SIMON FRASER UNIVERSITY PROJECT ON
STUDENT EVALUATION OF TEACHING AND COURSES

SUPPORTING DOCUMENTATION ITEM 1
REPORT ON KEY RESEARCH FINDINGS

PREPARED FOR THE SIMON FRASER UNIVERSITY SENATE COMMITTEE ON UNIVERSITY TEACHING AND LEARNING

15 JANUARY 2012

JOHANNE PROVENÇAL, PH.D.
# TABLE OF CONTENTS

**INTRODUCTION** .................................................................................................................................................. 2

  Brief history of the project ........................................................................................................................................ 2

  International context and broad issues identified in the research ............................................................................. 3

    The multi-dimensionality of teaching and the question of “good teaching” ............................................................. 4

    Understandings about student learning and the shift to learner-centred curriculum .................................................. 6

    Issues of validity and reliability in student evaluation of teaching ........................................................................... 9

**EVALUATION OF TEACHING AND COURSES** ....................................................................................................... 12

  Peer evaluation of teaching ........................................................................................................................................... 12

  Teaching portfolios ......................................................................................................................................................... 13

  Student evaluation of teaching and courses ................................................................................................................... 15

**STUDENT EVALUATION OF TEACHING AND COURSES: ISSUES FOR FACULTY, ADMINISTRATORS, AND STUDENTS** ............................................................................................................................... 16

  Issues for faculty ............................................................................................................................................................ 16

    Recognizing possible student bias ............................................................................................................................... 17

  Issues for administrators and the university administration ....................................................................................... 21

  Issues for students ........................................................................................................................................................ 24

**REFERENCE LIST** ...................................................................................................................................................... 29
INTRODUCTION

The purpose of this report is to provide the Simon Fraser University (SFU) community – faculty,1 administrators, and students – with a summary of research findings in support of the project overseen by the Office of the Vice-President Academic and Provost to update the SFU process and form used for student evaluation of teaching and courses. In the interest of effective and responsible use of student evaluation of teaching and courses at SFU, wide use of this report is encouraged and therefore, the main text shall remain relatively brief with footnotes and further reading in the reference list for those interested in further details.2

BRIEF HISTORY OF THE PROJECT

The Senate Committee on University Teaching and Learning (SCUTL) report, “Evaluating How We Evaluate” (2009) and the 2010 “Task Force on Teaching and Learning Recommendations Report” (which followed university-wide consultations) both called for a review of the process used by SFU for the evaluation of teaching and courses. In 2011, the Vice-President Academic and Provost, Jonathan Driver, made a request to Senate to approve a project to be led by SCUTL to review and update the instrument used for student evaluation of teaching and courses at SFU and to provide documentation to support faculty, administration, and students in their use of the new instrument and in the interpretation of instrument data. After Senate approval of the project in the summer of 2011, SCUTL and the project team3 began work on the project in the first semester of the 2011/2012 academic year, with the administrative support of the SFU Teaching and Learning Centre.

From the outset of the project, the Vice-President Academic and Provost, as well as the project team, have aimed to communicate clearly that this project did not arise from dissatisfaction with teaching and learning at SFU. In addition, those involved in the project recognize the importance of a set of evaluative process for teaching and courses in which student evaluation of teaching and courses is only one component. The purpose of the present project is to develop an updated form and process for student evaluation of teaching and courses at SFU that reflects new understandings from educational research on the widespread use, nationally and internationally, of student evaluation of teaching and courses.

---

1 Throughout this report, the term “faculty” is used to represent tenured and tenure-track faculty members, lab instructors, lecturers and senior lecturers, limited term appointments, and sessional instructors.
2 While the main text of this report, admittedly, contains more than 8000 words, it is relatively brief, given the vastness of the literature on the relevant questions for SFU’s project on student evaluation of teaching and courses.
3 At the time of writing (January 2012), the project team members include: Corinne Pitre-Hayes (Project Manager), Stephen Spector (SCUTL Chair), Hui Niu (Researcher, Instrumentation), Johanne Provençal (Researcher and Writer), and Chris Groeneboer (Manager, Applied Research, Teaching and Learning Centre, SFU).
INTERNATIONAL CONTEXT AND BROAD ISSUES IDENTIFIED IN THE RESEARCH

In the past 30 years student evaluation of teaching has become widely adopted internationally.\(^4\) The evaluation of universities, more broadly, has become of increasing importance as accountability for post-secondary institutions and concerns regarding student learning have drawn the attention of researchers in the field of higher education, of governments providing funding to universities, of industry and community organizations employing university graduates, and of students (as well as parents) investing time and tuition dollars in post-secondary education. Evaluation and quality assurance for universities in Europe, for example, is an important part of the Bologna Process.\(^5\) Quality assurance organizations in the United Kingdom (UK)\(^6\) and Australia\(^7\) include in their mandates evaluation of teaching and learning, and the Council for Higher Education Accreditation in the United States (US) and the Association of Universities and Colleges of Canada also support quality assurance.\(^8\) In Canada, there are post-secondary education quality assurance measures within the provincial jurisdiction for education in several provinces.\(^9\) The following is a useful description of the context for universities in Canada:

Global competition necessitates that, as Canadian institutions, we maintain our competitive edge and remain cognizant of major international developments beyond our borders; in particular, those shepherded by the Bologna process in the European Union, the Programme on Institutional Management in Higher Education (IMHE), and the Forum on Higher Education of the Organization for Economic Co-operation and Economic Development (OECD). It also necessitates that we keep an open mind and ear to what competencies are required for our students to be successful as life-long learners and professionals anywhere they choose to fulfill their aspirations.\(^10\)

\(^4\) Chen and Hoshower (2003, p. 71) cite, for example, studies by Seldin (1985), Abrami (1989), Wagenaar (1995), Abrami et al (2001), and Hobson and Talbot (2001) on the international use of what is widely referred to in the research literature as “student evaluation of teaching” (SET) or “student evaluation of teaching effectiveness” (SETE).

\(^5\) See the Council of Europe (http://www.coe.int/t/dg4/highereducation/ehea2010/bolognapedestrians_en.asp): “The Bologna Process is a European reform process aiming at establishing a European Higher Education Area by 2010” and follows from the 1999 Bologna Declaration, which “states the following objectives:

- adoption of a system of easily readable and comparable degrees;
- adoption of a system essentially based on two main cycles, undergraduate and graduate;
- establishment of a system of credits – such as in the ECTS [European Commission Transfer and Accumulation];
- promotion of mobility by overcoming obstacles to the free movement of students, teachers, researchers and administrative staff;
- promotion of European co-operation in quality assurance;
- promotion of the necessary European dimensions in higher education” (2011).

\(^6\) See the Quality Assurance Agency for Higher Education: http://www.qaa.ac.uk/.

\(^7\) See the Tertiary Education and Quality Standards Agency: http://www.teqsa.gov.au/. Also, as Gapp and Fisher describe, “Since the mid-1980s the university sector in countries such as Australia has seen increasing political interventions, an outcome of which has been the demand to implement course and teaching evaluations. These government initiatives have clearly been made a requirement of various funding models” (2006, p. 156).

\(^8\) Further, as Hallinger notes, “Although the use of accountability tools in higher education began as a largely ‘Western’ phenomenon, in recent years their adoption has become global in scale. The impact of quality assurance processes driven by government and accreditation agencies has been further accentuated in recent years by the dissemination of World University Rankings” (2010, p. 253).

\(^9\) In British Columbia, see the BC Education Quality Assurance organization: http://www.bceqa.ca/

\(^10\) See Saroyan 2010, p. 103.
An important part of quality assurance for universities, as can be expected, is in the area of teaching or instruction. As Algozzine et al report, “By far, the most common method of evaluating instruction in university classes is to have students provide feedback on instructional ‘effectiveness’ using rating scales”\textsuperscript{11} and citing Berk,\textsuperscript{12} the authors note that, “student ratings have had a greater impact as a source of evidence for assessing teaching than all of the remaining dozen sources combined.”\textsuperscript{13} That said, precisely because of the significance of student evaluation of teaching and courses, it has been subject to scrutiny and criticism (if not cynicism) both informally – by instructors and faculty, as by administrators and students – and more formally investigated by researchers either within their own disciplines or by those in the field of higher education. As a consequence, a significant research literature on student evaluation of teaching (and courses) has formed in the past 30 years\textsuperscript{14} and there is much that it can offer not only to the project team in updating the university’s instrument for student evaluation of teaching and courses, but also, for faculty, administrators, and students as they use the instrument and interpret the data it produces. An overview of key issues from the research literature is presented, below, with focus on some of the most significant of these issues in the sections that follow.

**The multi-dimensionality of teaching and the question of “good teaching”**

The complexity and multi-dimensionality of teaching are widely recognized by scholars and practitioners in the field of education, as by faculty in their experiences in university or college classrooms.\textsuperscript{15} It is perhaps not surprising, then, that “higher education has yet to establish a universally accepted definition of the characteristics and skills necessary for teaching excellence.”\textsuperscript{16} This is in part because what is expected from a “good teacher” can “vary from faculty to faculty since different are the needs, the disciplinary contents and the possible practical activities that show an applicative use.”\textsuperscript{17} In one study, “components of effective teaching identified from analysis of student ratings include six common dimensions...to as many as twenty-eight dimensions.”\textsuperscript{18} An understanding of teaching excellence clearly extends beyond the questions used on student evaluation of teaching forms and it is worthwhile to note that:

\textsuperscript{11} See 2010, p. 28. Here, Algozzine et al also cite the following: “Adams, 1997; Algozzine et al., 2004; Berk, 2006; Blunt, 1991; Braskamp, Caulley, & Costin, 1979; Bryson, 1974; Centra, 1977; Cohen, 1981; Darby, 2007; d’Apollonia & Abrami, 1997; Feldman, 1989; Platt, 1993; Remmers, 1927; Rifkin, 1995; Soyjka, Gupta, & Deeter-Schmelz, 2002; Sproule, 2000; Starry, Derry, & Wright, 1973; Theall, Abrami, & Mets, 2001; Theall & Franklin, 1990” (2010, p.28).

\textsuperscript{12} See Berk 2006, p.15.


\textsuperscript{14} As an example, in the literature review by Al-Issa & Sulieman (2007), they found 2988 articles on student evaluation of teaching (following the widely cited review by Cashin in 1988, which found more than 1300 articles).

\textsuperscript{15} See, for example, Ghedin and Aquario 2008.

\textsuperscript{16} See Arreola 2007, p. 98.

\textsuperscript{17} See Ghedin and Aquario 2008, p. 595.

\textsuperscript{18} See Paulsen 2002, p. 8.
Although researchers use different terms, there is a consensus regarding six dimensions [of “good teaching”]: rapport with students, course value, course organization and design, fairness of grading, difficulty, and workload. Ramsden (1992) lists six key principles of effective teaching in higher education as: (a) Interest and explanation, (b) Concern and respect for students and student learning, (c) Appropriate assessment and feedback, (d) Clear goals and intellectual challenge, (e) Independence, control and active engagement, and (f) Learning from students. Feldman (1976) identified 22 constructs of effective teaching, concluding that the two most highly rated dimensions were stimulation of interest and clarity of presentation.¹⁹

Perhaps more important than common questions on student evaluation of teaching forms is, as Ghedin and Aquario remind us, that how a “good teacher” is defined depends very much on conceptions about teaching²⁰ and this holds for faculty and students, both.

As the establishment of teaching centres within universities has initiated discussion about teaching and broadened faculty approaches to and understandings of teaching,²¹ studies in the field of higher education have attempted to identify and characterize faculty approaches, beliefs, and theories about teaching. The typology developed by Ramsden, for example, identifies what he refers to as “different theories of teaching represented in lecturers’ attitudes to teaching and their instructional strategies” and these are: teaching as “telling and transmission…teaching as organising student activity…[and] teaching as making learning possible”²² Similarly, Pratt et al developed a “Teaching Perspectives Inventory”²³ based on a series of 13 studies in which faculty classified their approach to teaching according to learning focus, with five approaches identified as follows:

1. The Transmission Perspective: Effective Delivery of Content;
2. The Apprenticeship Perspective: Modeling Ways of Being;
3. The Developmental Perspective: Cultivating Ways of Thinking;
4. The Nurturing Perspective: Facilitating Self-Efficacy; and
5. The Social Reform Perspective: Seeking a Better Society.²⁴

Underlying each of the above characterizations of teaching are assumptions and perspectives not only about the purpose(s) of higher education, but also, about the roles and responsibilities of both teacher and student. Particular attention has been given in recent years to questions of student learning and as a consequence, there has been a notable shift in higher education from a

¹⁹ See Khandelwal 2009, p. 299. Khandelwal also notes that, “Murray, Rushton, and Paunnen (1990) reported that forty to seventy percent of the variance in student teacher ratings could be accounted for by six personality traits, namely leadership, extroversion, liberalism, supportiveness, intellectual curiosity, and changeableness” (p. 300).
²⁰ See Ghedin and Aquario 2008. The authors also cite studies by Kember and Wong (2000) and Lecouter and Del Fabbro (2001).
²¹ See Knapper 2010, p. 238.
²³ See http://teachingperspectives.com/
“teacher-centred” to a “learner-centred” classroom and curriculum. Among the contributing factors in this shift have been changed views about and approaches to knowledge (where, how, and by whom “knowledge” is produced, legitimized, shared, etc.) alongside growing understandings about student learning (in particular, recognizing the increasing cultural and socio-economic diversity of the student population), an appreciation for different learning styles, and implications of the 21st century technological and media environment.

Understandings about student learning and the shift to learner-centred curriculum

While there are serious concerns about student learning in college and university, Harris and Cullen (with their view shared by many others) note, “One of the oddities of the tradition of higher education is that professors are rarely provided any instruction or professional development in the role that represents a major element of their responsibility: teaching.” Yet, teaching faculty have a great deal of experience encountering issues of student learning. Weimer, in a somewhat light-hearted but also sincere description, presents the following as the most common concerns that faculty raise in regard to students as learners:

- students are passive
- students lack confidence as learners
- many students lack the basic study skills necessary to succeed in college
- the only thing that motivates students are grades, points, and marks.

---

25 It should be noted that the term “student-centred” can also be used here. In this document, “learner-centred” is being used for reasons that are summarized well by Weimer: “Being student-centered implies a focus on student needs. It is an orientation that gives rise to the idea of education as a product, with the student as a customer and the role of the faculty as one of serving and satisfying the customer. Faculty resist the student-as-customer metaphor for some very good reasons. When the product is education, the customer cannot always be right, there is no money-back guarantee, and tuition dollars do not ‘buy’ the desired grades. Being learner-centered focuses attention squarely on learning: what the student is learning, how the student is learning, the conditions under which the student is learning, whether the student is retaining and applying the learning, and how current learning positions the student for future learning” (2002, p. xvi).

26 Similar to Pratt et al, though focusing on a teacher- or student-centred spectrum of teaching, there is also Trigwell’s description of “Five qualitatively different approaches to teaching...Approach A: Teacher-focused strategy with the intention of transmitting information to students...Approach B: Teacher-focused with the intention that students acquire the concepts of the discipline...Approach C: A teacher-student interaction strategy with the intention that students acquire the concepts of the discipline...Approach D: A student-focused strategy aimed at students developing their conceptions...Approach E: A student-focused strategy aimed at students changing their conceptions” (2010, p. 117)

27 See Harris and Cullen 2008, p. 58. Although part of graduate training often involves opportunities for teaching assistant positions and professional development offerings to support faculty in teaching are available on most campuses, doctoral-level programs are widely recognized as placing focus on training for research skills within a particular discipline. This is likely a contributing factor for “faculty on average hav[ing] unrealistically high self-perceptions of their own teaching effectiveness” (Khandelwal 2009, p. 299, citing Marsh and Dunkin 1997, p. 276). Knapper identifies several problematic areas in university teaching: [1] Teaching remains overwhelmingly didactic and reliant on traditional lectures, and assessment methods are often trivial and inauthentic; [2] Curriculum development relies far too much on disciplinary tradition and faculty interests, rather than on student and societal needs; [3] There is still a ‘tyranny of the academic disciplines’ which mitigates against integration of knowledge and insights from different fields; [4] Evaluation of teaching effectively and learning outcomes is often superficial” (2010, p.238).

28 The description by Weimer, though certainly not to be taken as a fair description of all students, is quite likely to be familiar to many faculty: “Based on my own experience as an educational developer, [faculty] would identify one or several of these four problems. First, students are passive. They want to sit back and have education ‘done unto them,’ and they hope the experience will be pleasant and painless...They want the teacher to write the material on the board, or better yet show it on a PowerPoint presentation that can conveniently be downloaded. They really like it when teachers provide complete sets of notes from class.
Although Weimer recognizes the challenges that faculty confront in regard to student learning, she also calls on faculty to consider how teaching can contribute to these difficulties, and that research on student learning can be drawn on to understand why these kinds of difficulties occur and ways to overcome them. In turning to the research for possible ways forward, Meyer describes, for example, a shift in the late 1970s that resulted in part from asking students about their own experiences of learning, making reference in particular to the widely-cited research by Marton and Säljö, who, “disturbingly demonstrated…that students differed from one another in their learning intentions…reflecting a categorisation not of students themselves but rather of the variability in what they do and why the do it in terms of learning process and underlying intention” (italics in original). An important finding in this research is the significance of “metalearning” in which students, becoming aware of their learning, gain “a recognition of self in relation to ‘learning’ in…[a particular] context and a consideration of what, in their own minds, they actually do when they are ‘learning’ in that context.” A similar principle is also significant for faculty in their awareness of their teaching.

A learner-centred classroom requires awareness by faculty of how their teaching supports learning. Perhaps one of the most promising aspects of a learner-centred approach to teaching is that, “When the focus is on student learning, the assessment of teaching takes into account...”

Next, students lack confidence as learners. They do not like it when they have to make decisions related to learning. They want teachers to spell out exactly what they should do. Teachers should specify paper length, appropriate fonts, margin size, and the number of references for the bibliography. Students want to know whether they should write in the first person, or include examples or quotations. They want teachers to ‘go over’ what will be on the test, detailing those sections of the text that should be the focus on in-depth study. Sometimes they find it difficult to say for sure whether or not they understand something. Third, faculty might observe that many students lack the basic study skills necessary to succeed in college. They struggle to read technical material, even texts that have become simpler and better organized than they used to be. They depend on a limited repertoire of study strategies—flash cards, recopying notes, and underlining material in texts. If those strategies fail, students still use them, thinking that the only option is to use them more. Often they write and calculate poorly. Finally, faculty might complain that the only thing that motivates students are grades, points, and marks. Students rarely demonstrate any intellectual curiosity. They will volunteer but only if the learning opportunity involves points, even just a few of them. They engage in intellectual dialogue with passion only when the argument involves the possibility of more points or an increase in partial credit (2010, p. 81-82).

29 See Meyer 2010, p. 192 (citing Marton and Säljö 1976). The categorization the authors refer to is that of deep learning versus surface learning. Lindblom-Ylänne describes, for example, how a student “applying a surface approach to a reading assignment concentrates on the text itself. A deep approach, on the other hand, is based on a genuine interest in the subject matter and the aim is interpreting the meaning of the text...Previous research has shown that the deep approach to learning is more likely to be related to higher-quality learning outcomes than a surface approach” (2010, p. 64).

30 See Meyer 2010, p. 200. Meyer also notes, however, that teaching faculty often “require something to prompt [in students] an empowering initial realisation of self as learner in process terms, the emphasis being on ‘what am I doing?’ rather than ‘who am I?’ The distinction is crucial because empowerment for students begins with the realisation that their learning processes can vary” (2010, p. 200).

31 Kinzie notes, for example, the need to understand what students expect from their experience in order to use instructional approaches that support them in becoming intentional learners able to reach intended outcomes. Kinzie also notes that, “what faculty members emphasize and think is important to learning can influence what students do” (2010, p. 143). In the same vein, Meyer offers the reminder that from “assesssment cues (in lectures) and the ‘ticky box’ multiple-choice test at the end, a student forms a temporal ‘cram and dump’ conception of ‘learning’ and she provides the following excerpt from a student participating in one of her studies: ‘I’m more focused on achieving a high mark than I am in gaining more knowledge. If I were to shift more focus towards placing value on the learning itself, high marks might come as an added benefit. … I harbour some resentment towards students who consistently memorise information without understanding it, write it down on an exam without really thinking, and then forget it all afterwards, but I am also guilty of the same practice’” (2010, p. 195).
what students are doing because it is assumed that they are [or can be] active participants who are engaged in learning.”³² Weimer again provides a useful perspective on what this means in the day-to-day realities of the classroom, suggesting five changes that are needed in post-secondary teaching practice in order to overcome the challenges that faculty commonly confront:

(1) **The balance of power:** Motivation to learn is greatly affected if students feel a sense of empowerment in their learning.

(2) **The function of content:** The relationship between course content and the learner is significant to the learning process. Here, constructivist approaches to learning are useful for understanding how “learners’ actively construct…their own knowledge rather than passively receiving information transmitted to them from teachers and textbooks.”³³

(3) **The role of the teacher:** Faculty expertise, rather than being exclusive, encourages and legitimizes different ways for students to interact with and relate to course content.

(4) **The responsibility for learning:** Many students, by the time they reach college or university, have become dependent learners in teacher-centred learning environments in the sense of depending on the teacher “to identify what needs to be learned, to prescribe the learning methods, and…to assess how well they have learned,” which runs counter to the development of “independent, autonomous learners who assume responsibility for their own learning.”³⁴ These skills must be taught, however, and require an understanding (by teachers and students) of the responsibilities of the teacher for teaching and the learner for learning.

(5) **Evaluation purpose and processes:** Students learn throughout their education to be motivated to focus on learning material on which they will be tested or otherwise assessed. It is important for teachers and students, both, to consider the implications of this approach to evaluation and to build in students an ability to conduct a self-assessment of the extent to which they do or do not understand what is being taught.³⁵

Weimer cautions, however, that some level of student resistance can be expected in shifting from a teacher-centred to learner-centred classroom and recommends an article that she reads before the start of every course.³⁶ That said, it is also worth noting, as Prosser reports from a series of

---

³² See Higgerson 2006, p. 36. See also Combs et al (2008) and their discussion of the importance – for faculty and students – of clear learning objectives and the usefulness of using a learning objectives tool, both for student evaluation of teaching and for student learning: “Students benefit directly from the use of this learning objective tool in several ways. The use of the tool requires stated learning objectives that clarify what the course is to deliver, which contributes to students’ understanding of what the instructor views as the important components of the course content. The identification of learning objectives also helps show students how the different course elements link to one another” (p. 90).
³⁴ See Weimer 2002, p. 15.
³⁶ Weimer recommends and quotes from “Navigating the bumpy road to student-centered instruction” by Felder and Brent: “It’s not that student-centered instruction doesn’t work when done correctly—it does, as both the literature and our personal
studies, that teachers who adopt teacher-focused approaches to teaching have students tending toward surface approaches to learning; while teachers who adopt student-focused approaches to teaching have students tending toward deeper approaches to learning.37

Issues of validity and reliability in student evaluation of teaching

Although, as Arreola notes, “Decades of research have demonstrated that tools used in faculty evaluation systems, especially student rating forms, can be designed to be valid and reliable,” there remain persistent challenges of validity and reliability in developing and implementing instruments for student evaluation and teaching.38 Chen and Hoshower provide a helpful description of the distinction between reliability and validity studies:

Reliability studies (Marlin & Gaynor, 1989; Scherr & Scherr, 1990; Nimmer & Stone, 1991; Wachtel, 1998) generally address the question ‘Are student ratings consistent both over time and from rater to rater?’. On the other hand, validity studies (Howard et al., 1985; Byrne, 1992; Tagomori & Bishop, 1995) address the questions ‘Do student ratings measure teaching effectiveness?’ and ‘Are student ratings biased?’.

Although validity and reliability in student evaluation of teaching and courses is to be discussed in detail in a separate report,40 some attention is given here to these issues. As a starting point, a more specific definition of validity is as follows:

[Validity is] the effectiveness of the test in representing, describing, or predicting the attribute that the user is interested in. Content validity refers to the faithfulness with which the test represents or reproduces an area of knowledge. Construct validity refers to the accuracy with which the test describes an individual in terms of some psychological trait or construct. Criterion-related validity, or predictive validity, refers to the accuracy with which the test scores make it possible to predict some criterion variable of educational, job, or life performance.41
For student evaluation of teaching and courses, a key question for validity is, whether or not the instrument measures what it is intended to measure. This, then, is a key point of discussion for SFU faculty, students, and administrators, because if the instrument is to measure teaching effectiveness, a decision must be made, for example, about the components of effective teaching to include in questions (and options for responses) in the student evaluation of teaching instrument. Similarly, the findings in the research literature on both the components of effective teaching and instruments used to measure effective teaching are important to inform the update of the student evaluation of teaching instrument. As Marsh (2007) notes, “poorly worded or inappropriate items will not provide useful information, while scores averaged across an ill-defined assortment of items offer no basis for knowing what is being measured” (p. 321). In regard to reliability, the test phase of the development of the instrument is important for ensuring reliability in inter-rater results, re-test results, and internal consistency. Although concerns regarding reliability and validity are important to recognize and address, as Chen and Hoshower note:

Overall, the literature supports the view that properly designed student ratings can be a valuable source of information for evaluating certain aspects of faculty teaching performance (Cohen, 1981; Marsh, 1984; Calderon et al., 1994).

Although these kinds of findings from research on student evaluation of teaching can be reassuring, there are also studies identifying sources of student bias that can affect validity and reliability of student evaluation of teaching and courses.

Understanding student bias as a variable affecting evaluation of teaching is important for developing the instrument to be used and in interpreting the data it produces. Further, in preparing the questions used in the instrument and in interpreting the data, it is now understood from studies on student evaluation of teaching that there are some aspects of faculty teaching that students are not equipped to evaluate. That said, in the absence of formal and careful evaluation

---

42 Hence, the overview on pages 4-9 of this report.
43 As noted, Hui Niu, the project team member leading the development of the instrument, will discuss findings in the research literature on instruments used for student evaluation of teaching.
44 Inter-rater reliability refers to the consistency of results from different raters using the same instrument. Re-test reliability refers to the consistency of results from the same rater with the same instrument but in multiple tests. Internal consistency refers to items within an instrument being consistent in what they measure.
45 See Chen and Hoshower 2003, p. 72.
46 Khandelwal, for example, in a review of the literature notes that student evaluation of teaching can be affected by “students’ prior motivation or desire to take the course (Marsh, 1984), anticipated grades (Howard & Maxwell, 1982), workload (Greenwald & Gillmore, 1997; Marsh and Roche, 2000), course level (positive relationship - Braskamp et al., 1985), class size (negative relationship - Cashin & Slawson, 1977; Smith & Glass, 1980), and grading leniency of the instructor (Greenwald, 1997). Cashin’s review (1988) concluded that student motivation (willingness to participate actively in the learning process) has the greatest positive influence on student satisfaction than any other instructional factor like grade expectations, sex of teacher/student, age of teacher/student, time of day, etc.” (cited in Khandelwal 2009, p. 300).
47 The list that Pallett provides is representative of what is found elsewhere in the research literature: [1] The appropriateness of an instructor’s objectives; [2] The instructor’s knowledge of the subject matter; [3] The degree to which instructional processes and materials are current, balanced, and relevant to objectives; [4] The quality and appropriateness of assessment methods; [5] The appropriateness of grading standards; [6] The instructor’s support for department teaching efforts such as curriculum
of teaching, faculty can or “will be evaluated by hearsay evidence, gossip, an other shoddy means.”48 (p. 7).

It is perhaps useful in SFU’s endeavours to update the form and process for student evaluation of teaching and courses to be reminded that:

[P]ractically no campus is satisfied with the protocols it has in place for evaluating teaching. The quest for perfecting the teaching evaluation process is quite likely to be doomed to be unending as we fuss and fiddle with the terminology, the question wording, and compare strategies. But it is in the fussing and fiddling that we move toward a better understanding of what goes on in the college and university classroom.49

Keeping in mind that although a perfect process for the evaluation of teaching and courses may not be an attainable goal, members of the SFU community are encouraged to hold in sight the purpose and promise of the project, which is to support faculty, administrators, and students in creating an environment that is rewarding and conducive to successful teaching and learning.50

50 Although the purpose of this report is to support SFU faculty, administrators, and students, it is worthwhile to note that in the larger context of research on post-secondary education, universities have been criticized as having “forgotten their purpose; namely, creating educated adults who will take responsibility for their society” (Harris & Cullen 2008, p.57). The same authors, citing Bok, echo a widely shared concern about faculty as teachers, “that not enough attention is paid to pedagogy” (2008, p. 58), as faculty training tends to focus heavily on disciplinary knowledge and research. By the same token, in regard to students as learners, Meyer’s view is that part of the challenge for student learning “lay in the fact that most entering university students are not likely to have had an opportunity to talk to their teachers (or anybody else) at school about how they typically went about learning in process terms. For the most part, students have never really thought about themselves in this way and, if asked, they generally experience difficulty, beyond habitual or preferential activities, in describing what they know about themselves as learners” (2010, p. 200, italics in original).
EVALUATION OF TEACHING AND COURSES

Hardré and Cox (among others) offer the reminder that “faculty performance is critical to the health of institutions of higher education and to the education of citizens.” It is also important, however, that evaluation of teaching and courses is something “done with” rather than “done to” those involved in teaching. It is necessary, therefore, for SFU faculty to be provided with multiple opportunities to be involved in updating the process and form used for student evaluation of teaching and courses at the university. Similarly, the SFU community is encouraged to consult this report for an overview of findings in the research literature to support their involvement in the development and responsible use of the revised instrument.

One of the key findings that surfaces repeatedly in the research literature is the importance of using a set of evaluative processes. As mentioned earlier in this report, there are several approaches that are commonly used in the evaluation of teaching. Although the focus of this report is on student evaluation of teaching and courses, some attention is given here to two approaches to teaching evaluation that are most often used to complement student evaluation of teaching and courses: (1) peer evaluation of teaching, and (2) teaching portfolios. It should be noted, however, that there is a need for validity of “all indicators of teaching effectiveness, not just SETs, [to] be systematically examined before they are actually used.”

PEER EVALUATION OF TEACHING

Peer evaluation of teaching is most commonly conducted through classroom observations by faculty in the same department as the faculty member being evaluated. Peer evaluation of teaching can be valuable because there are clearly aspects of teaching that students cannot be expected to be able to assess, such as “mastery of content, course goals, course organization and materials.” Evaluation by peers is also understood as fundamental to academic work, as faculty are accustomed to peer review in submitting material to journals and scholarly publishers, for example. As Paulsen notes:

Faculty expect public review of the methods and products of their research. In contrast, methods and products of teaching are rarely discussed or shared with peers. Just as the quality

---

51 See Hardré and Cox 2009, p. 383. In discussing the importance of evaluating faculty work, the authors also cite Fairweather (1996) and Gappa, Austin and Trice (2007).
54 See Marsh 2007, p. 343.
55 See footnote 47.
of research improves due to dialogue and debate among disciplinary peers, so would the quality of teaching benefit from similar opportunities.57

As with peer review for publication, however, there are challenges to overcome in terms of the ability and willingness of a peer to provide informed, fair, and helpful evaluation. That said, peer evaluation of teaching, which has for some time been common in the UK, is seen by much of the teaching faculty, “as an integral part of their own professional development as teachers and [they] see value from the process, both in engaging as the observer and the one being observed.”58 Peer observation and evaluation of teaching offers the following:

- insight into what helps learners to learn and what happens in effective teaching sessions;
- feedback on individual teaching skills and style;
- discussion, collaboration, and the exchange of ideas;
- mutual support between colleagues;
- earmarked ‘quality time’ to talk about learning and teaching;
- feedback on piloting a new idea, method, or solution to a problem;
- triangulation with other evaluative procedures, e.g. student perception questionnaires, module and program evaluations;
- focused reflection on teaching sessions;
- the opportunity to see exemplary practitioners at work.59

Research on peer evaluation of teaching also provides useful guidelines. According to Harris and Cullen, for example, the “three components that should frame the overall evaluation are: [1] teaching philosophy or pedagogical stance; [2] competency in the discipline; and [3] teaching craft.60 Although a thorough overview of peer evaluation of teaching is beyond the scope of this report, it is an important component in the evaluation of teaching and courses, and therefore warrants further attention as a complement to student evaluation of teaching and courses.

TEACHING PORTFOLIOS

Teaching portfolios were first used in Canada in the 1970s and have since become widely used across North America.61 Items in a teaching portfolio can include the following:

- courses, modules and units taught;

57 See Paulsen 2002, p. 10. Paulsen notes that, “Specialists in teaching and its evaluation also agree that the work of an individual faculty member is valued more when it has been subjected to rigorous peer review” and suggests that this is a contributing factor to research being valued more highly than teaching (2002, p. 10).
58 See Fullerton 1999, p. 220.
59 See Fullerton 1999, p. 221-222.
60 See Harris & Cullen 2008, p. 62.
61 Citing Kappa (1995), Fry and Ketteridge identify Queen's University as a “one of the pioneers of the teaching portfolio” (1999, p. 235).
- information about the context of your teaching;
- self-authored documents and/or study guides;
- module descriptions including aims and objectives;
- curriculum development work;
- problem-based learning cases;
- AV materials;
- teaching strategies;
- student handouts or workbooks;
- documentation to support practical teaching;
- instruments for assessment of student learning;
- analysis of examination results;
- student evaluation instruments, views and results;
- quality assurance reports;
- external examiners’ reports;
- examples of student work;
- video, audio and photographic material;
- computing and IT material;
- written commentary from observed teaching sessions;
- details of courses and workshops attended on teaching and learning;
- educational publications.\(^{62}\)

A teaching portfolio serves as a “personal record of achievement and professional development” and contains a careful selection of material collected and developed over the course of a faculty member’s career and is a way of demonstrating “a level of attainment, progression, professional development and/or achievement.”\(^{63}\) Teaching portfolios (referred to also as “teaching dossiers”) have also become required in applications for faculty positions and are developed as a component of faculty earning credentials in higher education learning and teaching, which has become common in the UK.\(^{64}\) A detailed overview of teaching portfolios is not within the scope of this report but teaching portfolios also warrant further attention.

\(^{63}\) See Fry and Ketteridge 1999, p. 235.
\(^{64}\) See The Higher Education Academy in the UK, for example (http://www.heacademy.ac.uk/ukpsf): “The UK Professional Standards Framework (UKPSF) for teaching and supporting learning is for institutions to apply to their professional development programmes and activities thereby demonstrating that professional standards are being met. We support and guide both institutions and individuals as they engage with the UKPSF...The UK Professional Standards Framework provides a general description of the main dimensions of the roles of teaching and supporting learning within the HE environment. It is written from the perspective of the practitioner and outlines a national Framework for comprehensively recognising and benchmarking teaching and learning support roles within Higher Education.”
STUDENT EVALUATION OF TEACHING AND COURSES

Although the sections to follow in this document provide details of key concerns and research findings in regard to particular aspects of student evaluation of teaching and courses, the widely-cited work of Centra provides a useful overview of the historical trajectory of student evaluation of teaching and courses, and is therefore worth quoting at length:

In the universities of medieval Europe...students evaluated teachers in two ways. First, a committee of students appointed by the rector made sure they covered each topic by the date specified in advance; any irregularities were reported to the rector, who fined the professor for each day he (in those days, always a “he”) had fallen behind. Second, students paid fees directly to each teacher (called the collecta). Teachers’ salaries, therefore, were determined by the number of students attending their classes (Rashdall, 1936). Today’s student questionnaires seem tame by comparison.

The modern era of student evaluations can be broken roughly into four periods: the thirty-year period preceding 1960, the 1960s, the 1970s, and the period from the 1980s to the present. Before 1960, most of the research on student evaluations was conducted by Herman Remmers and his colleagues at Purdue University...Although Remmers and other researcher...promoted their use, student evaluations were rarely used in colleges and universities until the 1960s. The student protests that rocked so many [US] campuses in the last half of the decade were in reaction not only to the Vietnam War and related national policies but also to policies in effect on their campuses. An irrelevant curriculum and uninspired teachers were among frequently heard student complaints...At many universities, students administered their own rating systems—often haphazardly—and published the results. More often, institutions began to develop their own student evaluation systems, which faculty members could use if they wished...

In the 1960s, faculty members usually volunteered to administer the form in their courses themselves, so that they might make necessary improvements on their own. Administrative use of the results was infrequent because few institutions had centrally administered student rating systems.

Interest in pertinent issues—the validity of the ratings, in particular—inspired a new wave of research and the 1970s became the golden age of research on student evaluations. Well designed studies investigated questions of bias, validity (Do student evaluations really measure teaching effectiveness?), and utility (Can they help improve instruction?). The generally favorable findings helped support the use of the evaluations for tenure and promotion decisions as well as for institutional improvement. During the 1970s, more and more institutions adopted them, with a majority reporting their use by the end of the decade (Centra, 1979). Department heads ranked student evaluations among the top three sources of information on teaching effectiveness (colleagues and department [end of pg 50] heads were the other two) and, moreover, believed them to be the most important source (Centra, 1979).

The fourth period, from the early 1980s to the present, has included the continuing refinement of the research findings, and a series of reviews and meta-analysis has substantiated the findings on important issues.65

With this historical context in mind, the sections that follow in this report take up issues of student evaluation of teaching and courses that are particular to the concerns of faculty, administrators, and students.

STUDENT EVALUATION OF TEACHING AND COURSES: ISSUES FOR FACULTY, ADMINISTRATORS, AND STUDENTS

One criticism (shared by faculty, administrators, and students) of student evaluation of teaching and courses is that, if used ineffectively, it serves as little more than “a ‘fire alarm’ function.”\(^{66}\) If developed and used well, however, there are also advantages to student evaluation of teaching: “[1] Students know best what was effective for their learning; [2] Students observe the whole class; [3] Student samples are larger [than is the case with peer evaluation].”\(^{67}\) It is also worth noting that, “there are generally high correlations between students’ and faculty members’ evaluations of teaching.”\(^{68}\)

In order for the instrument used at SFU for student evaluation of teaching and courses to be as effective as possible in supporting faculty teaching and student learning, it is important for the SFU community to be aware of the issues and challenges particular to faculty, administrators, and students in using a student evaluation of teaching instrument and in interpreting the data it produces. This section of the report summarizes key findings in the research literature in order to support effective and responsible student evaluation of teaching and courses at SFU.

ISSUES FOR FACULTY

In the literature on student evaluation of teaching and courses, Ackerman \textit{et al} identify the two main concerns by faculty as centering around: (1) validity (and the extent to which evaluations are affected by student bias); and (2) the use of data in tenure and promotion, given “a lack of sophistication among personnel committees who use student ratings.”\(^{69}\) These two concerns are discussed in this report, but first, it is worth mentioning common perceptions (and misperceptions) among faculty about student evaluation of teaching and courses reported in the research literature, as they provide an indication of what is often the starting point (and first obstacle) in working toward effective and responsible use of student evaluation of teaching.

\(^{66}\) See Edström 2008, p. 95.
\(^{67}\) See Ackerman \textit{et al} 2009, p. 27.
\(^{68}\) That said, the authors also note that: “Student and faculty views of one another depend on how much they agree on the characteristics of excellent instructors (Goldstein & Benassi, 2006). Interviews with students reveal the attributes most associated by students with outstanding teaching to be rapport, delivery, fairness, knowledge and credibility, organization, and preparation (Faranda & Clarke, 2004)…In addition, some dimensions valued by faculty members such as facilitating the achievement of key learning objectives and modeling rigorous thinking may be unrecognized by students (Buskist, Sikorski, Buckley, & Saville, 2002; Schaeffer, Epting, Zinn, & Buskist, 2003).” See Ackerman \textit{et al} 2009, p. 22. With this, there is perhaps further support for the “meta-learning” suggested by Meyer (2010), the benefits of reflective teaching practice widely supported by researchers in the field of higher education, and the significance of communication between faculty and students about pedagogical approaches being taken.
Student evaluation of teaching and courses is sometimes seen by faculty as a “popularity contest…designed to ‘win votes’…[using a] surveillance approach in relation to students’ expectations [that] leads to outcomes that are not always educationally sound.”\textsuperscript{70} This view presents valid concerns that have been taken up in the research literature. For example, the “popularity” of a faculty member among students (and also among other faculty)\textsuperscript{71} can affect ratings and faculty members have been known to employ tactics to “win” votes.\textsuperscript{72} In terms of unsound outcomes, one of the most widely cited is unspoken leniency in grading traded for higher student evaluation of teaching and courses. These perceptions of flaws (and abuses) in student evaluation of teaching and courses are valid. Faculty misperceptions, however, rest in dismissing student evaluation of teaching and courses altogether, rather than taking into account that: (1) the instrument can be designed to reduce “noise” (which can be expected in any evaluation method) through careful crafting of the questions and options for responses; and (2) guidelines on interpreting the data can provide considerable support for responsible, effective and instructive use of student evaluation of teaching and courses.

**Recognizing possible student bias**

Before presenting an overview of sources of student bias discussed in the research literature, one key concern noted earlier in this report should be reiterated: a consumerist view of higher education is problematic in that it can result in student bias that reduces the learning and the experience of higher education to a product that is purchased by tuition dollars and for which students (as consumers) can be inclined to take a “customer satisfaction” approach to teaching and learning. This is problematic because a consumerist view of higher education tends to give little attention to such things as the level of effort and commitment required of students, the difficulty and the demands of the work to be undertaken by students, or in short, that although there are different views on the purpose(s) of higher education, post-secondary studies are, after all, meant to be challenging (and this does not always bode well for “customer satisfaction”).\textsuperscript{73}

This is not to dismiss the importance of ensuring quality in higher education. Indeed, maintaining quality standards is fundamental to the strength of higher education and the learning opportunities that colleges and universities provide. The issue is one, however, that deserves attention and discussion between faculty and students in the interest of supporting faculty teaching and student learning, and the goals of higher education more broadly.

\textsuperscript{70} See Gapp and Fisher 2006, p. 157.
\textsuperscript{71} This is discussed below in the subsection on student bias.
\textsuperscript{72} See Halonen and Ellenberg 2006, p. 154-156. The authors provide the following list of what they refer to as “faculty follies:” (1) the “forget” to conduct the evaluation in their course(s); (2) they “taint” the evaluation by failing to provide enough time; (3) they “pander” (by handing out “treats” on evaluation day; (4) they “wheedle” (guilt-tripping students by noting that the evaluations affect their salaries); (5) they “script” (by priming students in how to respond on the evaluation), (6) they “barter, linger, over-respond, denigrate, ignore” (based on the nasty or overly harsh evaluations from a minority of students). The authors also identify “student follies” and “administrative follies” (these are presented later in this report).
\textsuperscript{73} Germain and Scandura also note, for example, that “Consumerism results in bias due to information not relevant to teaching competency, but important to students such as textbook cost, attendance policy, and the amount of homework” (2005, p. 58).
According to the research literature, beyond the effects of a consumerist perspective, other factors that can bias student evaluation of teaching and courses include: “prior subject interest, expected grade and actual grade, reason for taking a course, workload and difficulty, class size, level of course or year in school, instructor’s rank, sex of instructor or student, academic discipline, purpose of ratings, administrative conditions, and students’ personality.” Views on student bias are not uncontested, however. Marsh (2007), for example, cautions that the “literature on potential biases in SETs is frequently atheoretical, methodologically flawed, and not based on well-articulated operational definitions of bias” (p. 346) and he offers the following definition: “Bias exists when a student, teacher, or course characteristic affects the evaluations made, either positively or negatively, but is unrelated to any criteria of good teaching, such as increased student learning.” The question is not whether the characteristics of the student, teacher or course have affects on evaluations – they inevitably do – but rather, the difficulty is in determining whether or not the effects of these characteristics (of the student, teacher, or course) on the evaluations are related or unrelated to the criteria of “good teaching” put forward in SET forms. Studies on student bias in evaluating teaching and courses have suggested, for example, that:

- elective courses tend to receive higher ratings than required courses;
- upper level courses tend to receive higher ratings than lower level courses;
- the time of day that a course is offered can affect student ratings;
- the relationship between class size and ratings from student evaluation of teaching is debated, though smaller classes are more conducive to building rapport;
- there is a relationship between gender of the student, gender of the instructor, and ranking that the student gives the instructor;
- ethnic/cultural background and beliefs can also affect student evaluation of teaching.

---

76 It should be noted here that the list provided is not comprehensive, but rather, reflects areas given the most focus in the research literature. It should also be noted that, as in other areas of research, findings are not always conclusive and different studies can provide conflicting results. That said, care and attention have been given here, in review of the research literature, to report on key findings that can support effective and responsible development and use of student evaluation of teaching and courses.
77 See Scherr and Scherr (1990), Marsh and Dunkin (1992), and Pounder (2007).
78 See, for example, Braskamp et al 1985, March 1987, and Buchert et al 2008.
79 See Zabaleta, for example, whose study showed that, “instructors who teach between 10 am and noon receive the best evaluations” (2007, p. 61). See also, Husbands and Fosh (1993), as cited in Denson et al (2010).
81 See Pounder 2007.
82 As Hobson and Talbot note, “A heated debate continues about the influence on student evaluations of a teacher’s sex and gender-role orientation (masculine, feminine, androgynous, or undifferentiated) (Basow and Howe 1987; Hobson 1997; Sidanius and Crane 1989). With respect to student characteristics, an interaction effect appears to exist between students and instructor’s sex/gender-role orientation” (2001). For example, Centra and Gaubatz, in their study of 741 classes, investigating male and female students in their evaluation of male and female instructors, found that “female instructors received a lower rating from male students” on several scales, while there was no significant difference in how the students of both genders evaluated male instructors” (2000, p. 26).
83 See, for example, Germain and Scandura: “Cultural beliefs strongly influence the value and behavior of the people who grew up in the culture, often without their being aware of it…More specifically, a student’s cultural background can influence how
• socioeconomic status of students can affect student evaluation of teaching;\textsuperscript{84}
• generation gap issues need to be taken into account as a possible source of student bias;\textsuperscript{85}
• there is debate about instructor rank and student ratings, with evidence suggesting only slightly higher student evaluation of teaching for more senior faculty;\textsuperscript{86}
• pre-existing knowledge about the instructor can affect student evaluation of teaching;\textsuperscript{87}
• although there is ample research on discrimination against gay, lesbian, bi-sexual, and transgender persons, research specific to student evaluation of teaching is limited;\textsuperscript{88}
• instructor characteristics, including appearance\textsuperscript{89} and charisma, can affect ratings;\textsuperscript{90}
• differences across academic disciplines can affect student evaluation of teaching;\textsuperscript{91}

they react to their learning environment. The effect of cultural context has long been recognized as having an affect on student learning (Halloway, 1988)\textsuperscript{84} (2005, p. 63-64). Similarly, Al-issa and Suleiman note that, “students’ cultural and linguistic backgrounds affect their responses to SET [student evaluation of teaching]” (2007, p. 302). Smith also reports from a study of 31,768 evaluation forms that for “overall value of course and overall teaching ability…Black male faculty received the lowest mean score; Black female faculty received the second lowest mean scores (2008, p. 622)

\textsuperscript{84} See Germain and Scandura (2005): “Socioeconomic status may also affect how students relate to faculty. From an economic standpoint, wealth or hardship of a student could affect grades obtained in the class and subsequent evaluations. Students who are on grants or scholarships are obligated to maintain A’s or B’s to maintain their benefits…and fearing the loss of financial aid might make students grade professors that give C’s on course assignments more harshly” (p. 63).

\textsuperscript{85} See Germain and Scandura (2005): “More mature students may be less judgmental with older faculty members. Similarly, younger students may be more likely to give good evaluations to young faculty members” (p. 63). Similarly, Sprinkle reports that older students are “more likely to take responsibility for their own learning and grade, rather than place the burden upon the professor/instructor (Nunn, 1994)."[and] younger students are likely to believe that younger professors/instructors are easier to relate to and more understanding of/sympathetic to their difficulties” (p. 286).

\textsuperscript{86} See, for example, Marsh and Dunkin (1992) and Kogan et al (2010).

\textsuperscript{87} See Germain and Scandura (2005): “[T]he degree to which the instructor is well known on campus or whether a student has had the instructor in a previous course…may affect rapport during the semester. Because the student has pre-existing knowledge about the instructor, faculty evaluations may be biased compared with students who have not had the professor previously” (p. 63). There are similar findings in Griffin (2001), who also cites studies by Kelley (1950) and Widmeyer and Loy (1988): “Students prompted prior to viewing a lecture that an instructor was either ‘very warm’ or ‘rather cold’ as a person…were enough to affect student ratings of the instructor in a positive or negative direction” (p. 535). That said, in a study by Buchert et al (2009), “students considered first impressions more important than professor reputation as determinants of their end-of-the semester evaluations…it appears that students form lasting impressions of faculty during the first 2 weeks of classes” (p. 397).

\textsuperscript{88} See Jennings (2010), for example: “Experimental projects have concluded that student assessments of speaker credibility were negatively affected when students perceived a male speaker as homosexual compared with when the speaker was perceived as heterosexual. Further, students believed that they learned less from homosexual male speakers than from those who were not perceived as homosexual (Russ, Simonds, and Hunt 2002). Ewing, Stukas, and Sheehan (2003) found that under some experimental circumstances students rated gay men and lesbian guest speakers lower than speakers whose sexual orientation was unspecified. However…these experimental studies only exposed students to single lectures by guest speakers. They did not study students’ evaluations of instructors who were actually teaching the courses and with whom students had repeated contact, experienced multiple opportunities to observe instructors’ full range of teaching behaviours, or came to know the instructor personally. Only one project has directly examined the relationship between instructor self-disclosure and student evaluations in a naturally occurring university class (Liddle 1997). In contrast to the two experimental projects by Russ, Simonds, and Hunt and by Ewing, Stukas, and Sheehan, Liddle’s research analysed actual classroom evaluations of a course instructor and concluded that student evaluations of the instructor were not affected by instructor self-disclosure” (p. 327).

\textsuperscript{89} See, for example, Morris \textit{et al}, whose study indicated that, “perceptions of instructor competence were highest…[with] formal professional attire…with casual professional attire a close second…and the lowest ratings for…casual attire” (1996, p.143). There is also a substantial research literature on the effects of appearance on likability, and as Gurung and Vespia note from their study, “By far the strongest indicator of self-reported learning [which is a common question in student evaluation of teaching and courses]…was the likability of the professor. Likability, in turn, was predicted by instructor attractiveness, approachability, and formality of dress” (2007, p. 8).

\textsuperscript{90} The most widely cited example of student response to instructor charisma or “enthusiasm” in evaluating teaching and courses is what has become known as the “Dr. Fox” experiment in which a trained actor was “coached to present his topic and conduct his question and answer period with an excessive use of double talk, neologisms, non sequiturs, and contradictory statements. All this was to be interspersed with parenthetical humor and meaningless references to unrelated topics” (Naftulin, Ware, and Donnelly (1973). In the experiment, Dr. Fox received higher ratings from students than actual instructors.
courses with a heavy workload tend to receive lower ratings than courses with a workload that students perceive as reasonable for the course level;\textsuperscript{92} whether the student is part-time or full-time can affect ratings;\textsuperscript{93} perception about the purpose of the evaluation can affect ratings;\textsuperscript{94} and there are issues to consider in regard to international students.\textsuperscript{95}

In addition to the above list of possible sources of student bias in evaluation of teaching and courses, one further concern commonly raised in the research literature is the relationship between students’ expected grades (and/or actual grades) and their evaluation of teaching and courses.\textsuperscript{96} In particular, the concern is in regard to the high correlation between grades and evaluation of teaching, for which the following three explanations are commonly given:

The validity explanation is that grades and evaluations are correlated because students learn more from superior teachers. In contrast, the leniency explanation is that students reward professors with high evaluations in exchange for high grades. Finally, preexisting student interest in course topics may explain both grades and teaching evaluations. Research on this topic continues, but clear evidence for one explanation has not emerged.\textsuperscript{97}

Another perspective on the relationship between grades and ratings is “the cognitive dissonance explanation [that] states that students attempting to reduce the dissonance caused by a poor grade are more likely to denigrate the teacher than their own ability, which leads to lower evaluations.”\textsuperscript{98} Aside from concerns about the damage done to the integrity of assessment of teaching by an unspoken trade of high ratings on teaching and course evaluations for high grades, at issue here also is a broader concern about grade inflation in universities.\textsuperscript{99}

\textsuperscript{91} See, for example, Erdle and Murray’s (1986) study on varying “teaching behaviours” varying between arts, social sciences, and natural sciences. Aside from student bias in response to teaching behaviours (the authors differentiate between “interpersonal orientation” and “task orientation”), there are also considerations here for the items included in the evaluation form and options for customization across faculties and departments.

\textsuperscript{92} See, for example Marsh (1982), Butdsal and Bardo (1986), Jackson et al (1999), and Marsh and Roche (2000).

\textsuperscript{93} See Denson et al, for example, who found in their study that, “part-time students rated satisfaction with quality of the course lower than full-time students” (2010, p. 346).

\textsuperscript{94} Gapp and Fisher (2006) describe, for example, that an instructor’s view on evaluation of teaching and courses can transfer to the students. Al-Issa and Sulieman (2007) note from their study that, “students believed that their assessments were an effective means of voicing their opinions about teaching, but that they were not fully aware of the implications of their evaluations for university administrators and teachers. This raises the question of whether students were motivated to take the evaluation seriously” (p. 304). Further, Burden notes that the evaluation process can be improved by “students learning how to do evaluations, or becoming ‘more sophisticated evaluators’” (2010, p. 107, citing McKeachie 1997, p. 1223, italics in original). A study by Oliver et al (2008) provides an excellent example of revising (and clarifying) questions on a student evaluation form to support students’ understanding of what they are asked to evaluate.

\textsuperscript{95} Other than ensuring that students (international or otherwise) fully comprehend the questions on the evaluation form, Burden (2010) notes also, for example, that, “If students have…a positive attitude toward foreign countries, then it is easy for students to ‘tick all the boxes down the right (i.e. choose a rating of 5 [out of 5]) without even reading the questions” (p. 105). Similarly, Al-Issa notes that students from some cultural backgrounds may not be “accustomed to ‘passing judgment’ on their teachers” (2007, p. 303).

\textsuperscript{96} See Howard and Maxwell (1982), Olshavsky and Spreng (1995), McKeachie (1997), and Eiszler (2002), among others.

\textsuperscript{97} See Boysen 2008, p. 218, citing the work of Marsh (1984).

\textsuperscript{98} See Boysen 2008, p. 218. As noted earlier, however, instructors are not immune to cognitive dissonance.

\textsuperscript{99} As Germain and Scandura note, “Grade inflation has become an issue in higher education; students' grades have been steadily increasing since the 1960’s (Astin, 1998). In June 2001, a record 91 percent of Harvard seniors graduated with honors, and 48.5 percent of grades were A's and A-minuses (2005, p. 58).
Awareness of possible bias in student evaluation of teaching and courses and knowledge of the research in this area can support the SFU community and the project team as the instrument for student evaluation of teaching and courses is updated, both in terms of deciding on questions to include and in the interpretation of the data that the instrument produces. Although details of the instrument itself are discussed in another report,\textsuperscript{100} the studies on possible areas for student bias are useful for considering, for example, the inclusion of a question on a student’s reason for taking a particular course. Similarly, in interpreting the data, taking into account, for example, variables ranging from course level, to gender, to disciplinary differences, are also important. Concerns about interpreting and using the data produced by student evaluation of teaching and courses tend to focus on issues for administrators or the university administration more broadly. This report therefore turns to key findings in the research on these issues, which include administrator interpretation and use of data derived from student evaluation of teaching (including issues of validity and reliability), support for teaching and learning, and broader concerns for the university administration.

**ISSUES FOR ADMINISTRATORS AND THE UNIVERSITY ADMINISTRATION**

One of the key concerns that faculty have in the use of student evaluation of teaching is about the possible bias of administrators in interpreting data produced by rating instruments and the actions taken (or in some cases, not taken) based on interpretation of the data. The focus in the research literature on administrator bias primarily concerns the use of student evaluation of teaching data in tenure and promotion decisions. The particulars of these concerns are discussed below, but in summary, the key administrative issues that faculty are concerned about involve the capacity (and sometimes the willingness) of deans, departmental chairs and peers on tenure and promotion committees to provide informed and fair assessment of teaching that is based on student evaluation of teaching. While deans, departmental chairs and peer committee members are knowledgeable in their own field(s) of research, they do not always have depth of knowledge in the area of student evaluation of teaching, or more broadly in the area of pedagogy and student learning in higher education. A second concern involves the capacity and commitment of the university administration to use the data produced by student evaluation to provide training and development that supports faculty teaching. While university mission statements commonly identify teaching and learning as priorities, not all universities commit resources and expertise to this part of their mission, and instead privilege research. As a consequence, the data derived from student evaluation of teaching is not effectively used and there is either little incentive or opportunity for faculty development in their teaching.

If student evaluation of teaching and courses is to prove valuable, as Gapp and Fisher note, “first and most critical…[is] the establishment of a safe and trusting environment for both staff and

\textsuperscript{100} Hui Niu, the project team member leading the development of the instrument, is also providing a report.
students,"\(^{101}\) in which “relevant and useful input from participants…leaves all involved feeling empowered within the process.”\(^{102}\) It is therefore important that there are multiple opportunities for faculty, administrators, and students to be involved in updating the process and instrument for student evaluation of teaching and courses. Also necessary, however, is a shared understanding of the importance of teaching and learning as fundamentally valued by the university, not only in formal documents such as the university mission statement, but also in how faculty and students are supported and engaged in teaching and learning at the level of the university’s courses, programs, and departments.\(^{103}\) This can be challenging for the university administration, however, as:

Faculty organisational commitment in higher education has declined over time (Cintrón 1999; Judy and D’Amico 1997) and that diminished commitment influences retention and tenure (Werbel and Gould 1984), motivation and involvement (Mowday, Porter, and Steers 1982; Smith, Organ, and Near 1983), task performance and compliance with organisational policy (Angle and Perry 1981) and adoption of organisational values and priorities (O’Reilly and Chatman 1986). For these reasons, it is important to examine the documents that communicate evaluative processes, task expectations and performance standards, in light of the messages they may contain and how these can influence employees’ thinking about their work (Braskamp and Ory 1994; Fairweather 1999; Schön 1983).\(^{104}\)

Although the above concerns may be valid, the same authors also note the need for support from the university administration for design and delivery of courses.\(^{105}\) Further, support is needed from the university administration to recognize and address what Halonen and Ellenberg refer to as “administrative follies” in student evaluation of teaching:

- **Neglect:** failing to ensure validity, reliability, and training for proper use;
- **Constraint:** placing too much weight on student ratings;
- **Confusion:** using a single item as total assessment of performance;
- **Obfuscation:** using unclear questions in the form used for student evaluation of teaching;
- **Jeopardizing:** failing to take adequate measures to protect confidentiality;
- **Delay:** returning results of student evaluation of teaching to faculty at a date too late for the next offering of the course;

\(^{102}\) See Gapp and Fisher 2006, p. 158.  The authors also note, however, that “No system or method is perfect and innovation and improvement are part of a normal creative improvement process” (2006, p. 161).  
\(^{103}\) As Gapp and Fisher (2006) note, the approach of the university administration can sometimes be limited, unfortunately, to an emphasis on student retention rather than engagement with teaching and learning. That said, there has for many years in the field of higher education been discussion of a “discrepancy between the attention that the teacher [and the university more broadly] pays to the students professional training and the importance that the student should acquire professional skills” and this has implications not only for student satisfaction, but also for changing views on the purpose(s) of higher education (Ghedin and Aquario 2008, p. 593).  
\(^{104}\) See Hardré and Cox 2009, p. 387.  
\(^{105}\) Also, as Marsh (2007) suggests, “ratings for a given instructor should be averaged across different courses to enhance generalizability. If it is likely that an instructor will teach many different classes during his or her subsequent career, then tenure decisions should be based upon as many courses as possible…[and] that a longitudinal archive of SETs is maintained for personnel decisions” (p. 335).
Quantifying: accepting figures from student evaluation of teaching as conclusive;
Assumptions: not questioning perspectives on or reasons for ratings;
Succumbing: giving in to somewhat simplistic notions of student satisfaction;
Shirking: not recognizing that other activities (such as peer observation and teaching portfolios) are needed to accompany the use of student evaluation of teaching;
Normalizing: expecting most faculty to be average and that few will be excellent; and
Stopping short: failing to help faculty make sense of the data produced by the student evaluation of teaching instrument.106

The above list is not meant to reflect the *modus operandi* of university administrators (whether deans, department chairs, or peer reviewers in their administrative activities) or university administration more broadly. It is meant, however, to caution the university community about common administrative flaws in the use of student evaluation of teaching. Effective and responsible use of student evaluation of teaching, it is clear, requires that staff undertaking administrative roles in the process are familiar with and sensitive to a number of issues.107 Further, as Pallett notes, regrettably, “so much emphasis on the summative component of student ratings—getting the right number—[can have the consequence] that what can be learned to improve teaching is often overlooked.”108

Support for faculty career development in teaching activities is one of the chief concerns taken up in the literature in regard to administrative issues. Beyond the responsible and informed collection and interpretation of data produced from student evaluation of teaching (complemented by information from other sources on evaluation of teaching), “for real gains in teaching skill to occur, support and mentoring needs to be provided” to faculty.109 As Khandelwal observes, for example, in the “research on student evaluations of teaching effectiveness…that most studies identify characteristics instead of behaviors reduces their utility. Merely knowing that a particular characteristic of effective teaching exists does not tell a teacher how to enact that characteristic.”110 Universities provide different kinds of support for faculty

---

106 See Halonen and Ellenberg 2006, p. 156-158. In addition, Pallett identifies the following "abuses or misuses of student evaluation" to be aware of and avoid: Abuse 1: Overreliance on student ratings in the evaluation of teaching....Abuse 2: Making too much of too little....While there is substantial evidence that student ratings are reliable, there is always some "noise" in survey data....Abuse 3: Not enough information to make an accurate judgment....Abuse 4: Questionable administrative procedures....If student ratings are taken seriously by faculty and administrators, it is likely that students will take them seriously as well....Abuse 5: using the instrument (or the data collected) inappropriately....Abuse 6: Insufficient attention to selecting/developing an instrument....Abuse 7: Failure to conduct research to support the validity and reliability of a student ratings tool" (2006, p. 56-60).

107 As Harris and Cullen note, for example, administrators “need to be sensitive to issues of power and control, about attitudes toward learning (both student and teacher), and to the many forms of assessment and evaluation and their uses in various contexts. They will need to know the language for writing learning objectives and learning outcomes. They will need to be familiar with techniques for active lecturing, problem-based learning, concept mapping, effective discussion techniques, and more” (2008, p. 61).


110 See Khandelwal 2009, p. 300. Khandelwal elaborates thus: “For example, knowing that rapport is a common characteristic of excellent teaching does not provide any detail of the behaviors that convey or communicate rapport. There is more potential for improvement of teaching if behavioral items such as ‘Uses humor’ or ‘Gives real life examples’ are used, instead of abstract, general statements like ‘good presentation’” (2009, p. 300-301).
career development in their teaching and the data derived from student evaluation of teaching can prove useful for individual faculty members in seeking support in a particular area, as for the university administration in identifying where support is most needed for faculty.\textsuperscript{111}

Although a review of the literature on career development in faculty teaching activities is beyond the scope of this report, it is perhaps worth noting what Åkerlind identifies as “five qualitatively different approaches to developing as a teacher:

(1) Building up one’s content knowledge (improving what to teach).
(2) Building up practical experience (improving how to teach).
(3) Building up a repertoire of teaching strategies (becoming more skilful as a teacher).
(4) Finding out what strategies work for the teacher (becoming more effective as a teacher).
(5) Increasing one’s understanding of what works for the students (becoming more effective in facilitating student learning).\textsuperscript{112}

In the above list, as in the brief discussion of the shift to learner-centred curriculum earlier in this report, it becomes clear that in thinking about and updating the instrument for student evaluation of teaching, there are also student issues that need to be understood. As the research literature makes evident, it is important to involve students in the discussions and the process of updating the instrument used for student evaluation of teaching. This report therefore includes, in the section to follow, an overview of findings from the research literature that can support students in their involvement in the project to update SFU’s student evaluation of teaching instrument.

ISSUES FOR STUDENTS

If, as the research literature suggests, “student ratings are the most, if not the only, influential measure of teaching effectiveness, active participation by and meaningful input from students can be critical in the success of such teaching evaluation systems.”\textsuperscript{113} Gaining meaningful input from students, however, involves an understanding of student perspectives on evaluation of teaching. One of the most significant findings from the research on student perceptions of teaching evaluation is that “improvement in teaching” is the most valuable outcome of their participation in such evaluations.\textsuperscript{114} Although there is a substantial body of research on student

\textsuperscript{111} For the support provided to SFU faculty, see the SFU Teaching and Learning Centre: http://tlcentre.sfu.ca/. With the support of the Office of the Vice-President, Academic, SFU also offers teaching and learning grants to “recognize teaching development as a scholarly activity and to stimulate the development, implementation, and investigation of innovative teaching and learning” (http://www.sfu.ca/teachlearn/tlgrants.html).

\textsuperscript{112} See Edström 2007, p. 103, citing Åkerlind 2007.

\textsuperscript{113} See Chen and Hoshower 2003, p. 71. The authors also note that, “Several studies…have observed a significant link between student attitudes toward the evaluation of teaching effectiveness and the success of a teaching evaluation system (Hofman & Kremer, 1980; Marsh, 1984, 1987; Douglas & Carroll, 1987; Tom et al., 1990)” (2003, p. 72). Campbell and Bozeman (2008) also note from their study that students strongly believe that they should complete teaching evaluations.

\textsuperscript{114} See, for example, Chen and Hoshower (2003), Al-Issa and Sulieman (2007), and Algozzine \emph{et al} (2010). Algozzine \emph{et al} (2010) also note that the “least important use of teaching evaluations, from the students’ standpoint, is for the professor’s tenure,
views of teaching effectiveness in higher education, the title of an article by Remedios and Lieberman – “I liked your course because you taught me well” – captures an important finding in research on student views of teaching and teaching evaluation.115 Following from this, one of the key areas of discussion for students as they are involved in updating the instrument for student evaluation of teaching is about their views and understandings of effective teaching. Here, there are important links to the shift from a teacher-centred to a learner-centred classroom and curriculum, discussed earlier in this report.

The shift to a learner-centred classroom and curriculum presents challenges to students, as it does for teachers, given what often can be the limited understanding that students have of themselves as learners in this new context,116 as well as resistance that some students can have to what they have come to expect as the status quo of a teacher-centred classroom in which they are more passive than active participants.117 It is because of such challenges, shared by students and teachers (as by the university administration more broadly), that communication between faculty and students (and between administrators and faculty) about a diversity of teaching methods and defined learning outcomes is particularly important, as understandings about teaching and learning in higher education have changed with both research and the realities in classrooms across the disciplines. Arriving at shared understandings about the responsibilities of both faculty and students in teaching and learning is also important for efforts by the university to support innovation and diversity in teaching methods.118

Although a shared understanding between faculty and students about their respective roles (and responsibilities) in teaching and learning is invaluable, knowledge of the reasons students’ give for their high and low evaluations of teaching is useful for identifying student views on effective teaching. Most published studies on student evaluation of teaching include a rating scale of some kind and data derived from the scale(s). Two studies in particular are cited below, however, because they are not only consistent with general findings on student evaluation of teaching, but also, they identify reasons that students give for low ratings and positive descriptors used in student written comments on teaching effectiveness.

115 See Remedios and Lieberman 2009, p. 91. The full title of the article is, “I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students’ evaluation of teaching” and reports on the authors’ study in the UK. From their study, they conclude that, “Although grades and course difficulty did have a small influence on end-of-semester course ratings, structural modelling revealed that ratings were largely determined by how much students enjoyed or felt stimulated by the course content, which in turn depended on the perceived quality of teaching” (p. 91).

116 See, for example, Meyer (2010).

117 See, for example, Felder and Brent (1996) cited in Weimer (2002), and Edström (2008).

118 In keeping with the discussion by Weimer (and others, as cited above) about student resistance to change in higher education classrooms and curriculum, Struyven et al, for example, in their study on “students’ likes and dislikes regarding student-activating and lecture-based educational settings,” found that, “While lecture-taught students’ evaluations of the experienced teaching were generally focused and positive, students’ perceptions of the activating methods varied widely and both extremely positive and negative opinions were present” (2008, p. 295).
In his study on why students’ give poor evaluations of teaching, the following reasons (in descending order of frequency) were given:

- Poor teaching style or methods
- Rude, disrespectful, or uncaring
- Unprepared or unorganized
- Unfair in grading
- Unclear or poor communication skills
- Generally a bad professor
- Professor unknowledgeable
- Workload too heavy
- Did not learn enough
- Professor was unavailable
- Hard grading or no one in class did well
- Did not teach test material
- Objectives unclear
- Class was pointless
- Professor was not helpful
- Class did not match syllabus
- Professor had a poor attitude

The above list is significant for students and faculty (as well as for administrators) in a number of ways. For students and faculty, the list is useful for identifying where students and faculty share views on valid reasons for giving a poor evaluation of teaching and courses – such as the professor being unavailable or disrespectful, perhaps, as examples – while also providing insight on where discussion between faculty and students is needed to arrive at shared understandings of expectations – for example, as with workload and grading, or even explicit (rather than tacit) views on teaching methods and purposes of the course. The above list is also useful in identifying where students base evaluations on instructor characteristics or other aspects of the course (such as the amount a student felt he or she learned) rather than on instructor behaviours/actions and on what happens in a course, which is useful for taking steps in moving students, faculty, and administrators toward the use of evaluation of teaching that is both effective and responsible.

In contrast to the above reasons for giving poor teaching evaluations, the following were the most frequently used descriptors in written comments by students giving favourable teaching evaluations (presented here in descending order of frequency):

---

119 See Boysen 2008, p. 220.
Interesting
Approachable
Clarity
Ability to explain
Effective teaching
Knowledgeable
Willing to help
Aids understanding
Friendly
Patient
Delivery of concepts
Humorous
Stimulates thinking
Effective use of examples
Encouraging
Effective questioning
Engaging
Good lecture notes
Concise
Real-life applications

As with the previous list of reasons for poor teaching evaluations, the above list contains items that refer to instructor attributes as well as activities, which can be useful for students, faculty, and administrators, as with the previous list.

Pan et al also report that their study “indicate[s] that students value teaching quality more than teacher characteristics.” That said, as is the case for faculty and administrators, there are also “student follies” that can be detrimental to student evaluation of teaching. It is therefore important for students to understand that there is learning involved in becoming “more sophisticated evaluators” and contributing valuable and responsible input for evaluation of teaching and courses. For students, two other points warrant brief mention. The first is the usefulness of student self-evaluation in improving student achievement. Second, as Shuell

---

120 See Pan et al 2008, p. 83.
121 See Pan et al 2008, p. 73.
122 In addition to the “faculty follies” and “administrator follies” identified by Halonen and Ellenberg, the following are “student follies” that damage the usefulness and integrity of student evaluation of teaching: “they fade…they suck up…they go woebegone…they go blue…they go for the jugular…they mob…they calculate their [financial] losses…they lose focus…they use alien and imprecise language…they turn seductive…they delude themselves…[and] they become agents [of protest]” (2006, p. 153-154).
123 As cited earlier, see McKeachie 1997, p. 1223 (cited in Burden 2010, p. 107).
124 See, for example, Combs et al (2008, p. 93): “Meta-analyses of college students’ self-evaluation of learning find that it is positively correlated with student achievement (Cohen, 1986; Falchikov & Boud, 1998). More recent, individual studies from various content areas also find overall high correlations between self-assessment results and ratings based on a variety of external
notes, “It is helpful to remember that what the student does is actually more important in determining what is learned than what the teacher does.” In student evaluation of teaching and courses, in particular in the shift from the teacher-centred to student-centred classroom and curriculum, the roles and responsibilities of the student are as important as those of the teacher for successful teaching and learning.

criteria (Mehta & Danielson, 1989; Coombe, 1992; Oscarsson, 1997; Chesebro & McCroskey, 2000; Wortham & Harper, 2002; Fitzgerald et al., 2003).”

REFERENCE LIST


teacher/teaching using descriptors derived from qualitative feedback: Formative and
http://teachingperspectives.com/
Prosser, M. (2010). Faculty research and teaching approaches: Exploring the relationship. In
Taking Stock: Research on Teaching and Learning in Higher Education (J. Christensen
Hughes and J. Mighty, Eds.). School of Policy Studies, Queen’s University, pp. 129-138.
from: http://www.qaa.ac.uk/
Remedios, R. & Lieberman, D. (2008). I liked your course because you taught me well: the
influence of grades, workload, expectations and goals on students’ evaluations of
Remmers, H. (1927). The Purdue Rating Scale for Instructors. Educational Administration and
Supervision, 6: 399-406.
http://www.ericdigests.org/1996-1/status.htm
hazard?: The influence of sexual orientation on teacher credibility and perceived student
Taking Stock: Research on Teaching and Learning in Higher Education (J. Christensen
Hughes and J. Mighty, Eds.). School of Policy Studies, Queen’s University, pp. 95-112.
A practical guide to assessing teaching, research, and service (Peter Seldin and
Center for Applied Research, Lubin School of Business Administration, Pace University.
Simon Fraser University. (2012). Office of the Vice-President, Academic. Retrieved January 5,
2012, from: http://www.sfu.ca/teachlearn/tlgrants.html
Simon Fraser University. (2012). SFU Teaching and Learning Centre. Retrieved January 5,
2012, from: http://tlcentre.sfu.ca/


