reason to resist the evaluation of their teaching.

Sachs and Parcell (2014) describe the dilemma faced by institutions that want to incorporate peer review of teaching for routine use for both development and judgement purposes.

An informal chat over coffee between colleagues on improving assessment tasks, engaging large cohorts or innovative approaches to online pedagogy is clearly not an opportunity for management to monitor or control staff. It is only ever systematically supported peer review that can sensibly be perceived as a potential instrument of managerial control. This creates something of a dilemma. On the one hand, formal, systemically supported peer review can be viewed suspiciously leading to academic staff only engaging at the most perfunctory levels. The result is a system that fails to support enhancement. On the other hand, highly informal peer review is by its very nature almost impossible for universities to systematically support. The result is piecemeal improvements that are unsustainable. (p. 3)

Despite the fact that peer review is an effective way to provide feedback on learning and teaching and a useful strategy for academic development (Bell, 2005; Bell & Cooper, 2013; Nash et al., 2014), it is not widespread in Australian universities, nor is it systematically fostered or supported by policy frameworks (Harris et al., 2008). It is not considered to be a valuable development activity in which to engage, despite the evidence that it has been shown to be effective in enhancing the quality of teaching (D'Andrea & Gosling, 2005).

Peer review of teaching extends beyond classroom observations to different aspects of teaching. For example, the University of Western Australia recommends 10 dimensions of teaching as suitable for peer review, only one of which is observation of classroom practice. The other aspects of teaching include course and unit content; teaching and learning strategies; learning materials and resources; assessment practices; management; leadership roles; evaluation of teaching; scholarship related to teaching; and postgraduate supervision.

A teaching excellence framework developed to clarify expectations and types of evidence for teachers at different levels of appointment (Chalmers et al., 2014) identifies several aspects of teaching that are suitable for peer review. At the first levels of appointment, such as associate lecturer and lecturer, teachers might be expected to provide evidence that they have sought and responded to formative peer feedback to demonstrate that they are developing their teaching practices and appropriately supporting their students' learning. This formative/developmental engagement in peer review would normally be documented and reported in a teaching portfolio to inform the performance review process. It is predominantly at the higher levels of appointment that teachers would seek to be reviewed for a summative evaluation of different aspects of their teaching when applying for a promotion or teaching award. The teaching portfolio or dossier developed to document teachers' evidence and achievements itself relies on peer evaluation for its assessment (Hubball & Clarke, 2011).

Increasingly, peer review is required by institutions to support claims for teaching quality in performance review and applications for promotion. Where teachers must be peer reviewed, it is typically through a process of performance appraisal where supervisors monitor and provide feedback on staff performance (Blackmore & Wilson, 2005). This method of appraisal is described by Gosling as the "managerial approach" (Gosling, 2002). This process tends to be built around performance checklists, peer review of materials, accessibility of materials and availability of facilities for student interaction (McKenzie, 2011). It can involve classroom visits from senior academic managers, quality auditing teams or academic developers (Hammersley-Fletcher & Orsmond, 2004; McKenzie, 2011). Unsurprisingly, such top-down peer observation of classroom teaching behaviours has been found to increase anxiety (Bell, 2005)—described as a process "to be endured" (Blackmore et al., 2005, p. 227).

Several Australian national peer review projects have been supported over the past 15 years (Harris et al., 2008; Nash, et al., 2014; Sachs et al., 2013). Mostly, these have focused on peer review for teacher development and enhancement within an institution. A further two projects focussed on peer review for enhancement in online teaching contexts within and across institutions (McKenzie et al., 2011; Sachs et al., 2014). Crisp et al. (2009) trialled: *Peer Review of Teaching for Promotion Purposes* across four universities. It considered both intra-institutional teacher observation and external peer review of documented evidence. They concluded that "summative peer review of teaching has the ability to improve both the status and the quality of teaching at tertiary level, by encouraging the promotion of exceptional teachers and academics engaged in the scholarship

of teaching at all levels” (2009, p. 5). They recommended that “for a summative peer review of teaching program to be successful, peer reviewers must be trained and experienced” (2009, p. 5).

In Australia, Curtin and RMIT universities have established a pool of trained peer reviewers who can be called on to provide informal, formative peer review for developmental purposes and, at the teachers’ instigation, summative review. Initiatives such as these provide both institutional management and teachers with confidence in the peer review process and in the information it provides for developmental and judgemental purposes within an institution. However, to date, there has been little progress made in establishing and training a pool of teaching and learning experts external to the university, who can be called on to review teaching portfolios against institutional or external criteria for performance review and promotion purposes. The National Teaching Fellowship program of activities is now addressing this issue (Chalmers, 2016). The Higher Education Academy (HEA) Fellowship recognition program in England already provides an example of external peer review against a national standard of teaching.

Peer-review models and practices are not readily transferrable between institutions. They really need to be tailored to the needs and circumstances of particular institutions, disciplines, and curricula contexts. If they are to be successful, they must be fully supported by leadership at all levels of the institution, informed by a scholarly, developmental approach, integrated within a broader context of institutional and program-level reform initiatives, and recognised as providing critical, valid evidence for administrative decision-making about the effectiveness of teaching practices for tenure, promotion and/or teaching award adjudications (Hubball & Clarke 2011, p. 24).

Successful peer review models incorporate a focus that extends beyond individual teachers to include team and collaborative teaching in a range of teaching contexts (tutorial, laboratory, practicum, field, studio, on-line) as well as expanding the range of dimensions for review beyond classroom observation. Additionally, they incorporate casual and contract teachers in formative and developmental processes that recognise their value and contribution to the institution and to students’ learning.

2.3 Student Achievement

The evaluation and benchmarking of student achievement falls into the learner-learning dimensions of Braskamp’s model (2000). The inclusion of student achievement as a measure of teaching effectiveness began seriously



with the paradigm shift in the mid-1990s when the focus moved from what the teacher was doing to what the student was doing (Barr & Tagg, 1995). This was facilitated by different ways of gathering assessment data on student learning (Angelo & Cross, 1993) and has gained impetus from contemporary trends in learning analytics. However, the validity of some of the proxies used in learning analytics might be questioned in terms of the extent to which they are representative measures of students' learning outcomes as distinct from student activity.

What teachers can claim as evidence of student learning is confounded by issues of students' motivation, engagement and ability which are such influential factors in learning outcomes. Whilst it might seem reasonable to ask, "If there is no learning, has there been any teaching?" (Theall 2010, p. 86), in evaluative terms it raises the question whether students' failure to learn can be attributed to their teachers? Conversely, if students demonstrate outstanding learning, can that be attributed to teachers? According to Trigwell (2001, p. 65), there is a rich body of research which shows that good teaching is related to high quality student learning. He notes that the extent to which teachers and teaching behaviours are orientated to student learning can be evidenced and documented. For example, a student learning orientation is identifiable by the way in which teachers, or teaching teams, plan the course, design the assessment, provide feedback, and structure the program of learning.

The evidence of teachers' orientation to student learning is verifiable by peers, student input and self-assessment. The triangulation of these different sources of evidence for a student orientation provides valid indications of teachers' contributions to students' achievements. Claims for contribution require moderation in terms of the level of responsibility a particular teacher has for the course and curriculum, and for the setting and provision of feedback on the assessment tasks. Further moderating factors might include the extent to which a teacher either lectures and tutors or only lectures. Another factor might include whether a teacher manages tutors and provides oversight of their teaching or whether she or he develops and manages significant online learning experiences.

Different aspects of student achievement might be claimed as evidence of success by the teacher or the teaching team depending on their role and responsibilities in the oversight, design and management of a course or program of study. These might include:

- ▶ *Student progression.* The proportion of students who progress from first to second year, or from third year into honours or postgraduate study, may be an indicator of student achievement. If benchmarked

over time or with comparable courses, any improvements can provide evidence of teaching effectiveness. Similarly, evidence may be gleaned from the outcomes of any special interventions to enhance student progression. Reducing course failure rates is another example of evidence, but this needs to be treated cautiously, because it might be claimed that standards of achievement have been lowered in order to improve progression rates.

- ▶ *Student employment outcomes.* Teachers can claim evidence of successful graduate employment rates when a course has instituted a work integrated learning program if they have: designed a course that develops graduate capabilities; built relationships with employers and provided strong links to employers for their students; high levels of student employment following graduation; or if they have evidence that the course is well regarded by employers for the quality of their students. The Graduate Destinations Survey, which forms part of the Australian Graduate Survey, provides a nationally recognised source of information about employment outcomes.
- ▶ *Externally verified learning achievements:* Universities have a long tradition of using external examiners to appraise postgraduate theses and dissertations. Their comments can provide evidence of success for supervisors. Similarly, workplace supervisors of students on work placements or practica can provide feedback on the readiness of students for engagement in the placement or on the skills they have demonstrated. Evidence of successful teaching and supervision is strengthened if records of student achievements can be demonstrated over time and with a number of students.
- ▶ *Student grades:* Changes in grades attained by students may be evidence of teaching quality, particularly if these are sustained over time and involve some process of moderation or verification.
- ▶ *Graduate attributes:* Attainment of graduate attributes or capabilities can be identified through student self-reporting, and the CEQ teaching scale. For other strategies see the Assuring Graduate Capabilities website (Oliver, nd).
- ▶ *Student self-reported learning gains:* Students can reliably report on aspects their learning through self-assessment, learning journals or surveys in which students rate their learning gains and knowledge (Ross, 2006; Falchikov, 2005; Falchikov & Boud, 1989).
- ▶ *Evidence of learning:* Evidence of changes in learning or standards of learning can be gathered through the use of pre-post testing of

students' knowledge at the beginning and end of a course. There is an extensive range of classroom assessment techniques (Angelo and Cross, 1993) that provide evidence of learning to teachers and students.

- ▶ *Criterion referenced assessments, the SOLO taxonomy* (Biggs, 1999). Student's learning can be transparently and independently assessed and verified (Biggs & Collis, 1982) in standards or criterion referenced assessment, often supported by the use of explicit criteria rubrics, or a taxonomy, such as the SOLO taxonomy, which refers to the Structure of the Observed Learning Outcomes (see Biggs & Collis for more details).
- ▶ *Moderation and verification.* In Australia, the moderation and verification of assessment and grades is gaining momentum (Booth et al., 2015; Krause et al., 2013). Where standards of student work have been moderated and/or verified by others, then examples of student work that demonstrate different levels of achievement can be used to demonstrate standards set for student learning and grades achieved.
- ▶ *Surveys.* Questionnaires such as the Student Approaches to Study Questionnaire (Biggs, 1987), student engagement questionnaires (Krause & Coates, 2008), or Student Attitudes to Learning (Prosser & Trigwell, 1999) provide reliable indications of the ways students engage with their learning that have a relationship with the quality of outcomes and achievements.
- ▶ *Learning analytics.* The increasing use of learning management systems, online assessment and learning tools, and in-class response technologies, has enabled the tracking of learning through the smallest units of keystrokes while students are working on a task. This offers opportunities to aggregate data from multiple sources and to create data sets that can be interrogated to identify patterns of learning behaviours on a large scale. Given that learning analytics may provide robust evidence of student learning, it is likely that these will be used routinely in the future to inform teachers about their teaching and students about their learning, particularly for the evaluation of large scale online courses and Massive Online Open Courses (MOOCs).

These examples of student achievement, though by no means exhaustive, illustrate a range of sources on which to draw, from inside and outside the course, that provide evidence of university teachers' contribution to supporting their students' learning and achievements.

2.4 Self assessment

Light et al. (2009) note that reflection about learning and teaching is essential if we are to move beyond traditional, laissez-faire versions of the benign amateur teacher of the past. Reflection also counteracts any tendencies towards instrumental, behavioural-competence approaches to the evaluation of teaching. Critical reflection involves asking questions about disciplinary and professional roles as well as responsibilities in the university and more broadly in the higher education sector (p. 18). Typical self-assessment questions might include: “What professional responsibilities do I have for student learning? How or in what ways am I accountable for my teaching? What responsibility do I have for my growth as a teacher? What are the critical challenges of learning in my discipline?” (Light et al., 2009, p. 18). The answers to these questions can provide evidence that distinguishes excellent teachers (Kane et al., 2004; Trigwell, 2001). Evidence of self-assessment can include systematic documentation of:

- ▶ *Leadership roles*: This describes innovative practice, as well as support and mentoring for colleagues and sessional teachers and tutors (See Blackmore 2012).
- ▶ *Reflective Course Memos*: These can show, for example, the extent to which particular teaching strategies addressed students’ learning needs and the changes that were made in response to feedback.
- ▶ *Evaluation strategies*: This includes the strategies and responses to the outcomes of peer review, student feedback, and professional and industry feedback.
- ▶ *The scholarship of teaching and learning*: Publications and presentations about teaching which includes evidence of teachers adopting “scholarly, inquiring, reflecting, peer reviewing, student-centred approaches to teaching [as they are] likely to be achieving the purpose of improving student learning” (Trigwell, 2013, p. 102; Trigwell, 2012).
- ▶ *Purposive engagement in professional development*: The key word here is purposive for it provides a description of professional development undertaken with a rationale. Of particular interest is the development of activities undertaken in response to an issue, an interest or a personal development plan to enhance teaching and student learning.
- ▶ *Teaching practice inventories*: These are well-researched inventories for individual, or teams of, teachers to complete on their teaching

behaviours. They provide teachers feedback on the extent that the strategies and behaviours support student learning (See Wieman, 2015; Prosser & Trigwell, 1999; Hammersley-Fletcher & Orsmond, & Pratt, 2011).

Teaching portfolios

Teaching portfolios synthesise evidence that teachers have collected, drawing on the four sources related to their pedagogy, practices, planning, reflection, responses and development to illustrate claims of achievements. They can include, but are not limited to:

- ▶ *Teaching philosophy statement*: This is a reflective statement about how you teach and why, what you value in the teaching and learning process, what kind of teaching takes place in your classroom – and why, what guides your decision-making in curriculum, and how you have developed as a teacher over time.
- ▶ *Description of context, institutional and individual goals*: This description includes reference to institutional values and goals and the teachers' contribution towards these (Braskamp, 2000) as well as the extent, range, and duration of teaching.
- ▶ *Responses to institutional teaching criteria*: Typically, criteria are framed around good teaching principles. For example, the Australian University Teaching Criteria and Standards (AUTCAS) (Chalmers et al., 2014) framework draws on principles of teaching practices that focus on student engagement and learning synthesized into seven criteria. These are:
 - 1) Design and planning of learning activities
 - 2) Teaching and supporting student learning
 - 3) Assessment and giving feedback to students on their learning
 - 4) Developing effective learning environments, student support and guidance
 - 5) Integration of scholarship, research and professional activities with teaching and in support of student learning
 - 6) Evaluation of practice and continuing professional development
 - 7) Professional and personal effectiveness

The AUTCAS framework assists teachers in the development of portfolios by providing examples of practice for each criterion, and

Figure 2: AUTCAS Criterion 1. Design and planning of learning activities illustrating multiple sources of evidence

Criterion 1: Design and planning of learning activities				
Planning, development and preparation of learning activities, learning resources and materials for a unit, course or degree program; including coordination, involvement or leadership in curriculum design and development				
Lecturer (A)	Lecturer (B)	Senior Lecturer (C)	A/Professor (D)	Professor (E)
<p>Planned learning activities designed to develop the students' learning</p> <p>Sound knowledge of the unit content and material</p> <p>Unit outline that clearly details learning outcomes, teaching and learning activities and assessment</p> <p>Preparation of unit materials</p> <p>Peer review of unit materials by unit/course coordinator</p> <p>For relevant items in the student survey, average or above average scores for all units taught e.g.</p> <ul style="list-style-type: none"> • Appropriate teaching techniques are used by the teacher to enhance my learning. • The teacher is well prepared. • The teacher effectively used learning technologies to support my learning 	<p>Deep knowledge of the discipline area</p> <p>Well planned learning activities designed to develop the students learning</p> <p>Scholarly/informed approach to learning design</p> <p>Thorough knowledge of the unit material and its contribution in the course</p> <p>Effective and appropriate use of learning technologies</p> <p>Effective unit/ course coordination</p> <p>Effective preparation of tutors and management of teaching teams</p> <p>Peer review of unit materials by course coordinator</p> <p>For relevant items in the student survey, average or above average scores for two consecutive years and in all units taught</p>	<p>Meets the requirements for level B and</p> <p>Deep knowledge of the discipline area</p> <p>Innovation in the design of teaching, including use of learning technologies</p> <p>Effective preparation of tutors and teaching teams</p> <p>Leadership in curriculum development and design.</p> <p>Development of significant curriculum materials</p> <p>Benchmarking of a unit or course against similar units/courses</p>	<p>Meets the requirements for Level C and</p> <p>Leadership in effective curriculum development at a program level</p> <p>Contribution to the teaching or curriculum and/or discipline at a national level</p> <p>External expert peer review of unit/course materials/ curriculum/initiative curriculum</p> <p>Adoption of learning materials by other universities</p> <p>Nomination for a teaching award for curriculum contribution</p>	<p>Meets the requirements for Level D and</p> <p>Leadership role and impact in curriculum design and review, planning and/or development at a (inter)national level</p> <p>Significant curriculum or disciplinary contribution through published student learning materials/textbooks</p> <p>Leadership in mentoring and supporting colleagues in planning and designing learning activities and curriculum</p>
Indicative evidence: Unit/course outline and materials				

identifies clearly stated expectations of levels of performance and sources of evidence that could be used to demonstrate that an individual academic meets that criterion. (See figure 2). The evidence identified under each criterion draw from the four sources of evidence described in this paper.

- ▶ *Leadership in teaching and summary of overall achievements:* Teaching portfolios ideally close with an opportunity to highlight teachers' contributions to leadership and innovation, concluding with a summary of achievements and highlights.

There is now a growing body of work focusing on the assessment of teaching portfolios, particularly in relation to judgements for awards, promotion and performance review for teachers and teaching teams (Buckeridge, 2008; Trivett & Stocks, 2012). These studies have shown that sound judgements can be made by internal and external peer reviewers on the evidence presented by teachers in their portfolios.

Criteria and standards frameworks such as the AUTCAS have been criticised as leading to a check-box response from teachers (Probert, 2015). Yet, when teachers do make a tokenistic checkbox response and engage minimally with the indicators, it becomes immediately apparent in the documentation and review process. There are clear qualitative and quantitative differences between these portfolios and those in which teachers have presented their thinking and demonstrated the quality of their teaching through highlighting the multidimensional aspects of their work (Trigwell, 2001).

So far in this review, the focus has been on the four sources of evidence used to evaluate university teachers and their teaching. The purpose of these evaluation processes is to demonstrate the professionalism of university teachers and their teaching so that institutions can be confident in using the information to recognise and reward teachers throughout their careers and to inform the review of courses and programs of study.

3. Evidencing teaching - beyond the institution

The four sources of evidence can also be used to evaluate teaching (rather than teachers) at the levels of courses, schools or whole of institutions, but also across the sector—nationally and internationally. The different types of indicators and purposes for which they are used have been detailed in a number of reports (Chalmers 2007, 2008, 2010; Chalmers, Lee and Walker, 2008; Department for Business, Innovation & Skills, 2015). The further from the actual context of teaching that they have been drawn and aggregated,

the more they can only be considered as proxies for evaluating teaching quality. This is not a problem as long as it is understood that they provide a general indication or overview of the quality of teaching and at the broadest levels. The following section provides a brief review of the utilisation of the four sources of evidence to inform the quality of teaching beyond the institution.

Student input

The primary sources of evidence for student input beyond the teacher are from student feedback surveys. These are typically sourced from aggregated student feedback surveys, some collected at course level, and others collected at the discipline and institutional level. For example, the Course Experience Questionnaire (CEQ) collects data from all graduating students from universities, and increasingly, private higher education institutions in Australia. The UK equivalent is the National Student Survey (NSS). The University Experience Survey (UES) now seeks Australian student satisfaction of their experiences while enrolled in universities in their first and final year of undergraduate study. While these can be inappropriately used to compare disciplines and institutions, they are, nevertheless, a widely accepted indicator that utilises student input as evidence of teaching quality across the sector. The Quality University Learning and Teaching Indicators (QUILT) report these data on a publically accessible website in Australia and the MultiRank website provides similar information on programs of study across the majority of universities in Europe.

Peer and colleagues

Peers and colleagues have a widely accepted and respected role in the evaluation of teaching beyond each institution. This can be through review panels of courses, organisational units and institutions for quality reviews organised by quality and accreditation agencies, or through review of grants and awards offered by the Office of Learning and Teaching (OLT) and its antecedent organisations. External expert peer review of promotion applications, and publications are regularly used.

Student achievement

Student achievement is typically sourced from institutional completion, progression and attrition rates. Graduate destination surveys and starting salary data (both UK and Australia have a long history of collecting this



information) are used to indicate student achievement. Other examples of information gathered and reported about students' achievements include the widespread use in the USA of the College Learning Assessment (CLA) tests and tracking the proportion of students passing graduate and professional admissions tests. In the UK, external assessors provide a moderation role on assessment tasks as well as of student grades and achievement.

Self-assessment

Self-assessment plays a large part in evidencing teaching in quality reviews. It serves as a precursor to a peer review process for external quality review processes and for benchmarking across disciplines and institutions. Self-assessment and benchmarking rely on agreed standards and transparency. The Australian Higher Education Standards (Tertiary Education Quality and Standards Agency, 2015) are used by all Australian higher education institutions to guide their quality processes and are reviewed by TEQSA. The international standards for teaching and program quality in business schools is EQUIS (a European base accreditation agency) and the Accreditation Council for Business Schools and Programs (ACBS), a US based international accreditation agency. Both agencies utilise self-assessment and peer review to make judgements for accreditation. Further, self-assessment reports are routinely provided to the Australian government providing data and reports on aspect and achievements for indigenous education, diversity and staffing profiles, and international education activity. Self-assessment and reporting is the most commonly utilised quality model of evaluation internationally (Chalmers, Lee & Walker, 2008).

4. Conclusion

This review has explored the evaluation of university teaching from the perspective of teachers. It has described four sources of evidence: student input, peer review, student achievement and self assessment. While each may have weaknesses, used in combination of two or more sources, the strengths are enhanced and the weaknesses reduced. In brief, evidence from a number of sources currently provides the best opportunity for teachers to demonstrate teaching outcomes through formative and summative evaluations. The underlying premise of this holistic approach is that multiple sources of information provide the richest possible picture that describes the context, processes and outcomes of teaching.

In making an assessment of the future of the evaluation of university teaching we need to look to the past to see what has worked and what



hasn't. The question is, has the evaluation of teaching made a difference? Some say no. Newton (2002, p. 42), for example, suggests that the connection between data collection and subsequent quality enhancement measures has been lost: "The dons have outsmarted the government by turning the exercise into a game and playing it brilliantly". The evaluation game can lead to tokenism, reputation management and image control. Much of this can be attributed to the resistance of teachers and institutions to the external, even standardised, evaluation of what is seen to be their domain of responsibility. This sentiment is captured by Johnson (1994, p. 379)

It no longer really matters how well an academic teaches and whether or not he or she sometimes inspires their pupils; it is far more important that they have produced plans for their courses, bibliographies, outlines of this, that and the other, in short all the paraphernalia of futile bureaucratisation required for assessors who come from on high like emissaries from Kafka's castle.

Newton (2002, p. 41) worries that the culture of evaluation may have resulted in a quality assessment "sickness" characterised by "low trust, high accountability, competing models of quality and external quality monitoring, un-costed extra workload, and the absence of any real evaluation of the evaluators". In contrast, this review has argued that if professional, university teachers wish to be fully recognized and rewarded for the quality of their teaching (Chalmers, 2011b), they now have an opportunity to embrace and shape local policies and plans for the evaluation of teaching using the four broad sources of evidence that have been shown to provide valid information that substantiates claims to effective teaching and professional practice in local, national and international contexts.

The landscape of quality assurance and evaluation in universities is changing and evaluation processes may need to adjust accordingly in the future. For example, the Pricewaterhouse Coopers report (2016), on the changing nature of the Australian Higher Education workforce, makes it clear that increasing numbers of teachers are being employed on limited term contracts without traditional career progression opportunities. What investment will teachers on such contracts have in the outcomes of teaching evaluations on which they cannot act? As to the future, as Braskamp (2000 p. 30) noted:

The challenge is to figure out how to evaluate faculty within this new context, a context filled with external accountability demands, technological advances, student

needs often at variance with the lifestyle and values of the faculty, and diversity of institutional missions. In short, the future of students must be taken into account, and faculty cannot assume any longer that only they know what is best for the student.

In conclusion, this review of the evaluation of university teaching has raised questions about: what should be evaluated; the reliability of tools and information; and the potential for teachers to be de-professionalised if they do not take ownership of their evaluation. The argument is that we cannot keep locking our doors saying that the evaluation of teaching is all an alchemy or mystery. We know what good teaching looks like and, in the interests of our students and our own careers and professionalism as teachers, we can reasonably ask if we are doing it.

5. Notes

1. Chalmers (2011a) and Alderman et al. (2012) provide an overview of the development of student feedback surveys in the Australian national and university contexts.

6. Acknowledgements

Funding for a number of the project reports cited in this paper authored by Chalmers and colleagues has been provided by the Office for Learning and Teaching and its antecedents. The views in this paper do not necessarily reflect the views of the Australian Government Office for Learning and Teaching.

7. References

- Abrami, P., d'Apollonia, S. & Rosenfield, S. (2007). The dimensionality of student ratings of instruction: An update on what we know, do not know, and need to do. In R. P. Perry, & J. C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. (pp. 385-445). New York, NY: Springer.
- Alderman, L., Towers, S., & Bannah, S. (2012). Student feedback systems in higher education: A focused literature review and environmental scan. *Quality in Higher Education*, 18 (3), 261-280.
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for faculty*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.

- Atwood, C.H., Taylor, J.W., & Hutchings P.A. (2000). Why are chemists and other scientists afraid of the peer review of teaching? *Journal of Chemical Education*, 77 (2), 239-243.
- Barnett, R. (2103). *Imagining the university*. London: Routledge.
- Barr, R.B., & Tagg, J. (1995). From teaching to learning—a new paradigm for undergraduate education. *Change: The Magazine of Higher Learning*, 27(6), 12–26.
- Barrie, S., Ginns, P., & Symons, R. (2008). *Student surveys on teaching and learning*. Sydney, NSW: Australian Learning & Teaching Council. Retrieved from http://www.olt.gov.au/system/files/resources/Student_Surveys_on_Teaching_and_Learning.pdf
- Bell, M. (2005). *Peer observation partnerships in higher education*. Milperra, NSW: Higher Education Research & Development Society of Australasia.
- Bell, M., & Cooper, P. (2013). Peer observation of teaching in university departments: A framework for implementation. *International Journal for Academic Development*, 18(1), 60-73
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17 (1), 48-62.
- Berk, R. A. (2014). Should student outcomes be used to evaluate teaching? *The Journal of Faculty Development*, 28(2), 87-96. Retrieved from <http://search.proquest.com/docview/1673849348?accountid=14681>
- Biggs, J. (1987). *Study process questionnaire manual. Student approaches to learning and studying*. Hawthorn, VIC: Australian Council for Educational Research
- Biggs, J. (1999). *Teaching for quality learning at university*. Buckingham, UK: Open University Press.
- Biggs, J., & Collis. K.F. (1982). Evaluating the quality of learning: The SOLO Taxonomy (Structure of the observed learning outcome). New York, NY: Academic Press.
- Blackmore, P. (2012). Leadership in teaching. In L. Hunt, & D. Chalmers, (Eds), *University teaching in focus: A learning-centred approach* (pp. 268-281). Melbourne, VIC: ACER Press.
- Blackmore, P., & Wilson, A. (2005). Problems in staff and educational development leadership: Solving, framing, and avoiding. *International Journal for Academic Development*, 10(2), 107-123.
- Booth, S., Beckett, B., Saunders, C., Freeman, M., Alexander, H., Oliver, R., Thompson, M., Fernandez, J., Valore, R. (2015). Peer review of assessment networks: Sector -wide options for assuring and calibrating achievement standards within and across disciplines and other networks. Sydney, NSW: Office for Learning and Teaching. Retrieved from http://www.olt.gov.au/system/files/resources/Booth_Report_2015.pdf
- Braskamp, L. (2000). Toward a more holistic approach to assessing faculty as teachers. *New Directions for Teaching and Learning*, 83, Fall, 19-32.
- Buckridge, M. (2008). Teaching portfolios: their role in teaching and learning policy. *International Journal for Academic Development*, 13 (2), 117-127.
- Chalmers, D. (2007). *A review of Australian and international quality systems and indicators of learning and teaching*. Sydney, NSW: Australian Learning and

- Teaching Council. Retrieved from http://www.olt.gov.au/system/files/resources/T%2526L_Quality_Systems_and_Indicators.pdf
- Chalmers, D. (2008). *Indicators of university teaching and learning quality*. Sydney, NSW: Learning and Teaching Council. Retrieved from http://www.olt.gov.au/system/files/resources/Indicators_of_University_Teaching_and_Learning_Quality.pdf
- Chalmers, D. (2010). National teaching quality indicators project- final report. Rewarding and recognising quality teaching in higher education through systematic implementation of indicators and metrics on teaching and teacher effectiveness. Sydney, NSW: Australian Learning and Teaching Council. Retrieved from http://www.olt.gov.au/system/files/resources/TQI_Final_Report_2010.pdf
- Chalmers, D. (2011a). Student feedback in the Australian national and university context. In P. Mertova & S. Nair, (Eds), *Student feedback: The cornerstone to an effective quality assurance system in higher education* (pp. 81-97). Cambridge: Woodhead Publishing.
- Chalmers, D. (2011b) Progress and challenges to the recognition and reward of the scholarship of teaching in higher education. *Higher Education Research and Development Journal*, 30(1), 25-38.
- Chalmers, D., Cummings, R., Elliott, S., Stony, S., Tucker, B., Wicking, R., & Jorre de St Jorre, T. (2014). *Australian university teaching criteria and standards project. Final report*. Sydney, NSW: Office for Learning and Teaching. Retrieved from http://www.olt.gov.au/system/files/resources/SPI2_2335_Cummings_Report_2014.pdf
- Chalmers, D., Cummings, R., Elliott, S., Stony, S., Tucker, B., Wicking, R., & Jorre de St Jorre, T., (2015) *Australian university teaching criteria and standards project. Extension project report*. Sydney, NSW: Office for Learning and Teaching. Retrieved from http://www.olt.gov.au/system/files/resources/SPI2_2335_Cummings_report_2015.pdf
- Chalmers, D., Herbert, D., Hannam, R. Smeal, G., & Whelan, K. (2003). *Training support and management of sessional teaching staff*. Sydney, NSW: Learning and Teaching Council. Retrieved from <http://www.olt.gov.au/system/files/resources/sessional-teaching-report.pdf>
- Chalmers, D., Lee, K., & Walker, B. (2008). *International and national quality teaching and learning performance models currently in use*. Sydney, NSW: Office for Learning and Teaching. Retrieved from <http://www.olt.gov.au/resource-rewarding-and-recognising-quality-teaching>
- Chalmers, D. (2016). *Recognising and rewarding teaching: Australian teaching standards and expert peer review*. Sydney, NSW: Office of Learning and Teaching. Retrieved from <http://www.olt.gov.au/olt-national-senior-teaching-fellow-denise-chalmers>
- Chism, N.V.N. (2007). *Peer review of teaching: A sourcebook* (2nd ed.). Bolton: Anker.
- Collins, J.B., & Pratt, D. (2011). The teaching perspectives inventory at 10 Years and 100,000 respondents: Reliability and validity of a teacher self-report inventory. *Adult Education Quarterly*, 61(4), 358-375.
- Crisp, G., Sadler, R., Krause, K., Buckridge, M., Wills, S., Brown, C., McLean, J., Dalton, H., Le Lievre, K., & Brougham, B. (2009). *Peer review of teaching for*

- promotion purposes: A project to develop and implement a pilot program of peer review of teaching at four Australian universities.* Sydney, NSW: Office for Learning and Teaching. Retrieved from <http://www.olt.gov.au/resource-peer-review-teaching-adelaide-2009>
- D'Andrea, V. & Gosling, D. (2005). *Improving teaching and learning in higher education: A whole institution approach.* Maidenhead, UK: Open University Press.
- Davies, M., Hirschberg, J., Lye, J., & Johnston, C. (2009). A systematic analysis of quality of teaching surveys. *Assessment and Evaluation in Higher Education*, 35(1), 83-96.
- Department for Business, Innovation & Skills. (2015). *Higher education: teaching excellence, social mobility, and student choice.*, London, UK: Author. Retrieved from <https://www.gov.uk/government/consultations/higher-education-teaching-excellence-social-mobility-and-student-choice>
- Department of Education and Training. (2015). *Education exports hit a record \$17.6 billion* [Media Release]. Retrieved from <https://ministers.education.gov.au/pyne/education-exports-hit-record-176-billion>
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4), 395-430.
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions of aiding learning in higher and further education.* London: RoutledgeFalmer.
- Gosling, D. (2002). *Models of peer review of teaching.* York, UK: Learning Teaching Support Network (LTSN) Generic Centre.
- Gosling, D. (2014). Collaborative peer supported review of teaching. In J. Sachs & M. Parsell, M. (Eds.), *Peer review of learning and teaching in higher education: International perspectives* (pp. 13-31). Dordrecht: Springer.
- Hammersley-Fletcher, L., & Orsmond, P. (2004). Evaluating our peers: Is peer observation a meaningful process? *Studies in Higher Education*, 29(4), 489-503.
- Harris, K-L., Farrell, K., Bell, M., Devlin, M., & James, R. (2008). *Peer review of teaching in Australian higher education: A handbook to support institutions in developing an embedding effective policies and practices.* Sydney, NSW: Office for Learning and Teaching. Retrieved from <http://www.olt.gov.au/resource-peer-review-of-teaching-melbourne-2009>
- Harvey, M., Luzia, K., McCormack, C., Brown, N., McKenzie, J., & Parker, N. (2014) *The BLASST report: Benchmarking leadership and advancement of standards for sessional teaching.* Sydney, NSW: Office for Learning and Teaching. Retrieved from <http://blasst.edu.au/>
- Hildebrand, M., Wilson, R.C., & Dienst, E.R. (1971). *Evaluating university teaching.* Berkeley, CA: Centre for Research and Development in Higher Education.
- Hubball, H., & Clarke, A. (2011). Scholarly approaches to peer review of teaching: Emergent frameworks and outcomes in a research-intensive university. *Transformative Dialogues: Teaching & Learning Journal*, 4 (3), 1-32.
- Hunt, L. & Sankey, M. (2013). Getting the context right for quality teaching and learning. In D. Salter, (Ed.), *Cases on quality teaching practice in the social sciences* (pp. 261-279). Hershey, PA: IGI Global.

- James, R., Baik, C., Millar, V., Naylor, R., Bexley, E., Kennedy, G., Krause, K., Hughes-Warrington, M., Sadler, D., Booth, S., & Booth, C. (2015). *Advancing the quality and status of teaching in Australian higher education*. Sydney, NSW: Office for Learning and Teaching. Retrieved from http://www.olt.gov.au/system/files/resources/SPI2_2330_James_Report_2015.pdf
- Johnson, N. (1994). Dons in decline. *Twentieth Century British History*, 5(3), 370- 385.
- Kane, R., Sandretto, S., & Heath, C. (2004). An investigation into excellent tertiary teaching: Emphasising reflective practice. *Higher Education*, 47, 283-310.
- Krause, K. (2012). A quality approach to university teaching. In L. Hunt, & D. Chalmers (Eds.), *University teaching in focus: A learning-centred approach* (pp. 235-252). Melbourne, Australia: ACER Press.
- Krause, K., & Coates, H. (2008) Students' engagement in first-year university. *Assessment & Evaluation in Higher Education*, 33(5), 493-505.
- Krause, K., Scott, G., Aubin, K., Alexander, H., Angelo, T., Campbell, S., Carroll, M., Deane, E., Nulty, D., Pattison, P., Probert, B., Sachs, J., Solomonides, I., Vaughan, S. (2013). *Assuring final year subject and program achievement standards through inter-university peer review and moderation*. Retrieved from <http://www.uws.edu.au/latstandards>
- Light, G., Cox, R., & Calkins, S. (2009). *Learning and teaching in higher education: The reflective professional* (2nd ed.). Thousand Oaks, CA: Sage.
- Marsh, H. W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry, & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective*. (319-383). New York, NY: Springer.
- McKenzie, J. & Parker, N. (2011). *Peer review in online and blended learning environments*. Sydney, NSW: Australian Learning and Teaching Council. Retrieved from <http://www.uts.edu.au/sites/default/files/final-report.pdf>
- Moses, I. (1988). *Academic staff evaluation and development: A university case study*. Brisbane, QLD: University of Queensland Press.
- Nair, C.S., Adams, P., & Mertova, P. (2008). Student engagement: The key to improving survey response rates, *Quality in Higher Education*, 14(3), 225-232.
- Nash, R., & Barnard, A. (2014). *Developing a culture of peer review of teaching through a distributive leadership approach*. Sydney, NSW: Office for Learning and Teaching. Retrieved from http://www.olt.gov.au/system/files/resources/LE11_1980_Nash_Report_2014.pdf
- Nash, R., Bolt, S., Barnard, A., Rochester, S. Mcevoy, K., Shannon, S., & Philip, R. (2014). *Developing a culture of peer review of teaching through a distributive leadership approach*. Sydney, NSW: Office for Learning and Teaching. Retrieved from <http://www.olt.gov.au/project-developing-culture-peer-review-teaching-through-distributive-leadership-approach-2011>
- Newton, J. (2002). Views from below: Academics coping with quality. *Quality in Higher Education*, 8(1), 39-61.
- Oliver, B. (nd). *Assuring Graduate Capabilities*. Retrieved from <http://www.assuringgraduatecapabilities.com/>

- Pricewaterhouse Coopers. (2016). *Australian higher education workforce of the future*. Retrieved from <http://www.aheia.edu.au/news/higher-education-workforce-of-the-future-167>
- Probert, B. (2013). *Teaching-focused academic appointments in Australian universities: Recognition, specialisation, or stratification*. Sydney, NSW: Office for Learning and Teaching. Retrieved from <http://www.olt.gov.au/resource-teaching-focused-academic-appointment>.
- Probert, B. (2015). *The quality of Australia's higher education system: How might it be defined, improved and assured*. Sydney, NSW: Office for Learning and Teaching. Retrieved from <http://www.olt.gov.au/resource-why-scholarship-matters-higher-education-2014>.
- Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education*. Buckingham, UK: SRHE and Open University Press.
- Remmers, H. H., & Brandenburg, G. C. (1927). Experimental data on the Purdue rating scale for instructors. *Educational Administration and Supervision*, 13(6), 399-406.
- Ross, J.A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation*, 11(10), 1-13.
- Sach, J., & Parsell, M. (2014) Introduction: The place of peer review in learning and teaching. In J. Sachs & M. Parsell (Eds), *Peer review in learning and teaching in higher education: International perspectives*. Springer, Dordrecht. p.1-12.
- Sachs, J., Parsell, M., Ambler, T., Cassidy, S., Homewood, J., Solomonides, I., Wood, L., Wynn, L., Jacenyik-Trawöger, C., Probert, B., Lyons, J., Akesson, E., Irhammar, M., Brokop, S., Kilfoil, W., du Toit, P. (2013). *Social, communicative and interpersonal leadership in the context of peer review*. Sydney, NSW: Office for Learning and Teaching. Retrieved from <http://www.olt.gov.au/project-social-communicative-interpersonal-leadership-macquarie-2009>
- Shulman, L. (1993). Teaching as community property: Putting an end to pedagogical solitude. *Change: The Magazine of Higher Learning*, 25,(6), 6-7.
- Shulman, L. (1999). Taking learning seriously. *Change: The Magazine of Higher Learning*, 31(4), 11-17.
- Siemens, G., Dawson, S., & Lynch, G. (2013). Improving the quality and productivity of the higher education sector: Policy and strategy for systems-level deployment of learning analytics. Sydney, NSW: Office for Learning and Teaching. Retrieved from http://www.olt.gov.au/system/files/resources/SoLAR_Report_2014.pdf
- Simpson, R.H. (1967). Evaluation of college teachers and teaching, *Journal of Farm Economics*, 49(1) Part 2, 286-298.
- Smith, C. (2008). Building effectiveness in teaching through targeted evaluation and response: connecting evaluation to teaching improvement in higher education. *Assessment & Evaluation in Higher Education*, 33(5), 517-533.
- Spooren, P., Brock, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 18(4), 598-642.

- Surgenor, P. (2013). Obstacles and opportunities: Addressing the growing pains of summative student evaluation of teaching. *Assessment & Evaluation in Higher Education*, 38(3), 363–376.
- Tertiary Education Quality and Standards Agency. (2015). *TEQSA contextual overview of the new HES Framework*. Melbourne, VIC: Author. Retrieved from <http://www.teqsa.gov.au/teqsa-contextual-overview-hes-framework>.
- Theall, M. (2010). Evaluating teaching: From reliability to accountability. *New Directions for Teaching and Learning*, 123, Fall, 85-95.
- Trevitt, C., & Stocks, C. (2012). Signifying authenticity in academic practice: a framework for better understanding and harnessing portfolio assessment. *Assessment & Evaluation in Higher Education*, 37(2), 245–257.
- Trigwell, K. (2013). Evidence of the impact of scholarship of teaching and learning purposes. *Teaching & Learning Inquiry*, 1(1), 95-105.
- Trigwell, K. (2001). Judging university teaching, *International Journal for Academic Development*, 6(1), 65-73.
- Trigwell, K. (2012). Scholarship of teaching and learning. In L. Hunt & D. Chalmers (Eds.), *University teaching in focus: a learning-centred approach* (pp. 253-267) Melbourne, VIC: ACER Press.
- Wieman, C. (2015). A better way to evaluate undergraduate teaching. *Change: The Magazine of Higher Learning*, 47(1), 6-15. DOI: 10.1080/00091383.2015.996077
- Wilson, K.L., Lizzio, A., & Ramsden, P. (1997). The development, validation and application of the Course Experience Questionnaire. *Studies in Higher Education*, 22(1), 33-53.
- Wood, D, Scutter, S., & Wache, D., (2011). *Peer Review of Online Learning and Teaching University of South Australia*. Sydney, NSW: Office for Learning and Teaching. Retrieved from http://www.communitywebs.org/peer_review/