University of Victoria
Division of Learning and Teaching Support and Innovation (LTSI)

Teaching Excellence: A Briefing Paper

February 14, 2019

Prepared by Lesley Scott, MA, PhD(c), Research Associate, LTSI
in consultation with Dr. Laurene Sheilds, Executive Director

Teaching Excellence: A Briefing Paper

Executive Summary

Overview

This paper reviews the concept of teaching excellence in both global and local terms, including an examination of the many factors that influence the definition of teaching excellence, and the measurement of teaching effectiveness. A singular definition is difficult to ascertain, rather, teaching excellence is understood through multiple lenses including attributes of students, priorities, context and purpose. An extensive literature search was conducted to investigate two questions: What is teaching excellence within the post-secondary environment and secondly, how is teaching excellence most effectively measured?

Teaching Excellence

Teaching excellence can best be described as a set of practices designed to maximize (increase the likelihood of) student learning, and can best be recognized as a scholarly endeavour founded upon scholarship of teaching and learning research. Teaching excellence, further, is not at all distinct from either good teaching or quality teaching, but can be described as on a continuum, and is easily conflated in contexts where a competitive teaching excellence award is not the explicit goal.

Faculty, student and theoretical understandings of the highest standard in teaching in higher education, exhibit multiple common expectations and criteria. Effective teaching practices are frequently identified as including:

- strong and supportive student-teacher relations,
- student-centered pedagogies, such as highly-interactive or collaborative teaching methods; pedagogies which similarly ensure student engagement;
- expert and inspiring knowledge of one's discipline and subject;
- strong organizational skills, evident in class structure;
- strong explanatory skills, evident in clarity and student learning;
- appropriate assessment and timely provision of feedback
- commitment to personal pedagogic self-reflection and professional development.

Also, commonly cited in research on teaching excellence, is an emphasis on the scholarship of teaching and learning as foundational to quality teaching practices. Current research further reaffirms long-standing pedagogical emphasis and research on student learning outcomes, and the increasing importance of student learning gain as a measure of teaching effectiveness. There is finally, a focus on the importance of student and peer feedback for teaching enhancement, with additional considerations for innovation and the worthwhile integration of technology.

More recent models of teaching excellence increasingly move beyond the personal attributes of the teacher, to consider the individual teacher's impact on the wider learning environment, as in fostering wider pedagogic expertise, mentorship and innovation, or contributions to strategic institutional goals. Teaching excellence is supported through: the use of teaching awards and incentives; the formalization of institutional teaching excellence and evaluation policies; recognition and incentives involving salary, merit and promotion policies (teaching professors, teaching fellows, etc.); and the targeting of teaching through significant, and/or specific, award and grant initiatives.

Measurement of Teaching Effectiveness

The measurement of teaching continues to be one of the most noted contemporary challenges in the field of teaching and learning in higher education. The development of a shared understanding of what is actually meant by teaching excellence is an important starting place in the articulation of effective measures. Within this paper, the term *teaching excellence* itself was not singled out as always distinct in meaning from *good*, or *quality* teaching, but the concept can be considered, rather, through the broader lens of *teaching effectiveness* as a term representative of the most valued and highest levels of pedagogic expertise.

Teaching effectiveness (like teaching excellence) is considered difficult to measure with most scholars pointing to the importance of using multiple measures, including quantitative metrics (often in the form of student feedback), and qualitative measures such as self-reflection/evaluation, trained peer review, teaching dossiers, and evidence of student learning itself. The point is that a combination of multiple sources of quantitative and qualitative is better than simply quantitative alone or a singular source. In fact, the literature is clear, that a well-designed and broad set of measures will provide the best assessment of teaching effectiveness, identifying strengths and areas for improvement. The teaching dossier is repeatedly described as the most effective strategy through which teaching effectiveness can be evidenced and teaching enhancement goals identified.

Ideally, measures of teaching effectiveness must reflect student learning and be founded in the knowledge of best teaching practices. This encompasses the pedagogical practices that foster student engagement and maximize student achieved learning outcomes. While student feedback on teaching may be one component to be considered, there are multiple others sources of feedback including alumni ratings, employer ratings, administrator ratings, teaching scholarship, teaching awards, achieved learning outcomes, and more comprehensively, teaching portfolios. To further increase validity of teaching measures, different stakeholders ideally would also contribute to the shared understanding of good teaching practices. Notably peer review of teaching was viewed with some caution with scholars recommending that it is best undertaken by trained peer reviewers or educational experts.

Student evaluations of teaching (SETs) continue to be the most contested aspect of evaluation. While all scholars addressed the importance and essential nature of student feedback, they also consistently make clear that SETs should never be used as a sole measure of teaching effectiveness, and that the emphasis of SETs should be on improvement of teaching, rather than assessment for merit or promotion purposes. Further, when SETs are used, they should be designed by experts with knowledge of both effective teaching practices and instrument development (validity and reliability), and be used within the context of a multi-faceted evaluation. Strategies to maximize response rates are imperative to the effective use of SETs, particularly in the online environment. At the same time, it is broadly recognized that student feedback is integral to on-going teaching enhancement and the pursuit of teaching excellence.

Lastly, throughout the literature there is consensus that evaluation of teaching effectiveness should be located within the context of a faculty development model where faculty have input regarding evaluation processes, and where a supportive culture for teaching excellence is fostered, with a strong emphasis on teaching development and enhancement.

Teaching Excellence: A Briefing Paper

Introduction

Teaching excellence is "at the center of national and international higher education policy (HE) discourse" (Gunn & Fisk, 2013, p. 5). Teaching excellence is not only an internal issue for universities but, as Gunn and Fisk (2013) state, teaching excellence is also a matter for governments, employers, students, parents, and all stakeholders of the university. Much emphasis has been placed on higher education teaching practices over the past half century, with the establishment of learning and teaching centres at most institutions (paralleling the development of research services). However, significant debate continues as to what is meant by good, quality or excellent teaching in higher education.

The aim of the paper is to establish the current state of knowledge on teaching excellence within higher education, through examining different understandings, definitions, and models of teaching excellence, how teaching excellence is being recognized, and finally, how teaching excellence is being measured referred to as teaching effectiveness in the later portion of this paper. Understandings, definitions, and models of teaching excellence help to explain why, and how, the term is used, for although the term is in wide circulation, sources all confirm a general lack of consensus over what teaching excellence means (Brusoni et al., 2014; Greatbatch & Holland, 2016; Gunn & Fisk, 2013; Skelton, 2005, 2007). The idea of excellence in higher education may be as old as the university itself, but like the university, it is subject to ongoing and powerful forces of transformation and change (Rostan & Vaira, 2011b, p. 57) .

Describing learning and teaching in ancient universities, Skelton (2005) hypothesized that teaching excellence would be "associated with mastery of a discipline, the general sharpening of critical faculties, logical analysis and exposition and careful digestion of approved knowledge" (Skelton, 2005, p. 27). Today, however, the concept of teaching excellence is "witnessing changes in its contents, meanings, values, goals and tools" (Rostan & Vaira, 2011, p. vii). Gunn and Fisk (2013), pointing to gaps in the literature, describe understandings of "best teaching" as subject to two principal discourses: one is markedly critical focusing on wider higher education policy development while the other is more pragmatic, focused on "recognition and reward of teaching in a manner that implicitly acts as a catalyst, motivating academics towards excellent practices as well as increasing its parity of esteem with research" (p. 14). Broad terms are frequently used in the literature to label and differentiate teaching as good, best, quality or excellent; others refer to teaching quality on a continuum ranging from acceptable to exemplary; and still others articulate the relationship between quality teaching and student learning.

To begin to explore teaching excellence, it is first useful to consider the concept of quality in higher education, where related debates arguably precede those on teaching excellence (Harvey & Williams, 2010; Skolnik, 2010). Harvey and Green (1993) published a foundational paper on the meaning of quality in higher education, confirming the existence of different and competing definitions. The key point of their paper was that different higher education stakeholders—governments, employers, faculty, students, the public, etc., inevitably conceived of quality from different perspectives. Compounding this problem is the recognition that teaching is also highly context dependent, impacted by place, time, resources (human and other) and discipline (Sparrow, 2013). Within single institutions alone, there will be multiple perspectives, assumptions and expectations, as leadership, academics and students inevitably hold differing views (Gunn & Fisk, 2013; Sparrow, 2013). These entanglements are what Elton (1998) refers to as the multidimensionality of the concept of teaching excellence. It follows, as Brew (2007) observed, that if teaching excellence is not clearly defined within a higher education institution, conflicting demands may result. Gunn and Fisk (2013) relevantly query whether a single definition of teaching excellence can ever accommodate the complex and varied academic roles in universities today? A shared concern underlying all of this emphasis on teaching excellence is the belief that improved teaching will result in improved student learning (Elton, 1998; McAlpine & Harris, 2002).

Regarding both quality and teaching excellence, all stakeholders appear to share improvement of teaching and learning as a collectively recognized and agreed goal (Bartram, Hathaway, & Rao, 2018; Brusoni et al., 2014; Gunn & Fisk, 2013). As governments value their higher education sectors and wish to maintain quality, so higher education teachers clearly value their role as teachers (Nixon, 2007; Paulsen, 2002). Teaching is one of the two key pillars of the academic profession, and teachers "desire their craft to be recognized and valued by their institution" (Bartram, Hathaway, & Rao, 2018, p. 173). Agreement over the importance of teaching, however, does not resolve the questions of how *good* or *excellent* teaching should be carried out, or improved, nor does it confirm how good or excellent teaching should be defined, implemented, or ultimately assessed.

Before proceeding, it is important to note that the literature search on teaching excellence in higher education resulted in texts from within Canada, as well as a substantial international literature from countries where the debate has been most active to date. Available papers arise commonly from the UK (specifically England), Europe, Australia, and the United States. Canada's own literature employing the term, centers mostly around practical issues of implementation or the measurement of teaching excellence (Bartlett, 2013; De Courcy, 2015). Library searches particularly result in papers on Canadian teaching excellence awards. It is not the purpose of this paper to establish Canada's relative international position but literature does confirm distinctions of history, emphasis and approach (CMEC, 2007; Weinrib & Jones, 2014). A key difference may be that, compared to international peers, there has been relatively minimal government intervention to date in teaching matters within the Canadian context. As Land and Gordon (2015) observe in an extensive international review of teaching excellence:

> National approaches sit more readily in some countries… The US is an obvious example where federal influence on teaching excellence in higher education would not be the tradition… The UK and Australia both currently have governments prepared to be interventionist, if not dirigiste [excess control]…. The US, Canada, Germany, and Scandinavian countries seem less willing to give such strong steers and are perhaps still more trustful of their institutions to take responsibility for promoting teaching excellence. (p. 11)

Defining Teaching Excellence: What Is It?

The complex and contextual nature of teaching makes the goal of defining teaching excellence difficult. As already indicated, there is no universal definition of teaching excellence (Brusoni et al., 2014; Greatbatch & Holland, 2016; Gunn & Fisk, 2013; Skelton, 2005, 2007). Rather, "Excellence in higher education … depends on the person defining the term and their motivation for doing so" (Brusoni et al., 2014, p. 20). There are many candidates here:

> It is assumed that the different players in teaching excellence (senior academic management, disciplinary academics with a heavy research focus, disciplinary academics with a substantial teaching focus, clinical academics, generalist academics who want to be all-rounders, technicians, student services' providers, students, alumni, government bodies, and employers) come with different assumptions and expectations about necessary and desirable learning outcomes of university programmes. (Gunn & Fisk, 2013, p. 9)

In summary, definitions of teaching excellence will inevitably vary significantly, dependent upon students, discipline/content, leadership, teachers, learning and teaching contexts and resources. Definitions are, therefore, addressed here through differing lenses: first as an overarching discourse, then as faculty and student understandings, and finally, as theoretical and criterion-based models.

Discourses of teaching excellence. At a meta-level, there are broad discourses of excellence which function as over-arching frameworks for discussion and debate. Scholars here principally apply broad critical or pragmatic lenses according to purpose (Greatbatch & Holland, 2016; Gunn & Fisk, 2013). Whereas a critical reading may conceptualize teaching excellence as traditional, exclusive, elitist or ideologically-based (as in neo-liberal critiques) (Madriaga & Morley, 2016; Skelton, 2005), a more pragmatic reading focusses alternately on developmental pedagogic issues in daily institutional practice

(Gibbs, 2008, 2010; Gunn & Fisk, 2013). These descriptions tend to over-simplify, and often simultaneously contest and compete. Critical descriptions of teaching excellence as exclusive are linked to earlier, more traditional views, which link research pre-eminence and elite institutions with teaching excellence (Richardson, Moja, & Cohen, 2011; Skelton, 2005). An underlying assumption in early literature is that excellent teaching skills *inevitably* follow excellent research skills, a view conclusively disproved (Hattie & Marsh, 1996; Marsh & Hattie, 2002). Further critical conceptions of teaching excellence increasingly question global rankings and marketization as drivers of teaching excellence (Gourlay & Stevenson, 2017; Saunders & Blanco Ramírez, 2017; Skelton, 2005). However, teaching excellence equally preserves an alternative and pragmatic conception, more grounded in commitments to enhancement of teaching, the scholarship of teaching and learning (SoTL), teaching awards and incentives, and daily teaching practice with the intent of improving student learning.

Gunn and Fisk (2013) articulate this more pragmatic view where policymakers, institutions and academics seek an agreed way forward. This second discourse is perceived in "the integrity and authenticity of teaching enhancement as evidenced in the practice-based literature (and through the growth in student and staff led teaching excellence awards)" (Gunn & Fisk, 2013, p. 20). The "scholarship of teaching and learning" (SoTL), first named and encouraged by Boyer (1990), is embedded here, alongside sector and institutional efforts to level the playing field with research, partly by incentivizing teaching excellence. However, both critical and pragmatic discourses can be found existing concurrently in faculty understandings of teaching excellence.

Faculty understandings of teaching excellence. Faculty conceptions of good, quality, or excellent teaching, can be critical or pragmatic, and are not infrequently both. Critiques largely focus on neoliberalism (Rostan & Vaira, 2011b; Saunders & Blanco Ramírez, 2017), overly economic objectives (Gourlay & Stevenson, 2017) and the prevalence of "an empty rhetoric" (Wood & Su, 2017, p. 463). In practice, however, academics call for, and perceive, more nuanced and multi-faceted understandings of the term (Bartram et al., 2018; Gourlay & Stevenson, 2017; Wood & Su, 2017). Notions of teaching excellence, thereby, are not rejected by academics, but appear to co-exist within conflicting spheres, between the critical and pragmatic, as described by Gunn and Fisk (2013).

Wood and Su's (2017) study confirms academics hold multi-faceted understandings. A wide range of views are apparent through their study, but student-centered teaching styles, clear student-engagement with learning, high levels of both subject and pedagogic knowledge, and commitment to evaluating one's own practice, are commonly described. Individual comments further highlight concern for student learning gain, the presence of pedagogically informed teaching practices, and continuous intent to develop and improve: many participants "highlighted the importance of student feedback, self-reflection and peer review" (Wood & Su, 2017, p. 459).

Bartram et al. (2018) similarly comment on the complexity and diversity of faculty understandings, in a remarkably aligned set of findings from Australia and the UK. Critiques of "managerialism and marketing agendas" (p. 10) are balanced by thoughtful reflection on classroom teaching. The personal attributes of teachers (empathy, enthusiasm, openness, etc.) combine with particular teaching skills, which together emphasize: strong student-teacher interaction and relationships, student engagement, and both subject expertise and pedagogic knowledge. The authors (Bartram et al., 2018) also observe a commitment to professional improvement "evident in the high number of responses that suggested an important part of TE [teaching excellence] lay in being, and remaining, familiar with 'the evidence based continuum of methods which span the art and science of educational pedagogy to ensure constructive alignment of learning' " (p. 9). As faculty describe teaching excellence, a full spectrum of teaching practices become evident that will be further articulated under models of teaching excellence section (pg. 7). Faculty perspectives are next compared to, and augmented by, student perspectives, which will then be followed by those of respected educational scholars.

Student perceptions of teaching excellence. Faculty broadly emphasize personal attributes and teaching skills, with staff-student relations, high levels of interaction, participative pedagogies, student engagement, subject knowledge and professional pedagogic development, to the fore. Student conceptions are broadly in line with this, but with a clearer emphasis on emotional and social support roles. Studies indicate the following key overlapping factors constitute *excellence* for the student: the presence of emotional and social support, a student-centered teaching style, commitment to engaging students, and finally, views of learning as a partnership between diverse groupings of students and teacher. All such factors will similarly be found in theoretical definitions of *good*, *quality*, or *excellent* teaching. Each of the key factors mentioned by students is now expanded on below.

The presence of emotional and social support. In relevant studies, students place a clear emphasis on the personal qualities and empathy of the teacher, as well as their passion for their subject (Bradley, Kirby, & Madriaga, 2015; Greatbatch & Holland, 2016; Hill, Lomas, & Macgregor, 2003; Lubicz-Nawrocka & Bunting, 2018). In one study (Bradley, Kirby, & Madriaga, 2015) the top characteristic is described as, being supportive to students facing difficulties in their personal life or studies (p. 234). In the same study (Bradley et al., 2015) four of the subsequent seven characteristics include terms such as *helps*, *feel comfortable with*, *approachable* and *encourages* (p. 234). In line with this, students particularly want to "be taught by staff who are enthusiastic and knowledgeable about their subject, empathetic, approachable, helpful and patient, and encourage students to develop their full potential" (Greatbatch & Holland, 2016, p. 5). Hill, Lomas and MacGregor (2003) add that social and emotional support systems can be institutional, or involve relations between students and individual teachers. Additional peer support can also arise through shared, collaborative or small group learning in class (Hill et al., 2003; Jensen, Adams, & Strickland, 2014). To conclude, students appear to expect a strong level of welfare and personal support and emphasize this factor more strongly than do faculty. However, faculty, as indicated above, also clearly recognize student-teacher relations are an important element in teaching.

Teaching style and commitment to student engagement. The key element here is that the teaching style is student-centered. Students value teachers whose pedagogic style fosters the student's own engagement with learning (Hill et al., 2003; Lubicz-Nawrocka & Bunting, 2018). Contributing factors include: a motivating enthusiasm for the subject, learning experiences designed to stimulate interest and deepen student understanding, and a clear concern for a student's individual progress, as in timely provision of formative group, or individual feedback. Hill et al. (2003) prioritize a teacher's understanding of how the student learns and how knowledge is constructed in terms of the student's own mental models of reality (p. 38). Such teaching might integrate real life examples, or challenging questions and problems, in order to secure student engagement. Good organization and planning also contribute, through well-structured classes with clear aims and goals, which clearly confirm the teacher's effort and final intent (Lubicz-Nawrocka & Bunting, 2018). There is also a direct corollary here with academics' views of teaching excellence, regarding student-centered teaching, interactive pedagogies and the emphasis on student engagement.

Students as learning partners. This feature is also described as "Breaking down student-teacher barriers" (Lubicz-Nawrocka & Bunting, 2018, p. 8). In contributing practices, both students and teachers are seen as potential partners in learning, as, for example, in the creation of learning communities, or where the environment is structured to foster group activities or collaborative learning. It is suggested, in or out-of-class events might also be arranged to encourage interpersonal contact and constructive discussion between, either students and teacher, or students and other students. The provision of personal feedback, and also acting on student feedback, are finally seen as important factors here (Lubicz-Nawrocka & Bunting, 2018). All these points are arguably matched in the strong student-teacher interaction valued by academics, and reported by Bartram et al. (2018). Also repeated is the emphasis on self-reflection and pedagogic self-development.

To conclude, academic and student understandings of teaching excellence do not appear to be very different. The personal attributes of faculty, a teaching style that fosters student engagement, a strong student-teacher relationship, high levels of interaction and an interest in improving personal teaching practice, each have their equivalents in student understandings of the best teaching. It is notable, however, that each of these descriptions focuses on the perceptions and experiences of 'good teaching' and as such could/should be viewed as indicators or approximations of teaching excellence.

It is notable that the above understandings do not directly address, or focus on, student learning gains. Scholars and theorists describe many of the same features of teaching excellence but also draw attention to, and emphasize student learning. Learning gain will be discussed further, later in the paper.

Definitions and models of teaching excellence. Teaching excellence is somewhat difficult and elusive to define. As noted earlier, teaching excellence is contextual, dependent, for example, upon students, discipline, learning and teaching context and resources. Therefore, teaching excellence should not be reified as either absolute or universal by inferring that one best teaching practice might actually exist. Attributes of teaching excellence are more correctly interpreted as strategies that increase the likelihood of learning achievement, while equally recognizing that teaching must always consider the characteristics of the learners, as well as the context and content through which teaching occurs. However, none of the preceding has prevented educational scholars and theorists from seeking to study, define and advance proposed models of teaching excellence. These questions have led to models that define teaching excellence through specified lists of criteria. However, there is a prerequisite question that must first be addressed: can teaching excellence be easily or clearly distinguished from *good* or *quality* teaching? The uses of good, quality and excellent must therefore first be explored, before models of teaching excellence can finally be reviewed.

"Good teaching", "quality teaching" and "teaching excellence": Are they distinct*?*

Within the literature good teaching, quality teaching and teaching excellence are used to describe best teaching practices. Each has a particular focus, but each also consistently includes the following key features: research-informed pedagogic knowledge (SoTL); student-centered and collaborative teaching methods; the fostering of student engagement, as well as independent student learning; supportive and empathetic teacher-student relationships; appropriate assessment and feedback; and finally, teacher self-reflection based on varied forms of feedback and intent to progress and improve.

Good teaching. Good teaching has been described often and from varying perspectives, and it becomes clear, the *excellent* is not easily distinguished from the *good.* As Brusoni et al. (2014) write, "Excellence in teaching is determined by factors such as the inspirational nature of individual lecturers, the organization of presentations, the interaction with students as participants, and how well the information provided meets the learning objectives of the course" (p. 12). Here, there is little categorically separating *excellent* from *good.* Educational scholars overall, advocate teaching methods grounded in sound pedagogic research, as in educational psychology or SoTL (Barr & Tagg, 1995; Biggs & Tang, 2011; Chickering & Gamson, 1987; Marton & Saljo, 1976; Ramsden, 2003; Trigwell, 2001, etc.). Contemporary methods also generally embody what has been seen as the seminal shift from *instruction* or *teachin*g, to student *learning* (Barr & Tagg, 1995). The core intent of student-centered (or learner-centered) teaching design is "to create environments and experiences that bring students to discover and construct knowledge for themselves, to make students members of communities of learners that make discoveries and solve problems" (Barr & Tagg, 1995, p. 15). Today, student-centered approaches can be seen to foster: student engagement with content and materials (Ramsden, 2003a), *deep* as opposed to *surface* approaches to learning (Marton & Saljo, 1976), experiential or collaborative learning design (Chickering & Gamson, 1987), the appropriate use of learning outcomes, and finally *constructively-aligned* design of content, outcomes, teaching methods and assessment (Biggs & Tang, 2011). The focus, in short, is on *what the student does,* as opposed to what the teacher does (Biggs &

Tang, 2011). This latter principle is generally shared by well-regarded descriptions of *good, quality or excellent* teaching. Trigwell (2001) confirms "good teaching is the effective application of a combination of a scholarly approach to teaching, and teaching plans that are derived from (in alignment with) a student-focused conception of teaching" (p. 72). However, despite Biggs and Tang's (2011) assertion, *what the teacher does* is also clearly important.

Ramsden (2003) stresses that good teaching is not indefinable or elusive: "Good teaching and good learning are linked through students' experiences of what we do" (p. 84). Six key principles underpin Ramsden's (2003) "idea of good teaching" (p. 84), as follows:

1. *Interest and explanation*: stimulating student interest and making clear explanations foster student engagement, and result in deeper approaches to learning
2. *Concern and respect for students and their learning*: consideration, benevolence, availability, accessibility, generosity and humility, are each words used here
3. *Appropriate assessment and feedback*: the best and worst courses are separated by the quality and helpfulness of the feedback provided
4. *Clear goals and intellectual challenge*: a clear structure, and high expectations, which particularly support students to achieve higher levels of performance
5. *Independence, control and engagement*: the correct level of task, choice, as well as problem-solving and cooperative learning methods, ensure student engagement and foster deep learning
6. *Learning from students*: an evident and ongoing effort to diagnose and learn from student responses to instruction constitutes the most important action the teacher can take (Ramsden, 2003, pp. 93-99)

What is clear from Ramsden's (2003) text is the already familiar emphasis on student-centered teaching design described earlier, but there are also notably strong emphases on: the personal attributes and organizational skills of the teacher; the respectful and supportive quality of the student-teacher relationship; and the importance of evaluation, self-reflection and learning, specifically through student feedback. As indicated already, all these features re-appear in differing combinations throughout this paper.

Quality teaching. This term is more easily separated from the others, although not always. The term *quality* has mostly been applied in the context of higher education quality assurance policy, generally determined at government and sector level (Santiago, Tremblay, Basri & Arnal, 2008). National (or state, or provincial) higher education quality assurance policies are now common across the globe, and are aimed at maintaining and improving an agreed threshold standard of quality at institutional level. *Quality* here refers to teaching and learning, and not to research. Understandings of teaching quality are to be found in the detail of a quality assurance policy as it lays out its broad expectations and goals (Lewis, 2009). However, descriptions are, of necessity, relatively broad or generic, with emphases more on the overall quality of resources, faculty teaching practice, and the satisfactory student achievement of program, and course, outcomes and goals (see, for example: BC Government, 2016; OUCQA, 2010). Well-established definitions of teaching quality in this context are more reflective of different stakeholder perspectives, rather than specific or detailed pedagogies (Bogue, 1998; Harvey & Green, 1993). Further, as was stated, it is broadly accepted that *quality* in this context, and to date, is associated with a threshold, or fully acceptable standard, and not (yet) excellence (Brusoni et al., 2014; Greatbatch & Holland, 2016). As such, *quality teaching* is already distinguished from *teaching excellence.*

Teaching excellence. The word, excellence, implies the exceptional and "is generally defined as outstanding or as a quality that surpasses a defined threshold in a particular field" (Brusoni et al., 2014, p. 22). For example, McAlpine and Harris (2002) describe teaching as on a continuum, from acceptable to exemplary, with the culmination known as teaching excellence. Teaching excellence, as such, might be expected to belong to the province of the elite; to relativism and competition. However, other models demonstrate that teaching excellence is not, of necessity, confined in this way. As Tavares,

writing in Brusoni et al. (2014), asserts "An alternative version to the relative concept of excellence, far from the implied competition, elitism and exclusivity of some higher education institutions, should make it possible for everyone in principle, to attain that status" (p. 26). Greatbatch & Holland (2016) further explain:

> Excellence can also be interpreted using either norm-referenced definitions of excellence, which define excellence relative to the performance of others and imply competition and elitism, or criterion-referenced definitions of excellence, which define excellence in relation to a standard such that people are not in competition for it. (p. 5)

What this indicates is that, if criteria are identified and defined for teaching excellence, all institutions or individuals can aspire to excellence too. The existence of accepted criteria may, therefore, satisfy both competitive and more egalitarian developmental demands, according to the purpose to which the criteria are put. Teaching awards, for example, have a distinct purpose for faculty development. Teaching excellence can, therefore, refer to the elitist or the best, but equally may encompass the developmental or aspirational.

To conclude and summarize: good teaching, quality teaching and teaching excellence can be most easily distinguished when the contexts in which they are used are clearly identified. However, good teaching and teaching excellence are terms much more likely to be conflated in general use, as each can be applied ubiquitously or vaguely, to describe what could ultimately be very similar ends. In a competitive context, teaching excellence would mostly be "the best to be found"; in contrast, in an aspirational or developmental context, teaching excellence would be more of an agreed standard to aim at, in which case the application of good or excellent could easily be interchangeable. Gunn & Fisk (2013) observe the underlying paradox: in contexts where all are capable of becoming excellent, then "excellent" arguably becomes the new collective goal or agreed standard (p. 22). Excellence then becomes the new norm (or even threshold standard).

It is time to review theorists who have specifically defined the term *teaching excellence*, and this further confirms that *good* and *excellent* will continue to be conflated.

Models and criteria for teaching excellence. Successive models are presented chronologically here on the basis that the theorist (or group of theorists) has specifically addressed *teaching excellence,* as opposed to *good* or *quality* teaching. The theorists are: Sherman, Armistead, Fowler, Barksdale and Reif (1987), Elton (1998), Gibbs (2008), Sparrow (2013), and Gunn and Fisk (2016). Criteria for teaching excellence can be seen to develop and evolve through time.

***Sherman* et al. (1987**). This early and frequently cited model arises from a literature review of research into faculty and students' views of teaching excellence. This model can function competitively or developmentally. It is relativist in that "there appear to be qualities and /or conditions that separate excellent teachers from those who are very good, competent, or incompetent" (Sherman, Armistead, Fowler, Barksdale, & Reif, 1987, p. 66). However, it is also made clear that all teachers may develop teaching excellence, with the right conditions, motivation and training.

The model identifies the five key characteristics "regularly and consistently attributed to college instructors identified as excellent (Sherman et al., 1987, p. 67). These are:

1. Enthusiasm
2. Clarity
3. Preparation/organization
4. Stimulating
5. Love of knowledge (Sherman et al., 1987, p. 67)

All characteristics can be connected to ideas of teaching excellence already discussed. Teaching excellence here is assigned to the individual teacher in terms of personal attributes and their ability to organize. The individual teacher remains core to later models too, but it will be seen that wider criteria are progressively added in later models. In Sherman et al.'s (2013) model the itemization of key

characteristics alone does not identify what each characteristic looks like in practice. Each characteristic, therefore, is analyzed in more detail and additional factors in teaching excellence emerge, including:

- a deep knowledge of the subject,
- clear attention to student priorities and needs
- clear planning and clarification of course objectives
- clear planning and clarification of assessment
- stimulation of students' critical thinking
- making content relevant to students (Sherman et al., 1987)

All criteria, again, are familiar from faculty and student views of teaching excellence, and from earlier-cited principles of *good* teaching (e.g. Biggs & Tang, 2011; Ramsden, 2003).

Sherman et al. (1987) explicitly discount "talent or 'natural' ability" (p. 72), but then emphasize experience in conjunction with a developmental model for professional growth. Teaching excellence is here presented as something all can achieve, but only when the right factors are present. Experience alone is also explicitly discounted, and instead, research-informed pedagogic knowledge, and a personal prioritizing of teaching are proposed. In other words, teaching is a scholarly endeavour. One final factor discouraging teaching excellence might be an unsupportive, or otherwise non-ideal, environment. If the environment, for example, prioritizes research over teaching, achieving concurrent teaching excellence is more difficult (Skelton, 2005, p. 148; Gunn & Fisk, 2013, p. 9). To summarize, the excellent teacher is set apart only through their own efforts, through engagement with the scholarship of teaching and learning (SoTL), by being intrinsically and/or institutionally, motivated to aspire to teaching excellence, and by being situated in an ideal environment or context. It is notable that Sherman et al. (1987) incorporate what is an early description of SoTL.

Sherman et al. (1987) expand their study into a progressive plan for developing teaching excellence. Each of four progressive stages escalates towards a fourth and final requisite level (Sherman et al., 1987, p. 79). Excellence is achieved when teaching design, student learning, and teaching activities, exhibit a "recognizable and attainable" (p. 78) stage of complexity, interaction and understanding. Excellence here is found in teaching design, teaching methods and the evaluation of student learning. The developmental model also emphasizes: significant engagement with new teaching strategies or actions; the familiar self-reflection on one's teaching; and also familiar, evidence of progressive individual improvement.

To conclude, the authors' (Sherman et al., 1987) core point is that teaching excellence can potentially be attained by all. This model guides the achievement of excellence, but does not categorically set it apart from *good* teaching. As such it offers a pragmatic and practical concept of excellence. However, all depends on the purpose to which the model is put. Teaching excellence here is developmental, motivation dependent, and evident through an implicit set of criteria, where the individual teacher demonstrates:

- a deep knowledge of the subject
- clear attention to student learning priorities and needs
- clear planning and clarification of course objectives
- clear planning and clarification of assessment
- stimulation of students' critical thinking abilities
- content made particularly relevant to students
- engagement with the scholarship of teaching and learning
- significant engagement with new (innovative) teaching strategies and activities
- self-reflection on one's teaching
- evidence of personal, progressive improvement

Whereas these criteria apply to the individual, the next model provided by Elton (1998) frames teaching excellence more widely in terms of the institution.

Elton (1998). Elton (1998) adds another perspective by focussing on the levels of teaching excellence within a single institution: "first, classificatory, distinguishing the three levels of institution, department and individual, and second, substantive, describing the different ways in which each of the three levels can exhibit excellence" (p. 3). Teaching excellence is recognized, for example, as occurring individually with teaching awards, departmentally through recognition schemes developed by disciplinary or other organizations (such as specialized funds to support teaching), and finally, at the institutional level by their stated mission and values. Each level is subject to different criteria that would be considered as representative of teaching excellence.

The criteria evident in the next model (Gibbs, 2008) also demonstrate a wider institutional perspective. As Devlin and Samarawickrema (2010) stressed, understandings of teaching excellence must change over time; criteria must be reviewed and revised to incorporate new developments and imperatives.

Gibbs (2008). Gibbs' (2008) model draws wholly on teaching awards, with his underlying research addressing the weak rationales and inconsistent criteria evident in a myriad of UK award schemes. Brusoni et al. (2014) cite the "patterns of excellence in teaching and learning" (p. 25) emerging from Gibbs' (2008) study.

Gibbs' (2008) model presents a menu from which different types of awards may be constructed. This model reiterates and broadens criteria earlier seen in Sherman et al.'s (1987) scheme. Criteria are now not only focused on the personal act of teaching, but are wider in scope, including: the design of overall learning environments, the development of teaching skills in others, teaching-related leadership activities, fostering of related teamwork, community building, and more. The collective criteria are titled "Conceptions of excellent teaching underlying teaching award schemes" (Gibbs, 2008, p. 6). Gibb's (2008) principal criteria are paraphrased below, with author observations indented in italics:

- Exhibits observable, skilled teaching behaviors that correlate with student learning outcomes, and with student, or colleague, feedback.
  This is teacher-centered, outcome-centered and specifically reliant on feedback reflecting student expectations of excellent teaching. Peer feedback may equally contribute.
- Exhibits a clear focus on a student's learning, including how students learn (the process), and what is actually learned (the outcomes). Such teaching methods encourage deep rather than surface learning in students.
  This is familiar but additionally emphasizes the student learning process and student learning outcomes.
- Evidences engagement in SoTL. Prerequisites of excellence include: evidence of reflective thinking; a sophisticated personal philosophy of teaching; clear knowledge of pedagogic literature; engagement in pedagogic research.
  This is familiar and is developmental, emphasizing SoTL and personal professional growth.
- Focuses on personal development in the student as an alternative to discipline-focused teaching. Encourages wider attributes such as community awareness, good citizenship, transferable skills or employability, personal persistence etc.
  This expands conceptions of teaching excellence beyond subject content and discipline, to incorporate broader much broader and generic student learning outcomes.
- Creates wider learning environments where effective contexts enable educators or students. As examples: teaching teams effectively collaborate, students are supported by improved resources, or exceptionally creative assessment techniques, or exceptionally supportive monitoring of progress, etc.
  This pattern of teaching excellence could apply to course teams or departments, not only to individuals. This criterion further expands teaching excellence again, to another context beyond the individual teaching subject knowledge in a classroom.

- Disseminates teaching expertise or innovation to others. The emphasis is shifted from individual teaching performance to the encouragement of beneficial change in others.
  This also broadens the concept of teaching excellence beyond the individual and classroom, emphasizing sharing of expertise and wider personal influence. It does not include leadership which appears as a separate criterion (not covered here).
- Evidences improvement and innovation in teaching methods, which can be class, course or curriculum based. Innovation should specifically be valued and related to desired kinds of educational change. Related research may attract large scale funding.
  Improvement or innovation in teaching are familiar criteria, but are further qualified here, regarding whether each is well-conceived and appropriately directed.
- Evidence of teaching excellence which advances the approved institutional "mission" or "vision" statements.
  This adds contributions to overall organizational change. Teaching excellence here is connected to institutional aims, goals and priorities; these priorities might involve the teaching and learning strategy, or encompass wider political or social goals. (Gibbs, 2008, pp. 6-27).

Gibbs (2008) here has presented an expanded range of criteria and foci for teaching excellence, well beyond what the teacher does in the classroom in order to teach the subject discipline. New emphases include student and peer feedback, student achieved learning outcomes (including wider generic outcomes), wider learning environments, improvement and (qualified) innovation, and contribution to overall institutional strategies and goals. This is a considerably expanded view of teaching excellence acknowledging wider activities, environments and influence beyond a single classroom. Sparrow's (2013) model next incorporates now familiar criteria, but particularly adds a sense of impact extending into to the wider community and society.

Sparrow (2013). Sparrow's (2013) model follows work by McAlpine and Harris (2002) who earlier concentrated on the practical matters associated with teaching excellence; what they referred to as *effectiveness*. Through detailed analysis of the many factors that affect teaching, as well as the multiple perspectives that are required to evaluate teaching, these scholars provide a comprehensive and useful framework. Following an extensive literature review, Sparrow (2013) described teaching excellence as encompassing seven qualities:

- Teaching excellence as virtue addresses the values and goals of society as represented through the conduct of the post-secondary institution.
- Teaching excellence as quality focuses on how to ensure teaching practice is using quality tools and strategies that show learning improvement.
- Teaching excellence as scholarship highlights the necessity of research into teaching in higher education.
- Teaching excellence as good teaching is represented by teachers who demonstrate utilizing evidence-based strategies appropriately and effectively.
- Teaching excellence as student learning addresses the consideration given to providing an inclusive, safe, and encouraging learning environment for all students that helps them reach their full potential.
- Teaching excellence as elite performance addresses the award systems in place at the individual, departmental, and institutional level.
- Teaching excellence as community highlights collaboration and an expanded view of teaching in higher education.

Sparrow (2013) concludes with recognition that this breadth of elements, each and all, need to be considered when discussing teaching excellence.

Gunn and Fisk (2013). Finally, through an extensive and detailed analyses of international literature and models Gunn and Fisk (2013) summarize excellence in teaching practice, encompassing four broad dimensions within a simple table (reproduced as Table 1 below). This model provides an overview for easier comparison with previous models, outlining four principal "Dimensions of Excellence" (Gunn & Fisk, 2013, p. 34). These dimensions constitute the four main headings describing key areas where teaching excellence should be sought, in: planning and delivery, the design of assessment, contributing to the profession, and self-reflection and evaluation of one's teaching. These headings immediately echo criteria and emphases already encountered in Gibbs (2008, and others, above. For example: the organization of teaching and assessment; a personal impact on wider pedagogic reform; and self-reflection on one's own teaching. Sub-statements also include familiar criteria, including a central focus on how the student learns, engagement in SoTL, or engagement with, and in response to, student and peer feedback. Only one entirely new element is evident: "Innovation in delivery, assessment, feedback, evaluation, technology" (Gunn & Fisk, 2013, p. 34). Innovation itself is not new as an element, but the reference to technology is. However, technology is viewed as only one single element of teaching excellence here. This model, however, omits wider reference to the community and society, as in Sparrow (2013) above.

Table 1.
Dimensions of excellence in teaching practice

Planning and delivery • Curriculum design • Knowledge of the subject matter • Ability to Inspire and motivate learners • Respect, care and kindness for students as individuals • Active and group learning • Critical and scholarly	**Assessment** • Conscientious use of formative feedback • Creative and innovative approaches to feedback • Offering students a range of assessment to assess their mastery
Contributing to the profession • Innovation in delivery, assessment, feedback, evaluation, technology • Significant contribution to curriculum renewal and reform • SoTL • Participation in formal networks focused on teaching excellence • Broader leadership in teaching	**Reflection and evaluation** • Reflecting on inadequacies of own teaching • Degree of diligence in actively engaging with and responding to student and peer feedback and evaluations

Note. From "Dimensions of excellence 3: Excellence in teaching practice", by V. Gunn and A. Fisk, 2013, *Considering teaching excellence in higher education: 2007-2013*, p. 34.

To conclude, visions of teaching excellence arising from student and faculty views, and competitive teaching awards, lead to a set of criteria describing where teaching excellence can be found, as well as what it might look like in the university. Even when the context is competitive, the key point is that such criteria establish standards which can equally be d*evelopmental* and *aspirational*. As such, whether described as *good, quality* or *excellent* teaching, the practical and pragmatic contribution of these criteria at institutional and individual level, is to set achievable goals for improvement. The next question, therefore, is how are understandings of teaching excellence currently being implemented, at different levels, in different contexts, in practice?

Strategies to Recognize Teaching Excellence

Teaching excellence is fostered through multiple initiatives and strategies, at different levels of the system. The focus of teaching excellence may be individual, institutional, national, or even global, but the most recognized incentives are teaching excellence awards targeted toward different levels of the higher education system as a whole, including the well-established, faculty-specific, teaching awards. As Table 2 confirms, there is a multiplicity of purpose, approach and criteria, but also more than simply awards (Greatbatch & Holland, 2016; Land & Gordon, 2015).

Land and Gordon's (2015) comprehensive study of the *modalities* of teaching excellence indicates an extensive potential range. *Reward* here does not only refer to awards per se, but also to remuneration and other incentivizing approaches. Table 2 below, adapts Land and Gordon's (2015) own table, based on research into *modalities* from six continents and more than a dozen nations. Land and Gordon (2015) cite (and adapt) Dreyfus and Dreyfus (1986), who identified four levels of incentive or reward between novice and expert (p. 5). Increasing degrees of excellence were applied indicating development and progression, as demonstrated in Table 2 below. Competence, as the precondition for excellence, functions as the threshold or quality standard, set at "an agreed institutional or national minimum" (Land & Gordon, 2015, p. 5).

Land and Gordon's (2015) study details and discusses examples from each *Level of excellence* in Table 2. The following are but a few:

- *Proficiency level. Rewarding excellent programs.* 12 Australian Awards for Programs that Enhance Learning (APEL) have a prize value of AUS $25000 each (p. 7).
- *Proficiency level. Rewarding collaboration.* Significant injections of government funding to support teaching and learning, e.g. the addition of HK$7 million to existing Teaching Development Grants from the Hong Kong, University Grants Committee, to support collaborative projects on teaching and learning in the sector (p. 7).
- *Advanced proficiency*. Recognition of performance improvement through a national Professional Standards Framework (PSF). In the UK, the PSF establishes accreditation at *competen*t, *proficient* and *advanced proficiency* levels for individuals, as well as accreditation for professional development schemes at institutional level. Australia is doing the same, and Europe may follow (p. 10-11).
- *Expertise/high recognition*. Achievement across an entire career, as in Canada's STLHE Christopher Knapper Award, or the US AAUT Career Achievement Award (p. 12).
- *Expertise/high recognition. Rewarding great teaching ideas.* The Wharton-QS Stars Awards are at global level, open to all institutions and aimed at the transformation of H.E. teaching. (p. 11)

These are simply a few examples of how countries support and foster teaching excellence.

Initiatives and strategies not addressed by Land and Gordon (2015) further include: formalization of institutional teaching excellence and evaluation policies; recognition and incentives involving salary, merit and promotion policies (teaching professors, teaching fellows, etc.), and the targeting of teaching through significant government funding, major national government policy initiatives (Skelton, 2005) or specific grant initiatives.

Table 2.
Common initiatives, purposes and criteria for implementing teaching excellence

Level of excellence	Modality (or purpose)	Nature of criteria/indicators
Competence (Precondition of excellence)	Courses of initial training including professional teaching certificates, post-graduate certificates, as an element in doctoral programs, etc.	Evidence of performance that meets an agreed institutional or national minimum set of standards to enter a teaching career in higher education
Proficiency	Rewarding individual practitioners or teams Rewarding excellent programs Rewarding SoTL Student-led awards Rewarding collaboration Rewarding internationalization Rewarding and recognizing disciplinary teaching excellence Rewarding and recognizing inter-disciplinary teaching excellence	Performance beyond routine or habituated practices, exceptional or distinctive performance
Advanced proficiency	Rewarding innovation Citations for outstanding contributions to student learning Rewarding leadership	Performance indicating a senior level of practice or experience, leadership, or recognized authority in a given domain
Expertise/high recognition	Rewarding impact Rewarding great teaching ideas Creating new institutions of excellence Rewarding lifetime achievement National teaching award schemes Prime Minister's/Presidential awards Partnership awards	Distinguished performance which is widely acknowledged and merits a special degree or reward or ceremonial recognition

Note. Adapted from "*Teaching excellence initiatives: modalities and operational factors",* by R. Land and G. Gordon, 2015, p. 5. Copyright 2015 by The Higher Education Academy.

Measuring Teaching Excellence

What metrics and indicators are commonly used to evaluate teaching excellence? Having reviewed the meanings of teaching excellence and recognition strategies, this final section will first examine some metrics, indicators and measurement strategies in common use to determine teaching effectiveness**.** The paper will then consider specific aspects of ongoing debates concerning how teaching effectiveness (or excellence) is, or should be, measured. Finally, the paper will address the specific metrics and principles behind faculty teaching evaluation, ultimately focussing on student ratings of faculty teaching. Note that the latter are often referred to as student evaluation of teaching (SET), or more recently, as student ratings of instruction.

Metrics and indicators commonly used to evaluate teaching excellence. It has long been asserted that teaching is extremely difficult to measure (Bas, Tarantola, Carot, & Conchado, 2017; Gibbs, 2010; Harvey & Williams, 2010; Schleicher, 2015). What, therefore, are appropriate metrics (equivalent to *measures*) for evaluating teaching excellence, and is it helpful to separate good, or quality, from excellent in the measurement process? G*ood teaching*, *quality teaching* and *teaching excellence*, as has been seen, are easily and often conflated. To quickly review, therefore: quality is normally threshold standard, while teaching excellence in different contexts, can be normative or criterion –referenced, and have competitive, or aspirational and developmental purpose (Brusoni et al., 2014; Greatbatch & Holland, 2016). Governments, institutional leadership, faculty and students may each apply varying terms according to circumstance (Bartram et al., 2018; Bradley, Kirby, & Madriaga, 2015; Skelton, 2005; Wood & Su, 2017). Both measurement and terminology are therefore context and purpose dependent here regarding: what is meant by teaching excellence; what, or who, is being evaluated; why the evaluation is required; and the purpose for which the evaluation will be used? For these reasons, this account of measurement will largely disregard distinctions between quality teaching, good teaching and teaching excellence, to simply focus on measures seeking to identify the most valued and highest levels of teaching performance, referred to as *teaching effectiveness.* It should be noted here that scholars increasingly argue that the focus should be on student learning, rather than teaching excellence as performance (Gibbs, 2010).

There is extensive literature covering approaches to measurement, which addresses different levels of the system, as well as the multifaceted debates. It is impossible to detail every approach to the evaluation of teaching, or all arguments for and against. The focus will therefore be on contemporary thinking, and what is now broadly recommended for the measurement of teaching effectiveness, at institutional and faculty levels. This section of the paper, therefore, addresses the metrics and indicators commonly used to measure teaching effectiveness, introducing underlying rationales, before proceeding to ongoing debates. Note that the term *metrics* here describes indicators or other measures collectively, while i*ndicator* means something singular or particular, as in graduation rates, retention rates, etc.

Quantitative metrics*.* Quantitative metrics include *input* and *output* indicators, and sometimes, *process* measures too, although process metrics can also and often be qualitative. The terms, input, output and process are further explained below. Quantitative metrics also may include systemic measures such as student retention rates, graduation rates, graduate employment rates referred to as Key Performance Indicators (KPIs) by government systems (Finnie & Usher, 2005). Quantitative metrics are more easily collected than qualitative metrics, and being numeric, are well-suited to comparative assessments across institutions making them an efficient and popular administrative and student-friendly "entity". Quantitative metrics, however, are generally thought to provide inadequate information on the quality of teaching (Chalmers, 2008; Gibbs, 2010, 2016). Qualitative measures, in direct comparison, involve more extensive and detailed judgements, as in self-evaluation, peer review, teaching portfolios, academic review, and quality assurance review (Greatbatch & Holland, 2016, pp. 32-40).

Qualitative metrics. As inferred, the qualitative measurement of teaching generally involves lengthier and more complex processes of personal judgement, observation and assessment, submitting to greater language variations and interpretation. Gibbs (2010) proposes that the "right" subject for qualitative metrics are the *processes* of teaching and learning, concerning how teaching is done, how it is experienced, and particularly, how students respond to it (p. 16). As indicated also, qualitative measures are widely viewed by academics as better methods for evaluating teaching (Chalmers, 2008; Greatbatch & Holland, 2016). As Greatbatch and Holland (2016) stress, many consider it best if quantitative metrics are "supported by some form of peer review, accreditation, visit or audit" (Greatbatch & Holland, 2016, p. 40). *The point is that a combination of the quantitative and qualitative is better than quantitative alone*. Chalmers (2008) expands: quantitative metrics are felt to constrain "the

investigation of instructional, interactive and learning processes crucial to the quality of an institution, its educational programmes and its graduates" (p. 4). Qualitative approaches are viewed as more effective in capturing *what is going on* with teaching and learning.

A typology of metrics and indicators. Table 3 below is an amalgam of detail from Greatbatch and Holland (2016), Chalmers (2008) and Gibbs (2010). Table 3 includes only the better-known metrics and indicators, primarily providing examples to illustrate points. Distinctions here are not straightforward but broadly adhere to the following.

Gibbs' (2010) description of *presage*, *process* and *product* metrics has been applied in Table 3 to help clarify distinctions between the adjectives *input, output, outcomes* and *process* which appear in table headings. In brief: "Presage metrics define the context before students start learning, process variables describe what goes on as students learn, and product variables relate to the outcomes of that learning" (p. 4). There is a further distinction also between quantitative and qualitative metrics, as Table 3 also demonstrates.

The majority of process variables are qualitative, but following Gibbs (2010), a few are numeric too, as when indicators are coincident with student learning. For example, class size, contact hours and student/faculty ratios, follow Gibbs (2010) in being identified as numeric (or quantitative) process indicators, because they create conditions parallel to ongoing student learning. Chalmers (2008) explains that qualitative process indicators contribute to qualitative judgements by fostering deeper understanding of current practices. Relevant qualitative indicators, at either institutional or individual level, can be wide ranging and include: faculty self-evaluations of teaching, learning and teaching policies, professional development policies, course and curriculum documents, etc. Following such logic, metrics and indicators appear in Table 3 as either quantitative or qualitative, and further as input, output, process, or outcomes.

Table 3.
Examples of common metrics and indicators for measuring teaching at the Institutional Level

Quantitative metrics		**Qualitative metrics**
Input & output metrics	**Process metrics**	**Process and outcome indicators**
Inputs or presage metrics *Student characteristics* -qualifications on entry *Faculty characteristics* -research indicators -teaching qualifications -teaching awards *Supporting resources* Library facilities Investment in teaching and learning centers Investment in faculty development ***Outputs as product metrics*** Graduation rates Persistence/retention rates Graduate employment rates Employer/alumni ratings *Degree results* -exam pass rates -GPAs - degree classifications *Student satisfaction surveys*, e.g. -National Student Survey (NSS), UK	**Process metrics** (coincident with student learning) Class size Contact hours Student/faculty ratio	*Self-evaluation reports* *Peer reviews* *Curriculum reviews* *Teaching portfolios* *Quality assurance audits* *Faculty development policies* *Curriculum documents* *Assessment policies* *Appointment and promotion criteria* ***Qualitative output, outcomes, or product measures*** -student learning outcomes -Advanced Higher Education Learning Outcomes (AHELO) -Collegiate Learning Assessment (CLA), US ***Student engagement surveys*** -National Survey of Student Engagement (NSSE), U.S -UK Engagement Survey -HEPI-HEA Student Academic Experience Survey, UK

Adapted from "Teaching quality in higher education: Literature review and qualitative research," by D. Greatbatch and J. Holland. Copyright 2016 by the UK Government; "Dimensions of quality," by G. Gibbs. Copyright 2010 by the Higher Education Academy; "Teaching and learning quality indicators in Australian universities," by D. Chalmers. Copyright 2008 by the OECD.

***Measurement of teaching effectiveness*.** Debates regarding the measurement of teaching effectiveness generally circle issues of: how best to capture teaching practices and student learning, the validity and reliability of metrics and indicators used, and the tensions underlying related issues of autonomy and accountability. As stated, the body of literature examining the strengths and weakness of each category of measurement (or metric, or indicator) is extensive. The approach here has been to examine meta-analyses, such as literature reviews, or comprehensive studies by notable theorists or practitioners, for their conclusions on how teaching is best measured.

Multiple measures. The first observation has already been discussed. Researchers agree that, whether at institutional or individual level, qualitative and particularly process measures, create deeper understanding, while a combination of both qualitative and quantitative ideally informs the evaluation

of teaching effectiveness (Berk, 2005; Chalmers, 2008; Greatbatch & Holland, 2016; Knapper, 2001; Tam, 2001; Trigwell, 2001). A broad explanation of the underlying principle is that input metrics and process indicators are most useful when providing context, and when additionally combined with detail from output or outcome indicators. A well designed and combined set of qualitative, quantitative and particularly process metrics, will furnish the widest and best perspective on what is happening with teaching and learning, what the strengths and weaknesses are, and what needs to be done to improve (Chalmers, 2008, p. 6) at the institutional and individual level.

Student engagement and learning gain. Ideas coming to the fore in the literature on measurement increasingly stress student-achieved learning outcomes, teaching practices which ensure student engagement, and the idea of l*earning gain* (Evans et al., 2018; Gibbs, 2010, 2016; Kuh, Jankowski, Ikenberry, & Kinzie, 2014; Schleicher, 2015). For these reasons, Gibbs (2010) distinguishes between student satisfaction surveys, and student engagement surveys, only recommending the latter. In what is a comprehensive literature review on the presage, process and product dimensions of teaching quality, Gibbs (2010) asserts that the primary element in measuring effective, good or excellent teaching should be the notion of "educational gain" (p. 5). *Educational gain* means that the individual student, whatever their background or ability, and as a result of their degree, has personally gained or advanced, in terms of their own learning and attributes. This is the standard to which Gibbs (2010) holds each metric and indicator. Does the metric or indicator indicate a strong relationship with student learning gain? In agreement with Chalmers (2008), Gibbs (2010) confirms a preference for process variables, stating that "The process variables that best predict gains are not to do with the facilities themselves, but concern a small range of fairly well-understood pedagogical practices that engender student engagement" (p. 5). This view is already familiar from definitions of good or excellent teaching, and from criteria earlier examined (e.g. Chickering & Gamson, 1987; Hill et al., 2003; Ramsden, 2003). However, it should also be emphasized that one cannot simply equate student learning and student engagement, without clear evidence that fostering such engagement does lead to student learning gain. What, therefore, is the role and contribution of teaching effectiveness in *student learning gain*?

With learning gain at the core of the functional definition of teaching, it follows that assessments of teaching properly entail assessments of learning. As Evans, Kandiko Howson and Forsythe (2018) assert, "Learning gain approaches should be integral to curriculum design and delivery and not extraneous to it (p 1)." Similarly, the designation of teaching practices as "good teaching" should be predicated on evidence of the impact of those practices on learning [gain]. Below is a list of suggested teaching activities that arrange conditions to expedite learning and effect learning gain:

- Specifying intended learning outcomes that match the needs of individual learners, programs and the culture.
- Providing learners with models of cognitive domain (thinking, reasoning) and affective domain (feeling, caring) through examples and non-examples.
- Individualizing learning conditions to maximize learning for individual learners.
- Arranging motivating conditions that promote and support interest, practice, resilience, persistence, and independence.
- Arranging for contrived and genuine learning experiences.
- Attenuating conditions that interfere with learning.
- Arranging optimal sequencing of content streams and practice opportunities.
- Arranging conditions to promote "transfer" or "generalization" of learning to novel contexts.
- Arranging formative feedback conditions.
- Arranging mastery expectations and opportunities.
- Arranging active learning opportunities (hands on demonstrations, flipped classroom, assignments, presentations, *etc.*)

Such activities and conditions, it is therefore implied, might profitably inform appropriate and well-designed metrics and indicators which seek evidence of learning gain.

To recap, as the University of Alberta (2008) confirms, the overarching question has now changed from a focus on what resources an institution brings to teaching, to a focus on what the students are ultimately able to do and have learned (p. 4). In this light, simple quantitative input and output metrics are increasingly seen to be oblique of the mark as insufficiently related, or not related at all, to student engagement, or most significantly, to final student learning gain. To expand on this, what current metrics or indicators therefore, are presently evolving to now be regarded as perhaps better than others? Some straightforward examples will help to illustrate these issues.

It is widely accepted that the best qualified students are most likely to attend the highest ranked institutions, and then graduate with the best class of degree (Gibbs, 2010, p. 5). Prior student characteristics such as school qualifications, or other typical output indicators such as grade point average or degree classification are, therefore, increasingly considered misleading and of limited use: briefly, they are likely more indicative of the student and their background, confirming little about the quality teaching per se. Similarly, faculty input indicators, such as research citations or number of publications can be misleading too. While subject expertise undoubtedly contributes to good pedagogic practice, it is also well documented, as mentioned earlier, that excellent discipline researchers may or may not be excellent teachers: research confirms there is no inevitable correlation (Hattie & Marsh, 1996; Marsh & Hattie, 2002). Preferred indicators, as a result, could include quantitative process metrics such as class size, contact time, or faculty teaching qualification or development. The latter are each examples of indicators where evidence does exists that such factors do positively impact the quality of student learning, increasing learning gain (Gibbs, 2010; Knapper, 2010; Soilemetzidis, Bennet, Buckley, Hillman & Stokes, 2014).

To summarize and conclude, metrics and indicators for teaching evaluation remain the subject of intense debate, particularly where quantitative metrics predominate. There continue to be complex and ongoing arguments as to the best metrics to be used to evaluate teaching effectiveness in the university. Quantitative indicators are undoubtedly easier to collect and display (Spooren, Brockx, & Mortelmans, 2013)but current scholarship in the field suggests that a combination of qualitative and quantitative metrics is better, and that indicators which confirm learning gain, as well as appropriate types of student engagement, are increasingly valued and moving to the fore. The focus of the paper now turns to the second subject and an epicenter of debates; faculty teaching evaluations.

Faculty teaching evaluation: metrics and principles. Early evaluations of faculty teaching were largely based on student evaluations of teaching (SET), particularly in the US, but also in Europe (Berk, 2005; Knapper, 2001; Spooren et al., 2013). Spooren et al. (2013), in a European study, indicate that "Although SET was originally intended primarily for formative purposes, such evaluations came into use for faculty personnel decisions in the 1970s (Galbraith, Merrill, & Kline, 2012)" (p. 598). SETs are now widely used globally, but what is clear is that, in the past two decades, a wider range of metrics have been added (Berk, 2005; Cranton, 2001; Knapper, 2001). It is worth pausing here to briefly consider Canada's relative international position on teaching evaluation generally.

Higher education policy across Canada differs under the separate territorial and provincial governments, but generally individual universities (and particularly the research-intensive universities) appear relatively more autonomous in respect of teaching and learning, than many of their international peers. There is evidence of less "steering" from government (CMEC, 2007; Fisher, Rubenson, Shanahan, & Trottier, 2014; Weinrib & Jones, 2014) than would be found in the UK, Europe or Australia. Bartlett (2013), from the University of Toronto, provides a useful snapshot of Canada's relative international position on faculty evaluation for promotion purposes. With help from OISE (Ontario Institute of Studies in Education), Bartlett (2013) reviewed Canadian, Australian, New Zealand, South African and UK

practice, to establish how teaching excellence was measured for purposes of promotion and tenure. The national position in 2013 was summarized thus:

> The university systems in Australia, the UK, and New Zealand enjoy established and clearly defined national criteria for promotion. All require some assessment of teaching quality and factor this into the evidence for promotion and tenure, if appropriate. With national standards, it is somewhat easier to identify what is required and the nature of evidence to be determined. Other jurisdictions, such as the USA, have a variety of models, most of which identify teaching as a significant consideration, but few provide clear instructions as to what is to be collected and by whom. Finally, Canada, as in most things, falls between the two, with generally clearer guidelines but a great variation in evidence required and the means of assessment. (Bartlett, 2013, p. 357)

Where there is no singular national standard, how then do scholars in the field recommend evaluating the teaching practice of the individual instructor? As per measurement generally, the body of literature is again vast and the focus is once more on notable scholars and meta-analyses of the literature. Recommended approaches to teaching evaluation are reviewed first, before addressing the long-debated topic of student ratings and feedback, or SET.

Common metrics and indicators in faculty teaching evaluation. In an extensive literature review from the USA, Berk (2005) presents and analyzes 12 possible metrics for measuring teaching effectiveness:

a) student ratings
b) peer ratings
c) self-evaluation
d) videos
e) student interviews
f) exit and alumni ratings
g) employer ratings
h) administrator ratings
i) teaching scholarship
j) teaching awards
k) learning outcome measures
l) teaching portfolios (pp. 48-62)

Citing multiple studies from more than two previous decades, Berk (2005) reviews each metric in detail. Teaching evaluation is divided between *formative purposes,* as in evidence to support teaching development, and *summative purposes*, as in evidence to support performance review for salary, merit, promotion and tenure purposes. Berk's (2005) conclusions on the most valid sources of evidence are summarized in Table 4, below.

Table 4.
Comparing evidence to support teaching evaluation for summative and formative purposes

Best evidence for summative purposes	Best evidence for formative purposes
• student evaluations or ratings, • self-evaluation, • outcomes of interviews with students • administrator reviews (associate deans, program directors, department heads, etc.) using teaching criteria • examples of SoTL • individual faculty teaching portfolios	• student evaluations or ratings • peer ratings • self-evaluation • videos of instructor's own teaching • instructor-led quality control circles • structured class interviews (not instructor-led) • faculty teaching portfolios

Adapted from "Survey of 12 strategies to measure teaching effectiveness," by R.A. Berk, 2005, *International Journal of Teaching and Learning*, 17(1), pp. 48-62. Copyright 2005 by Creative Commons.

As evident in Table 4, but also from the text, student evaluations and teaching portfolios are emphasized by Berk (2005) for both formative and summative purposes, and there is further, the now

familiar emphasis on multiple measures including both the quantitative and qualitative. Berk (2005) stresses, however, that for summative purposes, student evaluations *should not* be used alone but in combination with other qualitative sources, particularly the teaching portfolio, quality circles or structured student interviews. Berk (2005) particularly acknowledges the heated debates, emotion and controversy around faculty evaluation, and specifically around student ratings and feedback (see also: University of Alberta, 2008; Berk, 2005; De Courcy, 2015; Gibbs, 2010; Knapper, 2001; Marsh, 2007; Spooren, Brockx, & Mortelmans, 2013). Student ratings are reviewed in more detail further below. Key theorists cited by Berk (2005) to support his conclusions include: Arreloa (2000), Braskamp and Ory (1994), Cashin (1989, 1990), Centra (1973, 1999), Knapper and Cranton (1997), McKeachie and Kaplan (1996), Seldin (1980, 1999), and multiple others.

Teaching portfolios, or dossiers (as they are commonly referred to within Canada), contain documentation about an instructor's experience teaching. The dossier begins with a teaching statement, or philosophy, including the instructor's teaching goals, claims, and strategies, and contains the evidence (in appendices) to support the instructor's teaching statement. The teaching narrative statement is a reflective narrative that is the key component of the teaching dossier through which the instructor makes claims about their teaching contributions. The teaching statement triangulates the claims with evidence from colleagues, trained peer reviewers and students that support those claims. The teaching dossier often includes evidence of the seven other metrics identified by Berk (2005), as well as evidence of engagement in teaching enhancement development activities.

Paulsen (2002) similarly reviews multiple studies of teaching evaluation, with the aim of establishing the reliability and validity of core sources of data, repeating that *multiple sources and types of data should be used.* Paulsen's (2002) main findings are that student ratings, qualitative group interviews with students, peer review of teaching (with reservations as noted below), self-evaluation, and peer review of the teaching portfolio, are common, reliable, and valid sources of data for faculty evaluation. With the exception of student ratings, these are all qualitative sources. As with Berk (2005), Paulsen (2002) has reservations about *untrained* peer evaluation. Peer evaluation should ideally not be used as a summative metric "in the absence of sound training or adequate numbers of observers" (p. 11). Careful selection of observers, appropriate training, more observers and more classroom visits, should, however, increase levels of reliability (Paulsen, 2002). Peer review of teaching portfolios should equally be based on sound training and involve portfolios with standardized requirements for content (Paulsen, 2002). Scholars finally highlight the benefits of discussing and negotiating metrics with faculty, and of providing ongoing support through faculty development programs and other resources (Knapper, 2001; Paulsen, 2002). Ideally faculty should be made familiar with all metrics and processes within a formative context, before summative evaluation is undertaken. Cranton (2001) finally repeats that qualitative and interpretive approaches, although time consuming, are most valuable in terms of depth and meaning, and can further be presented in easy-to-read formats (p. 16).

Reading all studies, however, there is one key issue acknowledged by all major scholars: student *ratings* of faculty teaching create more backlash and strong emotion than any other aspect of faculty evaluation, despite an extensive body of scholars who continue to support their use (Gibbs, 2010; Marsh, 2007).

Student ratings of university teachers. Student evaluations of teaching (SETs) are described as having three principal uses: to improve teaching and learning, to support salary and promotion decisions, and to inform institutional accountability reports (Spooren et al., 2013). As stated, SET is controversial. One of Canada's greatest teaching gurus, Christopher Knapper, wrote in 2001, it was a rare campus where student ratings were accepted with any equanimity (p. 3). At the time he wrote, however, student ratings had long been the dominant form of evaluation, particularly in the United States (Berk, 2005). There has since been considerable change, with a more eclectic approach and a wider range of metrics introduced, but yet there are ongoing and persistent concerns and debates.

More generally, the debate over student ratings remains intense and conflicting studies continue to come in (Rice Center for Teaching Excellence, 2018; Falkoff, 2018). Over the past several years there have been a number of empirical studies that have identified systemic gender bias, and bias against other designated equity groups, as a substantial concern in relation to student ratings of teaching (e.g. MacNell, Discoll & Hunt, 2014; Boring, Ottoboni & Stark, 2016; Wagner, Rieger & Voorvelt, 2016; Mitchell & Martin, 2018). Each of these studies, designed under specific conditions, provide evidence that systemic bias is, or may be, an intervening factor in student ratings. Findings from these studies further emphasize the problematic nature of focused reliance on student ratings and their potential impact on tenure, merit and promotion processes. Most recently, within the Canadian context, an arbitration decision between Ryerson University and the Ryerson Faculty Association (2018) (https://www.canlii.org/en/on/onla/doc/2018/2018canlii58446/2018canlii58446.html), addressed the summative use of SETs in the context of provisions in the Ryerson faculty collective agreement. The arbitration decision outlines the importance of SETs in capturing student experience and the value of that information to faculty and to the University, but cautions against blind use of SETs without attending to its limitations, such as issues of bias and unreliability. In particular, the arbitrator states: "The evidence is clear, cogent and compelling that averages establish nothing relevant or useful about teaching effectiveness. Averages are blunt, easily distorted (by bias) and inordinately affected by outlier/extreme responses. Quite possibly their very presence results in inappropriate anchoring" (p. 7). He then goes on to recommend the use of frequency distributions analyzed together with response rates (to assess the usefulness/reliability of the data). The award posits that teaching dossiers and peer evaluations are more effective in measuring teaching effectiveness.

This next section will, therefore, review the conclusions of those who have dedicated time and effort to reviewing SET related research accumulated over decades. Note that the language of student evaluation of teaching (SET) and student ratings of instruction (SRI) will be used. The latter reflects the more accurate role of students in providing feedback on their experience of teaching, in contrast to the notion that students are tasked with *evaluating* teaching. As Marsh (2007) stated, there are literally thousands of papers regarding student feedback on teaching. To date, there is no final singular conclusion but rather an informed debate that is ongoing. There is, however, consistent information and advice. Within this section, key meta-analyses are introduced chronologically, by title.

Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility (Marsh, 2007). Marsh's (2007) study is an updated review of the literature and research underpinning Marsh's own long career of research into student evaluations of teaching (SETS). Studies Marsh (2007) cites to support his conclusions are extensive and multiple, and broadly encompass the period from the 1970s to 2007. Marsh's (2007) conclusions are overwhelmingly supportive, that SETS are:

- multidimensional
- reliable and stable
- primarily a function of the instructor who teaches the 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 hypothesized as potential biases
- seen to be useful by faculty as feedback about their teaching, by students for use in course selection, and by administrators for use in personnel decisions (Marsh, 2007,)

(Note: the difference between Marsh's (2007) earlier findings and the later studies reported above in relation to bias.)

Marsh (2007) has conducted extensive research on SET and he is clearly a proponent. However, he *stresses the need for professional design knowledge, as well as strong views on appropriate purposes* for SET. Statistical and theoretical expertise is necessary to ensure any instrument is properly designed,

particularly regarding selection of "the components of teaching effectiveness that are to be measured" (p. 321). Marsh's (2007) study, overall, is critical of the extensive literature on bias in SETS results, blaming much on flaws of methodology, a lack of theory, and weak definitions of any bias itself. Areas of bias considered include the impact of prior subject interest, expected grade, workload difficulty, gender, year of course, academic discipline, etc. At no point does Marsh (2007) acknowledge a correlation of more than .30 between these characteristics and actual SET outcomes (p. 349). In Marsh's (2007) own view, however, SET *should be used to improve the teaching of individuals and programs, more than for salary or promotion purposes*. It is stated that much more research is needed and advised on when, and how, SET is used to inform salary or promotion decisions (Marsh, 2007).

The next meta-analysis to be reviewed was undertaken by a committee at the University of Alberta.

Evaluation of teaching at the University of Alberta. Report of the Sub-Committee of the Committee on the Learning Environment (CLE) (University of Alberta, 2008). The University of Alberta (2008), in line with Marsh (2007), concludes that any SET should be professionally designed by an expert, and implemented only in the context of multi-faceted evaluation, as one element amongst others. This literature review (University of Alberta, 2008) examined studies on: the validity of research results; the existence of bias in SET outcomes; whether students can effectively evaluate teaching; the impact of SET on teaching quality; and the use of SET in tenure and promotion decisions. What is clearest from the review is that there are conflicting conclusions between studies on each and every issue throughout; student evaluations are an issue which has been contentious since SET was first introduced (University of Alberta, 2008, p. 15). A further point is that studies are difficult to effectively compare as studies answer different questions through different methodologies. The University committee responsible, however, citing Arreola (2007,) does conclusively decide " 'properly constructed, appropriately administered, and correctly interpreted student rating[s] can be valid and reliable measures indicating the quality of teaching (Arreola, 2007, p. 98)' " (University of Alberta, 2008, p. 17).

Within its final recommendations, the U of A report notes the now familiar lack of agreement in the university, over what is actually meant by *excellent teaching* followed by already familiar advice: that faculty evaluation is supported by faculty development resources, and that faculty should have input regarding what is finally evaluated. Also recommended is a supporting reward structure, and that outcomes are provided confidentially to faculty, for teaching development and enhancement purposes. These same recommendations are also reflected in the next study; a large and particularly thorough literature review published at the end of 2013.

On the validity of student evaluation of teaching: The state of the art (Spooren et al., 2013)**.** If one meta-analysis is to be read, this is the paper recommended. The study is European but was published in the USA by the American Educational Research Association. The lens is one of validity through which a well-selected database of articles is viewed.

An extensive literature search evaluated 542 peer reviewed articles from authors in 12 countries, but the total was reduced to a final database of 160 texts which were evaluated as most appropriate for a study on validity. The texts themselves were then systematically analyzed to account for multiple types of validity, within three broad categories as follows:

1. *content-related* validity: involving detailed examination of the content of the instrument concerning, for example, what teaching effectiveness is, and whether the content itself makes sense to different stakeholders such as students, or faculty, or others.
2. *construct-related* validity: concerning the validity of the structure of the SET instrument, for example, as in the sub-set "convergent validity" where the correlation coefficient is sought by comparing final SET scores against student grades, or teacher self-evaluations, or the wide range of potential factors involved in bias.

3. *criterion-related* validity: for example, where the instrument is validated against one other criterion, as in a paper-based SET versus an online SET.

Study conclusions presented here, are those considered most useful or significant in the context of this paper. Ten main points are extracted, with the resulting advice from the report highlighted in bold:

- Content validity: the sub-set, *face validity*, is an issue, as different stakeholders can have different conceptions of good, or effective, teaching. **It is advised, that personnel decisions include other means of evaluation such as self-evaluation, peer reports, etc. To increase validity also, different stakeholders should ideally contribute to agreed definitions of good teaching.** This recommendation is already familiar but is widened to include all stakeholders.
- Construct validity: the sub-set, *structural validity*, of SET instruments varies. Although many are well founded on theory and rigorous validation, others are inadequately tested. Also, the majority of stakeholders, including faculty, administrators and students, generally have insufficient knowledge of the thousands of research studies published on SET. **Selection or design of instruments should be based on expert knowledge of both theory and validated instruments.** The study details ten validated instruments including: the Student Evaluation of Education Quality (citing Marsh et al. [2009]), the Course Experience Questionnaire (citing Ramsden [1991]), the Student Experience Questionnaire (citing Ginns, Prosser, & Barrie [2007]) and the Exemplary Teacher Course Questionnaire (citing Kember & Leung [2007]).
- Construct validity: the subset, *convergent validity,* demonstrates good correlation between SET scores and "teacher's self-evaluations, alumni ratings, and the evaluations of trained observers" (Spooren et al., 2013, p. 609). Student achievement is also correlated with SET scores. However, as stakeholders may hold different conceptions of what "good teaching" is, and faculty views of bias persist, **other metrics of teaching quality should be used, such as teacher's own reflections on their SET scores, and trained peer or educational expert reviews** (Spooren et al., 2013, p. 609)**.** These recommendations are also already familiar.
- Construct validity: *Discriminant validity* (indications of bias in student responses) of SET continues to be controversial with ongoing publication of new studies, and with results "genuinely mixed based on strong and less strong findings on both sides" (Spooren et al., 2013, p. 628). **The overall conclusion is that "SET remains a current yet delicate topic in higher education" (Spooren et al., 2013, p. 598) and should continue to be monitored when many indicators of effective teaching continue to be contested. SET should not be used alone to evaluate an individual's teaching and progress.** Although supportive of SET overall, the authors are dominantly sympathetic to faculty regarding their reaction to, and reservations about SET.
- Construct validity: the sub-set, *outcome validity*, indicates that teachers mostly agree with the use of SET for personnel purposes, but generally don't use SET to improve their teaching, despite their substantial agreement that SET outcomes provide useful insights. **It is advised that a more holistic approach be taken to encourage teachers to use SET outcomes for reflection and improvement, including support involving available expert consultation.**
- Criterion-related validity: studies here also indicate that SET scores are correlated with a range of quality teaching indicators, including achieved student learning outcomes, teacher self-assessments, and alumni assessments. Low student response rates, however, do affect validity. Criterion-based studies also confirm "results obtained with online SET instruments are similar to those obtained with paper-and-pencil instruments" (Spooren et al., 2013, p. 629), However, response rates are lower online. **Effort should be made to increase student response rates generally, and for online SETS in particular. There is some useful literature in this regard (e.g. Winer et al, 2016).**

Collectively, these meta-analyses provide varied conclusions but broadly consistent sets of advice, making recommendations in line with those detailed immediately above. No later, directly comparable and independent, academic meta-analyses have been located to provide conclusions distinct from those above. As mentioned, more recent writing clearly indicates that the debate continues (Rice Center for Teaching Excellence, 2018; Falkoff, 2018).

Conclusion

This paper set out to examine the concept of teaching excellence in both global and local terms, recognizing the various forces at play in both the definition of teaching excellence, and the measurement of teaching effectiveness. It is universally agreed, that to define teaching excellence is difficult: there is no single universal meaning. Rather, teaching excellence is understood through multiple lenses, shifting in response to attributes of students, priorities, context and purpose. Teaching excellence can perhaps best be described as a set of practices designed to maximize (increase the likelihood of) student learning, and can best be recognized as a scholarly endeavour founded upon SoTL research.

Faculty, student and theoretical understandings of the highest standard in teaching in higher education, exhibit multiple common expectations and criteria. Effective teaching practices are frequently identified as including:

- strong and supportive student-teacher relations,
- student-centered pedagogies, such as highly-interactive or collaborative teaching methods; pedagogies which similarly ensure student engagement;
- expert and inspiring knowledge of one's discipline and subject;
- strong organizational skills, evident in class structure;
- strong explanatory skills, evident in clarity and student learning;
- appropriate assessment and timely provision of feedback
- commitment to personal pedagogic self-reflection and professional development.

Current research further reaffirms long-standing pedagogical emphasis and research on student learning outcomes, and the increasing importance of student learning gain as a measure of teaching effectiveness. There is finally, a focus on the importance of student and peer feedback for teaching enhancement, with additional considerations for innovation and the worthwhile integration of technology.

Finally, the measurement of teaching continues to be, perhaps, one of the most outstanding contemporary challenges in the field of teaching and learning in higher education. The development of a shared understanding of what is actually meant by teaching excellence is an important starting place in the articulation of effective measures. Teaching effectiveness (like teaching excellence) is considered difficult to measure with most scholars pointing to the importance of using multiple measures, with a combination of multiple sources of quantitative and qualitative viewed as better than simply quantitative alone or a singular source. In fact, the literature is clear, that a well-designed and broad set of measures will provide the best assessment of teaching effectiveness, identifying strengths and areas for improvement. The teaching dossier is repeatedly described as the most effective strategy through which teaching effectiveness can be evidenced and teaching enhancement goals are identified. Ideally, measures of teaching effectiveness must reflect student learning and be founded in the knowledge of best teaching practices. While student feedback on teaching may be one component to be considered, there are multiple others sources of feedback including alumni ratings, employer ratings, administrator ratings, teaching scholarship, teaching awards, achieved learning outcomes, and more comprehensively, teaching portfolios. Notably peer review of teaching was viewed with some caution with scholars recommending that it is best undertaken by trained peer reviewers or educational experts.

Student evaluation of teaching (SET) continue to be the most contested aspect of evaluation. While all scholars addressed the importance and essential nature of student feedback, they also

consistently make clear that SETs should never be used as a sole measure of teaching effectiveness, and that the emphasis of SETs should be on improvement of teaching, rather than assessment for merit or promotion purposes. Further, when SETs are used, they should be designed by experts with knowledge of both effective teaching practices and instrument development (validity and reliability), and be used within the context of a multi-faceted evaluation. Strategies to maximize response rates are imperative to the effective use of SETs, particularly in the online environment.

Lastly, evaluation of teaching effectiveness should be located within the context of a faculty development model where faculty have input regarding evaluation processes, and where a supportive culture for teaching excellence is fostered, with a strong emphasis on teaching development and enhancement.

References

Barr, R. B., & Tagg, J. (1995). From teaching to learning : A new paradigm for undergraduate education. *Change*, *27*(6), 12–25.

Bartlett, K. (2013). Defining university teaching excellence in a globalized profession. *Cases on Quality Teaching Practices in Higher Education*, (chapter 21), 354–365. https://doi.org/10.4018/978-1-4666-3661-3.ch021

Bartram, B., Hathaway, T., & Rao, N. (2018). 'Teaching excellence' in higher education: a comparative study of English and Australian academics' perspectives. *Journal of Further and Higher Education*, *00*(00), 1–15. https://doi.org/10.1080/0309877X.2018.1479518

Bas, M. del C., Tarantola, S., Carot, J. M., & Conchado, A. (2017). Sensitivity analysis: A necessary ingredient for measuring the quality of a teaching activity index. *Social Indicators Research*, *131*(3), 931–946. https://doi.org/10.1007/s11205-016-1297-2

BC Government. (2016). *Quality assurance process audit (QAPA) handbook*. Victoria: BC Government, Ministry of Advanced Education. Retrieved from https://www2.gov.bc.ca/assets/gov/education/post-secondary-education/institution-resources-administration/degree-authorization/qapa/2_qapa_handbook.pdf

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. https://doi.org/10.1007/3-540-69876-0

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: what the student does* (4th ed.). Maidenhead, England: McGraw-Hill, Society for Research in Higher Education & Open University.

Bogue, E. G. (1998). Quality assurance in higher education: The evolution of systems and design ideals. *New Directions for Institutional Research*, (99), 7–18. https://doi.org/10.1002/ir.9901

Boring, A., Ottoboni, K. & Stark, P.B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research* (1) https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Boyer, E. L. (1990). *Scholarship reconsidered: priorities of the professoriate*. Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching.

Bradley, S., Kirby, E., & Madriaga, M. (2015). What students value as inspirational and transformative teaching. *Innovations in Education and Teaching International*, *52*(3), 231–242. https://doi.org/10.1080/14703297.2014.880363

Brew, A. Approaches to the scholarship of teaching and learning. In A. Brew, & J. Sachs, (Eds.). *Transforming a University: The Scholarship of Teaching and Learning in Practice* (pp. 1-10). Sydney: Sydney University Press.

Brusoni, M., Damian, R., Sauri, J. G., Jackson, S., Komurcugil, H., Malmedy, M., … Zobel, L. (2014). *The concept of excellence in higher education*. Brussels, Belgium. Retrieved from http://www.enqa.eu/indirme/papers-and-reports/occasional-papers/ENQA Excellence WG

Report_The Concept of Excellence in Higher Education.pdf
Chalmers, D. (2008). *Teaching and learning quality indicators in Australian universities*. IMHE Programme on Institutional Management in Higher Education. Retrieved from https://www.oecd.org/site/eduimhe08/41216416.pdf
Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education (AAHE) Bulletin*, (March), 3–7. Retrieved from https://files.eric.ed.gov/fulltext/ED282491.pdf
CMEC. (2007). *Ministerial Statement on Quality Assurance of Degree Education in Canada*, 1–14. Retrieved from http://www.cmec.ca/Publications/Lists/Publications/Attachments/95/QA-Statement-2007.en.pdf
Cranton, P. (2001). Interpretive and critical evaluation. *New Directions for Teaching and Learning*, *2001*(88), 11–18. https://doi.org/10.1002/tl.33
De Courcy, E. (2015). Defining and measuring teaching excellence in higher education in the 21st century. *College Quarterly*, *18*(1). Retrieved from http://collegequarterly.ca/2015-vol18-num01-winter/decourcy.html
Devlin, M., & Samarawickrema, G. (2010). The criteria of effective teaching in a changing higher education context. *Higher Education Research and Development*, *29*(2), 111–124. https://doi.org/10.1080/07294360903244398
Elton, L. (1998). Dimensions of excellence in university teaching. *The International Journal for Academic Development, 3*(1), 3-11, DOI: 10.1080/1360144980030102
Evans, C., Kandiko Howson, C. & Forsythe, A. (2018) Making sense of learning gain in higher education, *Higher Education Pedagogies*, 3:1, 1-45, DOI: 10.1080/23752696.2018.1508360
Falkoff, M. (2018, April 25). Why we must stop relying on student ratings of teaching. *The Chronicle of Higher Education*. Retrieved from https://www.chronicle.com/article/Why-We-Must-Stop-Relying-on/243213
Finnie, R., & Usher, A. (2005). *Measuring the quality of post-secondary education : Concepts, current practices and a strategic plan*. *Educational Policy*. Kingston, ON & Toronto, ON. Retrieved from http://immagic.com/eLibrary/ARCHIVES/GENERAL/CPRN_CA/C050407F.pdf
Fisher, D., Rubenson, K., Shanahan, T., & Trottier, C. (Eds.). (2014). *The development of postsecondary education systems in Canada: A comparison between British Columbia, Ontario and Quebec, 1980-2010*. Montreal and Kingston: McGill-Queen's University Press.
Galbraith, C.S., , Merrill, G.B. & Kline, D.M. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and bayesian analyses. *Research in Higher Education, 53*(3), 353-374, DOI: 10.1007/s11162-011-9229-0.
Gibbs, G. (2008). Designing Teaching Award Schemes. The Higher Education Academy. Retrieved fromhttps://www.heacademy.ac.uk/system/files/gibbs_final_manual_1.pdfGibbs, G. (2010). *Dimensions of quality*. York, UK. https://doi.org/10.1097/00005110-200109000-00006
Gibbs, G. (2016). Teaching. In *Response to the higher education green paper. HEPI report 81* (pp. 11–25). London, UK: Higher Education Policy Institute (HEPI). Retrieved from https://www.hepi.ac.uk/wp-content/uploads/2016/01/STRICTLY-EMBARGOED-UNTIL-7-JANUARY-2016-FINAL-GREEN-PAPER-RESPONSE-21_12_15-Screen.pdf
Gourlay, L., & Stevenson, J. (2017). Teaching excellence in higher education: critical perspectives. *Teaching in Higher Education*, *22*(4), 391–395. https://doi.org/10.1080/13562517.2017.1304632
Greatbatch, D., & Holland, J. (2016). *Teaching quality in higher education: Literature review and qualitative research, May 2016*. London, UK: UK Government. Department of Business Skills and Information. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/524495/he-teaching-quality-literature-review-qualitative-research.pdf

Gunn, V., & Fisk, A. (2013). *Considering teaching excellence in higher education: 2007-2013: A literature review since the CHERI Report 2007*. York. Retrieved from https://www.heacademy.ac.uk/system/files/resources/TELR_final_acknowledgements.pdf

Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education*, *18*(1), 9-34.

Harvey, L., & Williams, J. (2010). Fifteen years of quality in higher education. *Quality in Higher Education*, *16*(1), 3–36. https://doi.org/10.1080/13538321003679457

Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching : A meta-analysis. *Review of Educational Research*, *66*(4), 507–542.

Hill, Y., Lomas, L., & Macgregor, J. (2003). Students' perceptions of quality in higher education. *Quality Assurance in Education*, *11*(1), 15–20. https://doi.org/10.1108/09684880310462047

Jensen, K., Adams, J., & Strickland, K. (2014). Inspirational teaching: Beyond excellence and towards collaboration for learning with sustained impact. *Journal of Perspectives in Applied Academic Practice*, *2*(2), 37–41. https://doi.org/http://dx.doi.org/10.14297/jpaap.v2i2.88

Knapper, C. (2001). Broadening our approach to teaching evaluation. *New Directions for Teaching and Learning*, (88), 3–9. https://doi.org/10.1002/tl.32

Knapper, C. (2010). Changing teaching practice: Barriers and strategies. In J.Christensen-Hughes & J. Mighty (Eds.), *Taking stock: Research on teaching and learning in higher education* (pp. 229-242). Montreal and Kingston: Queen's Policy Study Series, McGill-Queen's University Press.

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014). *Knowing what students know and can do: The current state of student learning outcomes and assessment in U.S. colleges and universities*, (January), 1–50. http://www.learningoutcomeassessment.org/documents/2013%20Abridged%20Survey%20Report%20Fina

Land, R., & Gordon, G. (2015). *Teaching excellence initiatives: modalities and operational factors*. York. Retrieved from https://www.heacademy.ac.uk/system/files/resources/teaching_excellence_initiatives_report_land_gordon.pdf

Lewis, J., & Nicholls, C. M. (2014). Design issues. In J. Ritchie, J. Lewis, C. M. Nicholls, & R. Ormston (Eds.), *Qualitative research practice: A guide for social science students and researchers* (2nd ed.), pp. 47–76). London, UK: SAGE Publications Ltd.

Lewis, R. (2009). Quality assurance in higher education - its global future. In *Higher education to 2030. Volume 2. Globalisation* (pp. 323–355). Paris, France: Organisation for Economic Co-operation and Development (OECD). Centre for Educational Research and Innovation (CERI).

Lubicz-Nawrocka, T., & Bunting, K. (2018). Student perceptions of teaching excellence: an analysis of student-led teaching award nomination data. *Teaching in Higher Education*, *0*(0), 1–18. https://doi.org/10.1080/13562517.2018.1461620

Madriaga, M., & Morley, K. (2016). Awarding teaching excellence: 'what is it supposed to achieve?' Teacher perceptions of student-led awards. *Teaching in Higher Education*, *21*(2), 166–174. https://doi.org/10.1080/13562517.2015.1136277

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 707–754). Dordrecht and London: Springer. https://doi.org/10.1037/0022-0663.76.5.707

Marsh, H. W., & Hattie, J. (2002). The relation between research productivity and teaching effectiveness: Complementary, antagonistic, or independent constructs? *The Journal of Higher Education*, *73*(5), 603–641. https://doi.org/10.1353/jhe.2002.0047

Marton, F., & Saljo, R. (1976). On qualitative differences in learning: I—Outcome and process. *British*

Journal of Educational Psychology, *46*(1), 4–11.
MacNell,L., Discoll, A. & Hunt, A.N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*: 291. https://doi.org/10.1007/s10755-014-9313-4
McAlpine, L. & Harris, R. (2002). Evaluating teaching effectiveness and teaching improvement: A language for institutional policies and academic development practices, *International Journal for Academic Development, 7*(1), 7-17, DOI: 10.1080/13601440210156439
Mitchell, K.M.W. & Martin, J. (2018). Gender bias in student evaluations. *PS, Political Science & Politics. 51*(2). 648-652. https://doi.org/10.1017/S104909651800001X
Nixon, J. (2007). Excellence and the good society. In A. Skelton (Ed.), *International perspectives on teaching excellence in higher education* (pp. 15–31). Abingdon, UK & New York, NY: Routledge, Taylor & Francis Group.
OUCQA. (2010). *Quality Assurance Framework.* Ontario Universities Council on Quality Assurance (OUCQA). https://doi.org/10.1201/9781439834367.ch2
Paulsen, M. B. (2002). Evaluating teaching performance. *New Directions for Institutional Research*, *2002*(114), 5–18. https://doi.org/10.1002/ir.42
Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London, UK: Routledge.
Rice Center for Teaching Excellence (2018). *Research on student ratings continues to evolve* [Blog posting]. Retrieved November 4, 2018, from http://cte.rice.edu/blogarchive/2018/2/20/studentratingsupdate
Richardson, R. C., Moja, T., & Cohen, U. R. I. (2011). Defining and developing excellence in three national settings: Israel, South Africa and the U.S. In M. Rostan & M. Vaira (Eds.), *Questioning excellence in higher education: Policies, experiences and challenges in national and comparative perspective* (pp. 76–92). Rotterdam, The Netherlands: Sense Publishers.
Rostan, M., & Vaira, M. (2011a). Questioning excellence in higher education: an introduction. In M. Rostan & M. Vaira (Eds.), *Questioning excellence in higher education: Policies, experiences and challenges in national and comparative perspective* (pp. vii–xvii). Rotterdam, The Netherlands: Sense Publishers. https://doi.org/10.1007/978-94-6091-642-7
Rostan, M., & Vaira, M. (2011b). Structuring the field of excellence. In M. Rostan & M. Vaira (Eds.), *Questioning excellence in higher education: Policies, experiences and challenges in national and comparative perspective* (pp. 57–74). Rotterdam, The Netherlands: Sense Publishers.
Ryerson University v Ryerson Faculty Association, (2018). CanLII 58446 (ON LA)
Santiago, P., Tremblay, K., Basri, E., & Arnal, E. (2008). *Tertiary education for the knowledge society. Volume 1. Special features: Governance, funding, quality*. Paris, France. https://doi.org/10.1787/9789264046535-en
Saunders, D. B., & Blanco Ramírez, G. (2017). Against 'teaching excellence': ideology, commodification, and enabling the neoliberalization of postsecondary education. *Teaching in Higher Education*, *22*(4), 396–407. https://doi.org/10.1080/13562517.2017.1301913
Schleicher, A. (2015). *Value-added: How do you measure whether universities are delivering for their students? HEPI 2015 annual lecture*. Paris, France. Retrieved from http://www.hepi.ac.uk/wp-content/uploads/2016/01/Andreas-Schleicher-lecture1.pdf
Sherman, T. M., Armistead, L. P., Fowler, F., Barksdale, M. a, & Reif, G. (1987). The quest for excellence in university teaching. *The Journal of Higher Education*, *58*(1), 66–84. https://doi.org/10.2307/1981391
Skelton, A. (2005). *Understanding teaching excellence in higher education: Towards a critical approach*. Abingdon, UK & New York, NY: Routledge.
Skelton, A. (Ed.). (2007). *International perspectives on teaching excellence in higher education* (Paperback,). Abingdon, UK & New York, NY: Routledge.
Skolnik, M. L. (2010). Quality assurance in higher education as a political process. *Higher Education*

Management and Policy, *22*(1), 1–20. https://doi.org/10.1787/hemp-22-5kmlh5gs3zr0

Soilemetzidis, I., Bennett, P., Buckley, A., Hillman, N., & Stoakes, G. (2014). *The HEPI-HEA Student Academic Experience Survey 2014.* York, UK: Higher Education Policy Institute (HEPI), Higher Education Academy (HEA). Retrieved from https://www.heacademy.ac.uk/system/files/resources/hea_hepi_report_web.pdf

Spooren, P., Brockx, B., & Mortelmans, D. (2013). *On the validity of student evaluation of teaching: The state of the art*. *Review of Educational Research* (Vol. 83). https://doi.org/10.3102/0034654313496870

Sparrow, H. (2013). Teaching excellence: An illusive goal in higher education teaching and learning. Retrieved from https://ro.ecu.edu.au/theses/582

Trigwell, K. (2001). Judging university teaching. *International Journal for Academic Development*, *6*(1), 65–73. https://doi.org/10.1080/13601440110033698

University of Alberta. (2008). *Evaluation of teaching at the U of A: Report of the Sub-committee of the Committee on the Learning Environment (CLE)*. Edmondon, Canada. Retrieved from https://cloudfront.ualberta.ca/-/media/universitygovernance/documents/member-zone/gfc-standing-committees/cle-subcommittee-reports/teaching-evaluation-subcommittee-january-2009.pdf

Wagner, N., Rieger, M. & Voorvelt, K. (2016). Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams. *Economics of Education Review (54)* https://doi.org/10.1016/j.econedurev.2016.06.004

Weinrib, J., & Jones, G. A. (2014). Largely a matter of degrees: Quality assurance and Canadian universities. *Policy and Society*, *33*(3), 225–236.

Winer, L.R., di Genova, L., Costopoulos, A. & Cardoso, K. (2016). Addressing common concerns about online student ratings of instruction: A research-informed approach. *Canadian Journal of Higher Education, 46*(4), 115-131.

Wood, M., & Su, F. (2017). What makes an excellent lecturer? Academics' perspectives on the discourse of 'teaching excellence' in higher education. *Teaching in Higher Education*, *22*(4), 451–466. https://doi.org/10.1080/13562517.2017.1301911