Memo

V3: Edited to add appendices B and C, corrections
V2: Edited 2019-04-14 (incorporated feedback from CTAPT)

DATE: February 13, 2019 (original version)

TO: CTAPT Committee

FROM: Linda McNenly PhD, Research Associate

RE: Complementary Assessments of Teaching Report (Literature Review #2)

This report presents a summary analysis of current practices of assessments of teaching other than student evaluation surveys (SET or SET) in higher education institutions (HEI), based on a review of the literature and environmental scan of U15 practices, including the University of Waterloo.

Part 1 of this report outlines the current practices in Canada in terms of evidence supporting a complementary or multifaceted approach (MFA) to teaching evaluation, followed by a summary of the main concerns and benefits of using Peer Review of Teaching (PRT) and Teaching Dossiers (TD) as complementary approaches.

Part 2 summarizes evidence-based best practice for PRT and TD through a synthesis of key components of models and approaches for PRT and TD identified in the literature and environmental scan.

Part 3 offers a description of the use, benefits, and limitations of some of the validated tools and rubrics for PRT and TD found in the literature.
CTAPT Committee Report: Literature Review on Assessment Methods

Overview:

This report presents a summary analysis of current practices of assessments of teaching other than student evaluation surveys (SET) (also known as student ratings of instruction or SET) in higher education institutions (HEI), based on a review of the literature and environmental scan of U15 practices, including the University of Waterloo. The researcher used University of Waterloo’s tri-university library and the Centre for Teaching Excellence library for the literature review and created a catalogue of resources using Mendely and OneDrive. The scan is based on publicly available information on university websites; the researcher documented and compiled over 65 links and documents using OneDrive and an excel workbook called “CTAPT Overview.” Based on this literature review and scan, the aim of this report is to provide CTAPT with empirical evidence on current practices of complementary assessment methods and tools to guide CTAPT consultations with the university and to inform recommendations on best practices for complementary methods of teaching assessment.

The researcher found that the literature focuses on two complementary teaching evaluation methods: teaching dossiers (TD) and peer review of teaching (PRT). The teaching dossier is generally an accepted tool and the second most commonly used form of evidence after student surveys in Canada and USA (Wright et al. 2014: 40). Peer review of teaching is gaining in popularity in Canada and the USA (Chen and Yeager 2011: 207, 208), reflected by an extensive body of literature, albeit with a focus on its use for formative evaluation. Although TD and PRT are used or recommended in HEI, there continues to be challenges to the use of complementary assessment practices. A great deal of the literature, therefore, examines faculty perceptions of existing complementary evaluation methods in terms of the value (or potential value) and benefits on the one hand, and their concerns or the challenges with using these methods on the other. The main concerns relate to lack of resources and clear procedures or criteria while the main benefits include professional development, improved practice, and enhancement of collegiality. Interestingly, studies from Canada, the USA, Australia, the UK, and elsewhere report a similar ‘list of concerns and benefits’ (see Smith 2014; Teoh et al. 2016; Thomas et al. 2014; c.f. Iqbal 2014 in Canada; Bell and Cooper 2013 in Australia; Mager et al. 2014 in USA). Concerns and benefits are also consistent with perspectives from U-Waterloo faculty, based on faculty/Chairs feedback in workshops through CTE and email consultation.

Another substantial group of literature consists of studies reporting on the development, implementation, and evaluation of pilot programs, particularly of peer observations, generally referred to as peer review of teaching (PRT). These studies similarly discuss faculty participant concerns, but also share insights on how pilot programs addressed these concerns and overcame implementation challenges. The environmental scan also revealed several evidence-based PRT pilot study reports (U-Toronto, U-Alberta, UBC, and Faculty specific pilot at U-Waterloo) and more extensive teaching dossier guideline documents (U-Calgary, U-Toronto). This group of literature provides valuable evidence on: i) the usefulness, benefits, and impacts of using PRT or TD as part of a multi-faceted evaluation approach; ii) key components of an effective PRT or TD model; and iii) essential processes or procedures for developing, implementing, and sustaining a successful evaluation program using complementary methods. Some key components of PRT and TD are presented, from the basic “must haves” to more comprehensive models. Finally, while the literature is rich with studies examining PRT, there is a lack of research on instruments and tools, apart from some studies from STEM and health sciences that explain the development and validation of PRT instruments.
1. Current and Recommended Teaching Evaluation Practices

Assessment of teaching (or teaching evaluation) is done for two main purposes: i) providing feedback for faculty growth, teaching development or improvement, and ii) providing evaluative information for personnel decisions (Arreola 1995: 2; Wright et al. 2014: 36). In most cases, this is done via separate formative and summative evaluation process, but scholars suggest that these diverse purposes can be served by one teaching evaluation system. As Arreola argues, faculty evaluation and development are two sides of the same coin and should ideally work together, as long as there are "resources or opportunities which enable faculty to develop that performance" being evaluated (1995: 2, 3). Zubizarreta astutely notes that greater focus is being placed on professional development within "comprehensive" evaluation systems incorporating multiple forms of information and data with the aim of strengthening teaching (in Seldin 1999: 163). Some suggest “only when the elements of a faculty evaluation program are carefully integrated into a professional enrichment program does the institution obtain the greatest benefit from both” (Murphy et al. 2009: 232; Arreola 2007, 1995). Indeed, much of the literature on complementary methods advocate for a "improvement" or "growth-orientated" evaluation framework (e.g. Arreola 2007, 1995; Chism 1999; Seldin 2010, 2006, 1984; Wright et al. 2014). There is no one-fits-all system, rather, an evaluation system should be adaptable and tailored to specific institution needs, be responsive to a multi-disciplinary context, and consider "faculty values" (Arreola 1995: 1; Hubball and Clarke 2011: 1, 3). Nonetheless, research has shown there are core elements to an effective teaching evaluation system.

Core Elements of an Effective Teaching Evaluation Programs

The following is based on a thorough and fairly recent, mixed method study on Canadian teaching evaluation practices completed by Wright et al. in 2013, which sought "to develop a framework for improvement-orientated formative and summative assessment of teaching in Ontario" that would ultimately include policies, online tools, guides, and more, to support teaching evaluation and development (2014: 7, see also Gravestock 2011). Findings align with recommendations and approaches by other scholars (e.g. Arreola 2007; Chism 1999; 2007; Seldin 1999, 2007) and with University of Waterloo goals and future priorities.

Figure 1: Holistic Approach to the Review of Teaching (CTE)

---

1 This view is not held universally however; performance evaluation and development purposes can in some instances be incompatible, according to Michela (Appendix B in the OCUFA Report of 2019.).
1. A Multifaceted Approach

Consensus is that while student evaluation of teaching (SET) are useful when measuring certain aspects of teaching, they should not be the only source of evidence used to assess teaching and learning. It is generally accepted that best practices for teaching evaluation use a “complementary”, “comprehensive” or “multifaceted approach” (MFA), which uses multiple types of data - including evidence from students, peers, and self - and multiple approaches to gather and evaluate data, such as surveys, peer observations, and teaching dossiers, etc. (Wright et al. 5). This triangulation of data increases reliability and fairness (Berk 20414: 88; Wright et al. 2014: 16). Not only is MFA more effective than SET alone, especially in summative evaluation, it is also more likely to tease out the “contextual and structural factors that impact teaching performance,” thereby providing a more holistic picture of teaching and learning (Arreola 2007; Berk 2009; Wright et al 2014: 15, 16). This aligns with the “Holistic Approach to the Review of Teaching” advocated by the Centre for Teaching Excellence at the University of Waterloo (see Figure 1 above) and supported by scholars in the field (e.g. Arreola 2007; Chism 2007; Hubball and Clark 2011).

2. Shared Understanding of Teaching Quality

There must be a shared understanding of teaching quality, one which includes a definition of teaching effectiveness (TE), evaluation criteria of TE, and institutionally aligned goals and purpose (Gravestock 2011). This definition must be “contextual, evolving, and periodically reviewed” (Devlin and Samarawickrema 2010) and involve “faculty values” (Arreola 1995: 1; Berk 2006; Wright et al. 2014: 14). That is, disciplinary, departmental, and institutional contexts shape “what constitutes effective teaching and which aspects of teaching are valued and attended to” (CTE). Moreover, conceptions of TE should foreground the motivation and rational of how faculty approach teaching to balance traditional actions and behaviours (Wright et al. 2014: 15), which may be contextualized in a teaching dossier. Studies show there are essential or core dimensions of teaching effectiveness identified in Literature Review #1.

3. Robust Feedback Cycles and Support

Wright et al. recommend a teaching evaluation process with “...robust feedback cycles that are integrated into evaluation and instructional improvement programs,” and where faculty development programmes support feedback and improvement cycles (2014: 5, 17). A strong professional development culture influences the uptake of an improvement approach to evaluation (ibid: 17). Support and training, as well as effective stakeholder communication, are essential (ibid: 31).

4. Multi-level Leadership and Engagement

Multi-level leadership, consultation, and communication are key to an effective integrated approach (Wright et al. 2014: 17). Clear roles, responsibilities, and governance should be established (Gravestock 2011; see also Chism 1999, chapter 3). Wright et al. suggests administrators and program directors can lead a systemic review, evaluation, and development or redesign of the evaluation program in collaborative teams with faculty and with input from students (2014: 18). The authors also argue that “building cultures that value teaching and developing expertise in evaluation, and evidence-based inquiry and decision making, appear to have a positive impact on faculty engagement with teaching evaluation and on the quality of decision-making” (Wright et al 2014: 18; also Arreola 2007; Gravestock 2011). The University of Waterloo has already begun to take steps towards evidence-based inquiry on assessment of teaching and university engagement with the formation of several committees and working groups such as the Course Evaluation Project Team (CEPT and CEPT2), the Complementary Teaching Assessment Project Team (CTAPT), and a Graduate Supervisory Taskforce.

5. Scholarly Approach to Teaching

“Scholarly teaching is based on practice wisdom which is developed by reflection on experience and published research” (Martin 2007, from Allen and Field 2005).

---

2 Michela (Appendix B in the OCUFA Report of 2019) argues that this fact of consensus does not in itself warrant any evidentiary weight given the newer information that has accumulated since the appearance of most of the sources cited here. The remaining, most recent sources cited here do not in any way engage the empirical research and theoretical analyses that are held up as refutations of SETs as legitimate measures of instructional performance.
A scholarly approach to teaching refers to the process of consulting with the literature and peers on teaching methods and applying disciplinary approaches and appropriate knowledge to guide teaching-learning experience, and reflecting on teaching practices and outcomes. Potter and Kustra’s (2011) define scholarly teaching as “teaching grounded in critical reflection using systematically and strategically-gathered evidence, related and explained by well-reasoned theory and philosophical understanding, with the goal of maximizing learning through effective teaching” (in Crabtree et al. 2016: 1). It thus involves reflection, inquiry, documentation, and communication about teaching (Fong et al. 2013: 61). While studies often refer to the triad of lenses (student-peers-self) to increase validity of teaching evaluation, notice how in Figure 1 contains a fourth lens of “research” (see Brookfield 1995), showing how a scholarly approach to teaching completes the circle. Hubball and Clarke explain the value of this approach in research-intensive universities:

“Scholarly approaches to peer-review of teaching in research-intensive universities is key for successfully integrating formative and summative approaches to teaching; for providing reliable and valid evidence for administrative decision-making about the effectiveness of teaching practices for tenure, promotion and/or teaching award adjudications; and, is not only consistent with the research excellence and methodological rigor espoused by these institutions, it is responsive to their diverse needs and circumstances in multi-disciplinary contexts” (2011: 1).

It is important to note that instructors may engage in a scholarly approach to teaching but not necessarily the scholarship of teaching and learning (SoTL). In comparison, SoTL involves a systematic examination and analysis of research questions about student learning and teaching activities that are shared publicly with the aim of improving learning, strengthening teaching, and advancing the field of teaching and learning.

Sources of Evidence Used

Although Berk (2005) reviews 12 potential sources of evidence to measure teaching effectiveness, the literature review and environmental scan confirmed most HEI in Canada, the US, and elsewhere mainly use one or more of the following: student ratings (SET), self-evaluation (in particular teaching dossiers), and peer review of teaching. Therefore, this report focuses on two complementary methods:

1. **A Teaching Dossier** (aka Teaching Portfolio) is a factual, integrated summary of a teacher’s approaches, strengths and accomplishments. “It contains documents and materials that collectively suggest the scope and quality of a professor’s teaching performance. The portfolio is to teaching what lists of publications, grants, and honours are to research and scholarship” (Seldin et al 2010: 4). Some recommend developing more than one format: a reflective TP, and one used for evaluation (e.g. Schoenwetter et al 2002: 90).

2. **Peer review of teaching** is an informed collegial assessment of faculty teaching for either fostering improvement or making personnel decisions (Chism 2007: 3). Peer review of teaching is a holistic approach to evaluation involving peers, but the literature often uses this term to mean peer observation of teaching. This report similarly uses peer review of teaching (PRT) in reference to peer observation methods of assessment. Gosling (2002 in Teoh et al. 2016: 1-2) identifies 3 models:
   i) Evaluation Model for summative feedback (senior staff evaluates performance for assessment)
   ii) Developmental Model for growth and improvement (‘experts’ observe and evaluate teaching competencies)
   iii) Peer Review Model for formative feedback that involves mutual reflection (teachers observe each other to encourage reflection via discussion)

The Peer Review Model is prevalent in the literature. **All models involve at minimum three steps:** pre-observation meeting, observation, debriefing meeting with feedback. Peer Observation can take the form of dyads (one on one), triads, online environment, Teaching Squares, mentorship, or a trained “peer consultant,” for example.

3 “Teaching Squares is a concept created by Anne Wessely of St. Louis Community College. The Teaching Squares approach involves a self-reflective process about teaching gained through observation of one’s peers. It is not a peer evaluation exercise but rather a self-evaluation process which takes place in a confidential and mutually supportive environment” (Monica Vesley, Centre for
Studies from Wright et al. (2014) and Gravestock (2001) show that universities in Canada use multiple forms of data or evidence for teaching evaluation. The most commonly used form of evidence for summative purposes remains SET at 82%, followed by teaching dossiers (50%), other self-evaluation instruments (29%) and peer observation of teaching (20%) (Wright et al. 2014: 42). Complementary approaches tend to be used for formative evaluation: 67% of Canadian Universities use peer observation, 43% use self-evaluation instruments, and 31% use teaching dossiers (ibid). While multiple forms of evidence are being used, they are not always required for promotion and tenure at Canadian Universities; 39% of universities require teaching dossiers, and only 7% require in class observation by peers (Wright et al. 2014: 43). Wright et al.’s study of Canadian and Ontario practices points to one of the main barriers to using complementary assessment methods - the lack of standard materials and procedures for these processes (Wright et al. 2014: 43).

University of Waterloo

- Responses from the CTAPT Email Consultation (April 2018) reveal few Departments or Schools use PRT. Thirteen units indicated they "could or should" use PRT. Only 5 explicitly stated that they use PRT “on occasion” or “periodically” while 2 said they are used. The numbers of departments using PRT for tenure and promotion could potentially be higher. One said they used to use PRT but do not anymore.
- The “Departments/Schools with PRT in Place” table produced by CTE based on Chair’s Meetings (2012) show that 14 had formal PRT processes in place while 4 had informal policies about PRT; another table shows 6 reported no PRT process, 2 said no formal process, some said they were in the process of reviewing the use of PRT (see CTAPT Overview for links to all these documents).
  - This indicates that use could have be underreported in 2018, or that use has changed since 2012.
- Only one responded that they used to use TD but do not anymore. It is important to note, however, numerous responses spoke of "other methods they use" that ideally could be easily compiled into a TD, such as: unique approaches or innovations, curriculum development, contributions to teaching excellence, mentoring, professional development, evidence of self-evaluation leading to improvement, development of teaching, teaching materials, and context (Email Consultation, April 2018).

Evaluating Evidence

Arreola’s “rule of thumb” outlines an important principle for selecting the best sources to use: “…the important principle to follow in identifying sources is always to select the source which has the best opportunity to observe first-hand the performance to be evaluated (2007: 47 in Murphy et al. 2009: 228). Based on this principle, the table below outlines some examples drawn from Chism 1999; Mills in Seldin 1999; and Seldin 1999, 2006.

<table>
<thead>
<tr>
<th>Can Evaluate / Provide Evidence on</th>
<th>Instructors</th>
<th>Peers</th>
<th>Admin / Chair</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clarity of teaching philosophy</td>
<td>Subject matter expertise</td>
<td>Corroborate or supplement information about teaching contributions, supervision, mentoring</td>
<td>Personal impact on learning</td>
</tr>
<tr>
<td></td>
<td>Approaches to teaching and assessment (including objective's and appropriateness)</td>
<td>Course content - current, accurate, appropriate, alignment</td>
<td>Can evaluate comparatively</td>
<td>Teacher-student interactions</td>
</tr>
<tr>
<td></td>
<td>Adaptations of teaching</td>
<td>Teaching strategies or approaches used and how well delivered</td>
<td>Efforts of professional development</td>
<td>Student experiences and perceptions</td>
</tr>
<tr>
<td></td>
<td>Professional development and impact</td>
<td>Ethical behavior</td>
<td>Rapport with students</td>
<td></td>
</tr>
<tr>
<td>Methods, Source of Information</td>
<td>Teaching Journals, Video Recordings, Faculty Activity Reports, Video recordings, Teaching Dossiers - can</td>
<td>Classroom Observation (PRT) and review of materials and/or student products</td>
<td>Comparative and Individual Evidence</td>
<td>Appropriate SET questions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contextual information, TD</td>
<td></td>
<td>Other student feedback mechanisms</td>
</tr>
</tbody>
</table>

Teaching Excellence Blog, University of Waterloo, February 22, 2013. Depending on the definition and model of PRT, not everyone would consider Teaching Squares a form of peer review of teaching.
provide:
  a) Descriptive data (course materials, teaching philosophy, supervision, classroom research, teaching leadership, etc.)
  b) Evidence of adaptations of teaching, evidence of SoTL, professional development efforts, awards or honors received

Barriers to Using Peer Review of Teaching and Teaching Dossiers

Two prevalent themes emerged in the literature regarding PRT and TD: a lack of resources, and a lack of clear procedures and criteria (e.g. Hubball and Clarke 2011: 20-23; Iqbal 2014; Smith 2014; Teoh et al. 2016; Thomas et al. 2014). Based on the literature review, this section provides a summary of the main resource and procedural issues and related concerns about the usefulness and validity of these methods. When faculty expressed concerns or experienced challenges, some of the potential benefits of using PRT and TD were not fully realized. The following figure (Figure 2) consolidates these themes as derived from the literature.

- Barnard et al. (2011: 442) write that faculty are ‘time poor’ and professional development is a low priority, particularly when potential benefits and rewards are not emphasized. In fact, one of the most common concerns with TD and PRT is the time commitment required to collect evidence, observe, review materials, write narratives or feedback, and interpret evidence (e.g. Iqbal 2014: 113-4; Barnard 2001).
  - However, studies have shown faculty are willing to participate in peer review of teaching, despite required time commitment (e.g. Iqbal 2014: 113-4; Barnard 2011: 435). Bell (2001: 38) further suggests that the outcomes justifies the time commitment.
- Another challenge is that the ‘reflexive mode of writing’ may be unfamiliar or uncomfortable for some, both for those writing and those assessing the TD narrative (Murphy et al 2009: 231-2).
- A significant issue is the lack of clear guidelines, criteria, and tools to conduct PRT or develop and interpret TD (e.g. Iqbal 2014: 108; Schönwetter et al. 2002: 91; Teoh et al. 2016: 1). This often leads to concerns about validity, consistency (reliability) and bias.
- Critiques of PRT mainly focus on subjectivity, fairness and representativeness (only one observer, one time), and availability of appropriate rating scales (Berk 2005: 50).
DeZure states the primary sources of bias would be “disagreement about what constitutes good teaching,” interpersonal relationships, disagreement what to observe, lack of consistency in process, and lack of training of observers (in Seldin 1999:72).

Another concern is that evaluation (of either PRT or TD) would be bias towards the reviewers’ style and approach to teaching and learning (Gormally et al. 2014: 188; Seldin 2010: 43; Thomas et al. 2014: 150).

- Faculty are concerned with the lack of training in terms of how to conduct PRT observations or provide constructive feedback. Some are concerned that observers are not knowledgeable of pedagogical expertise or lack experience conduction teaching assessments (Iqbal 2014: 115; Chism 2007: 6).
- The quality of feedback is another issue; some studies show feedback can be too vague and not constructive, or too positive (Iqbal 2014:108; Richard et al. 2018; Theo et al. 2016: 4).
- In her study on faculty participating in PRT at a Western Canadian research-intensive university, Iqbal found that summative PRT strengthens validity of evaluation process but does not lead to professional growth in teaching whereas formative fosters self-reflection and generates insights into teaching (Iqbal 2014: 109; cites other studies with similar results, Bryne et al 2010; Kell and Annetts 2009; Peel 2005). Reasons cited was lack of quality feedback and the fact that PRT often involves unequal relationships where collegiality is at stake in an already high stakes process (Iqbal 2014).
- Others consider whether TD and PRT can be used for formative evaluation only, or if they can be used for both formative and summative purposes (see Arreola 2007: xxii; Knapper and Wright 2001: 25; Murphy et al. 2009: 230).
- Responses from the CTAPT Email Consultation (April 2018) suggests that faculty concerns are very similar; some comments include that PRT was “not useful” and “not critical” or that the process was not clear. Those who found PRT useful, said it was particularly so for new or probational faculty.

Benefits of Using Peer Review of Teaching and Teaching Dossiers

Studies examining the development, implementation, and evaluation of existing or pilot programs emphasize the benefits of PRT or TD and provide insight on the how barriers can be mitigated, thereby maximizing benefits for faculty. The three main benefits of using PRT and TD for a MFA highlighted in the literature include: increasing validity and reliability of teaching evaluation; encouraging professional development and change of teaching practices, and building collegial relationships.

- Studies show PRT and TD enhances faculty engagement in reflective practices and professional development of teaching and learning and encourages the development of a scholarly teaching community of practice (e.g. Bell 2001; Barnard 2011: 443-445; Crocker and Hoekstra 2015: 53; Pelger and Larsson 2018: 184; Thomas et al. 2014: 119-20). Reflection is an “essential skill of effective teachers” in their professional development (Bell 2001: 32).
  - As online teaching becomes more prevalent, PRT provides opportunities to make gains in teaching development through researching the process of online teaching and learning process (Bennett and Sandy 2009: 455).

- TD allows academics to engage in all components of a scholarly approach to teaching, which, like research, is based on rigorous evidence of teaching effectiveness involving reflection, inquiry, evaluation, and documentation (Fong et al 2014: 61; see Berk 2005: 56-7; Kenny et al. 2018 Knapper and Wright 2001: 24; Wright et al. 2014: 24).
  - TD serve a dual purpose: it provides a way for faculty to collect and document evidence of their professionalism, “teaching processes and outcomes,” and achievements; it also provides an effective framework to reflect on areas of improvement (Wright et al. 2014: 24).
  - Findings from an e-Portfolio pilot found academics collected more and different types of feedback than before the pilot (Crocker and Hoekstra 2015: 70).

- PRT can result in mutual benefits for both the reviewee and the reviewer (even if observation roles are not switched) in models emphasizing collaborative relationships (Hendry 2014: 325, 327; Hubball and Clarke 2011: 4; Mager et al. 2014).
  - For example, a nursing in comprehensive university in United States piloted a peer review program aimed to bridge the gap between formative and summative evaluation by developing a partnership and co-mentoring approach (dyad): “post project focus groups revealed that
faculty found their modified PRoT process to be a mutually beneficial experience that was more useful, flexible and collegial, and less stressful than their previous evaluation process” (Mager et al 2014: 113-115).

- PRT provides opportunities for dialogue about teaching and learning, facilitates collegiality, and builds confidence among faculty (Bell and Cooper 2013: 64; Hendry et al. 2014: 325, 327; Mager et al. 2014).
  - Pelger and Larson write faculty experienced increased confidence after using a TD (2018: 184).
  - Bell and Cooper reported overwhelmingly positive responses by all participants in the Engineering PRT program in Australia “with regard to teaching skills and ideas, confidence, collegiality, giving and receiving feedback, resources and training” (2013: 64).
  - Studies examining the use of PRT in an online environment found participants valued the opportunity to discuss online teaching and processes, learn from each other, and build community (Bennett and Barp 2008; Harper and Nicolson 2011).

- Significantly, studies have shown that faculty participating in PRT make changes or improvements to their teaching (e.g. Barnard 2011: 443; Bell 2001: 33; Hendry et al. 2014: 322; Hubball and Clarke 2011: 19; Shortland 2004).
  - The qualitative study by Bell (2001: 33) states 75% reported making immediate changes to practices that were technical, pedagogical, or critical in nature; half of the changes fell into the pedagogical category.
  - A mixed methods study by Hendry et al. (2014: 322) of a peer observation program in research intensive HEI in Australia indicates that of the 28 staff who observed a colleague, “82% said they had learned at least one new teaching strategy,” and 95% of them had tried it in their own class.
  - Gormally et al. summarizes best practices for providing peer feedback, arguing that faculty are more likely to make changes to teaching practices when supported by coaching and feedback (2014: 187).

- Using PRT and TD strengthens the validity, reliability, and fairness of the evaluation process by triangulating contemporaneous student-self-peer evidence, and by providing essential and important data students are not able or knowledgeable enough to assess (e.g. Bell 2001; Drew and Klopper 2014; Knapper and Wright 2001; Murphy et al 2009; Smith 2014; Wright et al 2014: 5, 16-7; see also Arreola 2007; Chism 1999 xi, 2007).
  - On triangulation in a multifaceted approach to teaching evaluation, Berk (2005: 46) argues: “By drawing on three or more different sources of evidence, the strengths of each source can compensate for weaknesses of the other sources, thereby converging on a decision about teaching effectiveness that is more accurate than one based on any single source.”
  - Teaching Dossiers are the most effective and systematic away to integrate and demonstrate multiple forms of evidence and data collected (MFA) (Wright et al. 2014: 17; Policy 77, University of Waterloo).
  - A TD meets the “twin aims of responsible evaluation and continual improvement,” and encourages reflection without minimizing need for evidence and accountability (Zubizarreta in Seldin 1999: 163).

2. Modal Approaches

What are the Essential Components of Peer Observations (PRT)?

This section describes some approaches to Peer Observations (PRT) as a complementary method highlighting the essential components studies showed contributed to a positive outcome. **Table 1** provides some examples of PRT models found in the literature. The examples have in common attention to resources and procedures to mitigate challenges or concerns, thereby enhancing faculty benefits of participation. Most of these examples are for formative evaluation and situated within an improvement-based framework. PRT models vary in terms of the purpose or use, process steps and range of activities (e.g. review of materials, form of feedback, action plans, etc.), frequency of observations, and roles of
participants. Given this variability, when implementing a model, all of these aspects should be described in detail so that procedures and processes are clear. Some key findings regarding these aspects are teased out below.

**Purpose / Use:**
The intended *actual use* of PRT must be clear; PRT done for formative purposes, but then submitted as evidence for tenure should be thought of as summative or evaluative (Chism 1999: 4). Nonetheless, the *outcomes* of formative PRT (e.g. SET scores improved, tried new innovative practice, successful change in practice, etc.) may be incorporated into a teaching dossier as evidence for summative evaluation. When PRT is for summative purposes, it *must include*: guidelines on how an observer is chosen, or choose a trained observer who is part of a PRT team; clear criteria with a standardized tool; frequent observations; and the report includes a description of the process used to gather feedback (Arreola 1995: 51; Chism 1999: 79).

**Process Steps:**
Scholars agree best practices should include a review of course materials (Murphy 2009: 228) as well as clear guidelines and checklists, rubrics, or guiding questions to ensure reliability of observations (Chism 1999: 76). Arreola discusses objectivity by introducing the concept of “controlled subjectivity [which] is the consistent application of a predetermined set of values in the interpretation of measurement data” (1995: xix; 2007). In addition, the literature consistently notes the importance of training for observers to enhance reliability and objectivity (Arreola 1995; Hubball and Clarke 2011: 23; Iqbal 2014: 119; Mager et al. 2014: 114; Smith 2014; Thomas et al 2014: 152; Wright et al. 2014: 50).

The basic PRT model includes at minimum three stages (Millis in Seldin 2006: 84):

1. **Pre-observation Meeting**: discussion of goals, purpose, expectations, and ‘rules’ or logistics of visit; course materials provided in advance.
2. **Observation in class (or monitoring online teaching)**: non-participatory peer observation of items discussed, classroom environment, and interactions; recording observations using a narrative log, checklist, and/ or rubric.
3. **Debrief / Post-observation Meeting**: peer provides reflective feedback, focusing on goals discussed; instructor provides reflection on how class went; discussion of strengths and possible actions for improvement or innovation; may include written report.

- Strong PRT processes involve opportunities for reflection by *both* the observer and observee, which is regarded an “essential skill” of effective teachers (Bell 2001: 32). Constructive feedback is critical, and faculty may benefit from training on how to give feedback (Thomas et al 2014).
- **Feedback** needs to be *timely, relevant, descriptive, sensitive, and constructive*; confidentiality is important. Mager suggests feedback should also be *self-referenced* versus norm-referenced (Mager et al. 2014: 114); however, a scan of processes show that some universities suggest that feedback should be norm-referenced, that is, based on known criteria (e.g. U-Alberta PRT document, CTAPT Overview).
  - Gromally et al (2014: 192) *vision for effective feedback* is voluntary and delivered immediately and on more than one occasion; feedback providers need to be perceived as sympathetic, credible, and unbiased. The authors argue feedback should identify areas for improvement in a positive but constructive manner and provide examples on how to improve.
  - Weimer’s two guiding principles to effective feedback are to give the type and quality of feedback you would like to receive, and to develop an agenda so peers can learn and reflect together on a specific problems (in Gormally et al. 2014: 193).

- See Table 1 for examples of PRT models in the literature.

**Examples from Scan (see CTAPT Overview):**
- The University of Toronto “Peer Observation of Teaching: Effective Practices” guide (2017) clearly outlines processes, procedures, and models, and provides tools in the appendix. This is a comprehensive, evidence-based document with information and sources in one document.
- University of Alberta PRT guide and toolkit is based on research they conducted. Comprehensive, evidence-based guide and several documents from one central easy to navigate webpage.
UBC outlines formative and summative PRT review processes, with accompanying guide. Multiple links and pages, but also a concise wiki page.

University of Calgary presents a CARRA model for formative feedback (based on 4 lenses) and tools (resources link) as an outcome of their Formative Feedback for Teaching Development Initiative.

The CTE at the University of Waterloo supports faculty interested in participating in Teaching Squares. CTE has also compiled examples of descriptors, checklists, and rubrics for PRT in a CTE Sharepoint (not publically accessible) folder called “Peer Review of Teaching: Observation of Classroom Instruction Possible Categories and Items.”

<table>
<thead>
<tr>
<th>Table 1: Examples of Approaches to Peer Review of Teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic Model</strong></td>
</tr>
<tr>
<td>1) Pre-observation Meeting</td>
</tr>
<tr>
<td>2) Observation</td>
</tr>
<tr>
<td>3) Feedback report</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Roles:
There is a shift from ‘traditional’ reviewer – reviewer model, to more collaborative and collegial approaches emphasizing reciprocity and mutual benefits (Thomas et al 2014: 113), what Gosling 2002 refers to as the “peer review model.” Whether traditional or collaborative, formative or summative, roles must be clear. Moreover, scholars suggest that formative and summative evaluation be linked in a goal driven, improvement focused framework, otherwise evaluation of either kind is viewed as punitive (Arreola 2007: xxii in Murphy et al 2009: 229-30).

For example, Barnard et al. describes roles in a collegial framework based on a community of practice as such: “The peer partner is not necessarily the person with all the answers; instead, the partner cooperates and offers opinion, explores new strategies together and looks toward solving future challenges” (2001: 437). This model also involves a fourth step to include ‘implementation’ through continued reflection and action, which serves to emphasize the importance of impact or outcomes of PRT.

“Peer partnership involves a reciprocal relationship whereby a peer is invited to eyewitness a partner’s teaching and learning activities and to provide supportive and constructive feedback (Bell 2005).... The peer partner is not necessarily the person with all the answers, instead, the partner cooperates and offers opinion, explores new strategies together and looks toward solving future challenges” (Barnard et al. 2011: 437).

Bell and Cooper’s (2013: 68) mixed method case study of PRT (variation of a triad) in Engineering in an Australian HEI discusses the significance of process and role of an educational developer (coordinator) in enhancing objectivity: “Most participants responded positively to the role of the coordinator and the importance of a facilitator who is seen as unbiased seems significant in supporting the program.”
Example from scan: Developed in Dublin, University of Calgary outlines in the “Genuine Peer Observation of Teaching” document a model that outlines the roles of observer and person observed with “I do x statements,” which aims to establish peer observation as a “interaction between equals” (see CTAPT Overview).

Frequency:
Multiple observations enhance reliability and validity of PRT (De Zure in Seldin 1999). A minimum of 2 to 3 observations are recommended, others recommend two observers (Chism 1999: 76). Some of the literature describes PRT models that feature multiple visits or cycles of observation-feedback-reflection-action (e.g. Barnard et al 2001; Bell 2001; Mager et al. 2014; see Table 1):

- For example, Mager et al. (2014: 115) write how their two round model was designed to bridge the gap between formative and summative evaluations through co-mentoring in a PRT process.
- Bell 2001 describes a structured PRT program using an action research framework with a robust feedback cycle to support reflective practice and professional development (Australia). The participant chooses a support college, and they complete 4 cycles of PRT over 2 years; in one cycle, the roles are reversed, and the participant observes the support college who explains their philosophy and methods. An educational developer monitors the program and provides feedback on reflections. This study highlights the importance of timely and relevant feedback in a structured cycle using an “action research framework” and the value of a “support triangle” acting as “critical friends,” which also includes an educational developer who monitors the program, provides feedback, and encourages reflection and action (Bell 2001: 36-8).

Key Components of an Effective PRT Process
(based on Thomas et al 2014; Bell and Cooper 2013; Smith 2014):
- Goal driven, improvement-based framework
- Collegial; can involve collaboration between early and late career as equal partners
- Involvement of educational developer and Centres for Teaching and Learning
- Clear and transparent structure with consideration of ethics and confidentiality
- Detailed pre and post meetings in a robust feedback cycle
- Faculty Ownership:
  - Tailoring PRT according to faculty and disciplinary needs
  - Control of process, e.g., voluntary participation driven by faculty goals, choice of reviewer, compilation of relevant course materials to give reviewer.
- Clear, evidence-based criteria (and guidelines) that are known and shared
- Resources: time, training, tools and templates
  - Time to complete process and to implement actions
  - Trained faculty and/or educational developers
  - Tools: instruments that can accommodate different teaching modes
- Reflective practice and the opportunity and support to implement action goals
- Professional documentation and development through Teaching Dossier

What are the Essential Components of a Teaching Dossier (TD)?

Teaching dossiers are the most effective and systematic way to integrate and demonstrate multiple forms of evidence and data collected in a complementary or multifaceted approach to teaching evaluation (Wright et al. 2014: 17; Policy 77, University of Waterloo). Using a TD, therefore, enhances validity of teaching evaluation by presenting a fuller picture of teaching and learning and by triangulating data. In addition, Centra (2001) found that TDs are reasonably reliably (inter-rater reliability) when reviewers had specified criteria to use (Knapper and Wright 2001: 26-7; Murphy et al. 2009: 230). Again, there needs to be agreement of what constitutes “teaching effectiveness,” but “beware of trying to force portfolios into a quantitative paradigm when one of their strengths is providing rich qualitative data that will be different from person to person” (Knapper and Wright 2001: 27). TD also need to be valued and rewarded to provide incentive for faculty to commit to the time it takes to develop a TD (Knapper and Wright 2001: 28).
Some scholars suggest developing two formats – a reflective version and one used for evaluation and dissemination (Schönwetter et al. 2002: 90). However, this seems redundant, as teaching dossiers are also, ideally, a self-reflective tool that also provides a method for teachers to link formative and summative evaluations (Knapper and Wright 2001: 27; c.f. Murphy et al. 2009: 230 also considers challenges). Smith, for example, proposes a “formative review processes that results in a teaching portfolio that would reflect a faculty member’s efforts and successes in a critically reflective PRT process, and contributes to ongoing teaching improvement” (2014: 1). Knapper and Wright echo this opinion, stating “it was envisaged that portfolios would be used mainly for major career decisions such as tenure and promotion, but it was also apparent that compiling a portfolio would stimulate a good deal of reflection about teaching by the individual concerned and by those who read the portfolio” (2001: 21). Moreover, while a TD for summative purposes might take shape differently than a ‘reflective’ TD, Knapper and Wright found that the differences between the two are not that different in practice (2001: 25). For example, a TD written for personnel decisions (summative) must include evidence, however, a TD for improvement also involves reflection through an examination of scholarship and evidence (Zubizarreta in Seldin 1999: 178).

**Two key characteristics of a TD** are that it is focused and “curated.” First, a TD is not a compilation of random information, it is focused and organized around principles, for example, “career development” could be an integrative theme (Seldin 1999: 110). The teaching philosophy statement serves as a central narrative that describes instructors beliefs about teaching, practices reflecting those beliefs, results of those practices, and future goals (Schönwetter et al. 2002: 85). “A teaching philosophy statement is a systematic and critical rationale that focuses on the important components defining effective teaching and learning in a particular discipline and/or institutional context” (Schönwetter et al. 2002: 85). Second, a TD is “curated” in that it represents an instructor’s “best work” rather than an exhaustive compilation of all materials related to performance and effectiveness (Seldin 1999: 110). This does not translate into a biased picture, but rather, a fair and accurate representation (ibid: 110-1). As Zubizarreta explains, a rigorously curated TD with proper guidance and review “cannot hide weak performance...because the evidence is just not there”; conversely, a TD provides an opportunity for a teacher to document excellence that has not been recognized (in Seldin 1999: 170).

As a well-established practice in North America, TD components are consistent in the literature, varying only in terms of how much is included, resulting in a either a briefer or more extensive TD. The length and content will also vary dependent on the instructor’s goals and stage of career and HEI policies and priorities. Seldin estimates one can complete a TD to second draft stage in 12-15 hours, spending most of this time thinking, planning, and compiling documents or evidence for the appendix (1999: 111). On average, dossiers that are eight to ten pages plus a manageably sized appendix would be enough for most professors (ibid:111). Because dossiers are compiled and composed by teachers themselves, teachers have a sense of ownership and control of the evaluation process (Knapper and Wright 2001: 27). Dossiers, therefore, balance academic freedom with rigorous evaluation.

The table below outlines the range of TD components from basic “must haves” to a comprehensive “deluxe” model. (The evaluation of teaching dossiers is discussed in Part 3).

<table>
<thead>
<tr>
<th>Basic - Must Haves (1)</th>
<th>Basic - Plus (2)</th>
<th>Focussed Dossier (3)</th>
<th>Comprehensive (1, *)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching Responsibilities</td>
<td>Teaching Responsibilities</td>
<td>Teaching Responsibilities (includes supervision)</td>
<td>Teaching Responsibilities</td>
</tr>
<tr>
<td>Teaching Philosophy (reflective statement of objectives, strategies, methods)</td>
<td>Teaching Philosophy</td>
<td>Teaching Philosophy</td>
<td>Teaching Philosophy</td>
</tr>
<tr>
<td>Efforts to Improve (Professional development - seminars and workshops)</td>
<td>Results and Achievements</td>
<td>Professional Development and Future Plans</td>
<td>Redesigned Courses</td>
</tr>
<tr>
<td>Teaching Goals</td>
<td></td>
<td></td>
<td>Efforts to Improve</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Service to Teaching</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Info from Students</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Info from Peers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Info from Other</td>
</tr>
<tr>
<td>Evidence (Appendix) At</td>
<td>Evidence (Appendix)</td>
<td>Evidence (Appendix)</td>
<td>Future Goals</td>
</tr>
</tbody>
</table>

---

**Table Notes:**
- Basic - Must Haves (1) includes essential elements that must be present.
- Basic - Plus (2) adds additional components that enhance the dossier.
- Focussed Dossier (3) is a condensed version, emphasizing key achievements.
- Comprehensive (1, *) represents a detailed, expansive dossier.

---

**References:**
- Schönwetter et al. 2002
- Knapper and Wright 2001
- Murphy et al. 2009
- Smith 2014
- Zubizarreta in Seldin 1999
- Seldin 1999
3. Instruments and Tools for Assessment of Teaching

While there is a gap in the literature of validated instruments adapted for cross disciplinary use, there are some studies that examine the development, testing, and validation of instruments for teaching evaluation, mainly from the health sciences and STEM. Most of the instruments tested are checklists or rubrics for peer observation. This section describes the use, benefits, and limitations of each of the following: TPI, COPUS, BERI, POET, PORTAAL, and PARF. Evaluation guidelines and rubric for a teaching dossier philosophy statement are also described. Additional Teaching Inventories and Classroom Observation Protocols may be found on the Yale University Poorvu Center for Teaching and Learning website. The CTAPT Overview workbook created by the researcher documents these and other resource links from Canadian Universities containing additional examples of PRT tools and checklists as well as TD guidelines.

Peer Review of Teaching Instruments

Wieman and Gilbert (2014: 552, 556) describe the development and validation of a Teaching Practices Inventory instrument (TPI) consisting of evidence-based factors of effective teaching in STEM. The instrument presents factors as items in a checklist with quantified values. TPI is similar to the PULSE Vision and Change course-level rubric in that it includes all seven factors listed in PULSE, but provides more “extensive and detailed characterization of teaching” (see Table 2, p558). The eight categories or factors in TPI are: course information provided; supporting materials provided; in class features and activities; assignments; feedback and testing; other (diagnostic); training and guidance of teaching assistants; collaboration and sharing of teaching (ibid: 553). TPI can be used i) for meaningful comparison of teaching used for different courses within and across departments, and ii) to gauge extent of use of research-based teaching practices but does not measure the quality of implementation (ibid: 553). They suggest the tool would be valid in STEM and social sciences; however, to be valid, the TPI must characterize the range of teaching practices used in a course (ibid: 553, 561). TPI’s main use is for guiding improvement by identifying practices they are not using, as well as demonstrating teacher’s effort to incorporate practices for promotion (560). TPI was tested for validity in that: items mean what they should, instructors interpret items in a consistent and accurate manner, and that the “inventory covered all teaching practices used by more than two instructors in our largest test sample” (ibid: 553, 555).  

The TPI may be found on the Carl Wieman Science Education Initiative Website at UBC. This website also provides a validated Classroom Observation Protocol for Undergraduate STEM or COPUS (adapted from TDOP), which “allows observers, after a short 1.5 hour training period, to reliably characterize how faculty and students are spending their time in the classroom” (LINK). In short, it provides abbreviations for activities to log on a time sheet that documents what students are doing compared with what instructors are doing for how long. The authors suggest that the protocol provides valuable feedback on how the students and instructors are spending their time and identifies areas for improvement (Smith et al. 2013: 626).

Also on this website is a student engagement observation protocol called BERI (Behavioral Engagement Related to Instruction protocol), which is a “classroom observation protocol for quantitatively measuring student engagement in large university classes.” It provides level of engagement descriptors (engaged-disengaged) for different activities, for example, descriptions for engaged listening, writing, reading,
Crabtree et al. 2016 tested whether the **Peer Observation and Evaluation Tool (POET)** developed in pharmacy could be used in occupational therapy for formative evaluation. The tool consists of 39 items divided into four sections (4 stages of PRT) - three sections are rubrics with Likert-type scale, the fourth section for the post observation meeting consists of instructor self-reflection questions (see Appendix). POET was tested for content validity (agreed upon which items were essential aspects of teaching) and inter-rater reliability. The authors suggest that results of testing indicate POET may be used to document instructor strengths, facilitate professional growth through “systematic, strategic, and constructive” peer review feedback, and offer suggestions to improve teaching. However, POET may have limitations as it was tested through observing videos of instruction rather than in person observations and only for one specific lecture; the authors also recommend cross faculty testing prior to use (Crabtree et al. 2016: 8).

**PORTAAL** (practical observation tool assessing active learning) is an observation rubric to assess active learning consisting of 21 elements based on discipline specific review of best practices that have been shown to increase student outcomes related to achievement, logic development, and other learning goals, clustered around four dimensions (Eddy et al. 2015: 1, 3):

- practice (e.g. alignment question level with assessments, immediate feedback),
- logic development (e.g. practice higher order activities, students explain answers)
- accountability (e.g. activities worth points, small group work, use random call vs volunteer), and
- apprehension reduction (praise work of class or individual, error rephrasing, emphasize hard work over ability)

Tables 1-4 in the article provide more description of the dimension elements, how they are observed, the outcomes, and the citations of supporting evidence for the dimensions. The authors suggest that the tool may facilitate the use of research based best practices by communicating these practices to instructors, provided that the tool ‘include[s] a clear description of each teaching intervention such that a novice could implement them’ (ibid: 2). The benefit of this tool is that it is supported by literature, easy to learn, and does not depend on deep pedagogical or content expertise (ibid: 2). However, like POET, the tool is ideal for video lecture and not feasible to use in real time, and the evaluation of quality based on time used in class is problematic and cumbersome (ibid: 2, 13). Moreover, this tool only measures classroom experience not materials (ibid: 2). The authors claim PORRTAL can be used in disciplines other than STEM (3); however, the researcher suggests the tool would have to be validated for that discipline's best practice and context.

Another tool for PRT that focuses on learner centred practices is the Peer Assistance and Review Form or **PARF**. PARF reflects “learner centered practices with four domains anchored on Danielson’s Components of Professional Practice principles: planning and preparation, classroom environment, instruction, and professional responsibility.” (Magno 2012: 107). The rubric was pilot tested with 183 randomly selected higher education faculty in 5 different major areas in a HEI in the Philippines, where participants were observed by two reviewers in their classroom (ibid: 104); observers were trained and provided class documents. Validity and reliability testing showed high internal consistency, indicating high reliability and concordance of the two reviewers across the four domains; this also shows inter-reliability and validity of items (consistent understanding of item meanings) (ibid: 109, 110, 115). The authors suggest this tool is suited for a PRT approach using a “constructive process” with the aim of facilitating improvement. Theoretically, this could be used across disciplines.

CTE at the **University of Waterloo** has compiled examples of PRT items and categories that can be used in labs, for team teaching, web-based instruction, and learner centered practices, for example, that may be used as a starting point for the construction of rubrics to be refined and tested, or compared with tested examples; there are also some department or school specific examples (see CTAPT Overview).

**Teaching Dossier Guides and Instruments**

Schönwetter et al. (2002) provides guidelines to **develop and evaluate** the central narrative of a teaching dossier, the **teaching philosophy statement**. "A teaching philosophy statement is a systematic and
critical rationale that focuses on the important components defining effective teaching and learning in a particular discipline and/or institutional context" (2002: 84-5). The authors point out crucial pitfalls to avoid in evaluation related to bias and subjectivity, which are echoed by Seldin (2010: 43).

"They need to be cognizant of separating their evaluation of the specific views represented in the teaching philosophy statements of others from their evaluation of the quality of the teaching philosophy statement" (Schönwetter et al. 2002: 84-5).

"Evaluation of the teaching philosophy reflected in the writer’s teaching philosophy statement is justified when the focus is to assess the extent to which it is grounded in teaching and learning theory and demonstrates fluency with theory. In some situations, congruence with a programme, departmental or institutional vision for teaching and learning may also be considered" (ibid).

As a solution, the authors provide a rubric grounded in the literature reflecting some of the categories of approaches to scholarship of teaching, validated through graduate students, faculty, managers, and faculty development professions in HEI across Canada. The rubric consists of text descriptors with Likert type scale of poor, average, and superior. Table 1 in the article outlines the categories “critical for the evaluation of each component of the teaching philosophy statement" (ibid: 91). Categories are:

- definitions of teaching and learning;
- view of the learner;
- goals or expectations of student-instructor relationships;
- teaching methods and evaluation;
- personal context of teaching;
- organization.

In short, the teaching philosophy statement, and the TD as a whole, is an evolving document and changes with knowledge and experience, thus evaluation focuses on cohesion and integration of knowledge, beliefs and practices (ibid: 91). Studies examining the reliability and validity of TD are non-existent, except for a much-cited study by Centra, who found that TD are reasonably reliable (inter-reliability) when reviewers had specified criteria to use (Knapper and Wright 2001: 26-7).

The environmental scan revealed a lack of specific guidelines, criteria, and tools for the evaluation of teaching dossiers. Most universities refer instructors to their faculty contracts or promotion guidelines. However, some resources were found:

- UBC provides descriptors for evaluation
- U-Toronto provides guiding questions
- U-Saskatchewan describes criteria such as the degree of completion of the TD, clarity of organization, includes a broad selection of evidence, and connections between the teaching philosophy and evidence provided
- Holmes from CTE at the University of Waterloo has developed a “Focused Dossier” template and rubric that could be circulated for feedback and piloted as an option.

Often cited as a reference, Seldin offers general questions to consider in evaluating a teaching dossier, as well as specific questions for each section that could form the basis of a rubric (2010: 44-50). These criteria should be adapted to already established criteria and suggested focus, and be shaped to meet institutional, faculty, and departmental contexts (ibid: 46). General questions to consider include inclusion of required items, clear profile of individual “style, priorities, and teaching achievements,” evidence of accomplishments supports reflection portion, alignment or consistency between reflection, actions, and evidence, currency and balance of evidence from triad (self, peers, students), evidence provided is valid documentation defined by department and institution, and efforts to improve through time (ibid: 44).

Conversely, the environmental scan uncovered a plethora of comprehensive, robust, and specific guidelines for developing teaching dossiers. These guidelines consist of self-reflection questions to guide development of a teaching philosophy statement, descriptions of key sections and components of a TD, examples of evidence to include, TD Templates, and examples of a completed teaching dossier itself; some also include questions and rubrics for evaluating teaching dossiers.

- The University of Calgary “Teaching Philosophies and Teaching Dossiers Guide” is an exceptional and recent example (Kenny et al. 2018, Taylor Institute for Teaching and Learning); it includes a section on evaluating a TD (with sample rubric) and a “Guide for Providing Evidence of Teaching” in the appendix. The document reiterates that this is not a checklist but a guide, as not every category may be relevant to an instructor’s context.
• Additional examples may be found at the University of Toronto and UBC (see CTAPT Overview workbook). Other universities have similar information but not in one comprehensive document or require faculty to follow a series of links and resources to other external websites to compile information piecemeal for themselves.
Conclusions

Research conducted for CTAPT confirms that the main complementary methods of teaching evaluation remain teaching dossiers, self-evaluation, and peer observation (PRT); however, PRT is becoming more prevalent and shifting towards collaborative and reciprocal approaches. The main concerns about PRT and TD focus on the lack of clear standards, criteria, and/or tools for using these methods, which relate to additional questions of validity, subjectivity, and bias. At the centre of much of the literature is an acknowledgement that the notion of ‘good’ or ‘effective’ teaching is ambiguous in teaching evaluation programs. Studies show faculty are also concerned about time commitments in an already ‘time poor’ environment, and often do not realize the benefits or purpose of engaging in PRT due to the lack of quality feedback. Conversely, other studies demonstrate how effective and well-supported PRT programs and/or the use of TD lead to positive outcomes. Some of the benefits of using complementary approaches identified in these studies include encourages reflective practice and professional development; facilitates change of teaching practices (new or improve); and, creates opportunities for dialogue and builds collegiality. Studies also show that effective PRT programs or using TD in fact strengthens the validity and reliability of teaching evaluation through triangulation of evidence. Wright et. al. 2014 provide a feasible framework for developing an effective teaching evaluation program that maximizes the benefits and addresses the concerns found in the literature on complementary methods of the assessment of teaching.

Best practices for teaching dossiers are straightforward. The literature on teaching dossiers confirms components and use of teaching dossiers are well established and vary only in terms of details and length (what is included based on institution specific policies or length depending on experience and career stage). Guides for developing a TD are plentiful whereas adequate tools and/or guides for evaluating a dossier are lacking. One debate questions whether a TD may be used for summative and formative evaluation. In contrast to the TD, it is evident in the literature that there are many possible models of PRT. Even so, there are studies evaluating the implementation and outcomes of various PRT models that provide insight on best practices. Most of these PRT programs are based on the 3 stage model (including a review of materials), and have in common clear procedures, multiple visits or an established PRT cycle, and robust training, and emphasize timely and constructive feedback as well as collegiality and reciprocity. Faculty ownership, control, and confidentiality also require for attention.

Recommendations

This section outlines recommendations on how findings from the research may be used to meet CTAPT objectives. In terms of developing complementary methods for assessing teaching and learning in general, consider these suggestions as a starting point for discussions. More detailed recommendations related to PRT and TD follow below.

1. The literature supports the use of a multi-faceted or complementary approach to the assessment of teaching as a way of triangulating evidence to increase validity and fairness. Wright et. al. (2014) identifies additional core elements of an effective teaching evaluation program that are relevant for CTAPT:
   - Shared understanding of teaching effectiveness (TE);
   - A scholarly approach to teaching (and potentially seeding a community of practice);
   - Obtaining and sustaining multi-level leadership, with attention to institutional culture, and the importance of multi-level leadership in fostering a culture that values and rewards teaching (see also Barnard et al 2015, and Hubball and Clarke 2011)
   - University community engagement and communication.

2. Address concerns about lack of clarity, validity, and bias by establishing clear processes and procedures (which outline frequency and supports as well as a call to value time dedicated to teaching akin to research) in addition to developing tools and methods. The purpose, use, and benefits of each method should be clear, and similarities or differences in use for formative and
summative purposes should be specified. Establish a balance between need for standardized
documents (guides and tools or template) and faculty ownership or control by recommending a
cascade approach and developing flexible and adaptable tools.

3. The need for training and resources is evident in the literature. Building on existing CTE programs,
recommend the implementation of a comprehensive and sustainable “evaluation training program”
consisting of supports for how to, for example, develop a dossier and incorporate evidence such as
PRT reports and/or outcomes, conduct peer observations and use PRT tools, give constructive
feedback, and evaluate PRT data and teaching dossiers. This may involve tweaking existing
programs and/or developing additional new ones and will be dependent on resources. A combination
of training, clear procedures and TE definition, and availability of tools would address concerns in the
literature about subjectivity, bias, and the quality of peer review feedback. What will be CTE’s role in
developing and maintaining a training program?

4. Participatory consultations with faculty and/or chairs in the development of flexible instruments and/or
piloting tools. Review and adapt existing examples of guidelines and tools from the environment scan
that are consistent with findings about lack of guidelines, criteria, and tools as well as clarifies questions about the purpose and benefits of using a TD. On that note, the
guide should emphasize benefits of using TD highlighted in the literature, for example, provides a
framework for reflective practices and a scholarly approach to teaching, documents academic
teaching successes, increases the validity of the evaluation process, and provides opportunities for
self-evaluation and the contextualization of SET. Two examples from the scan that illustrate what this
might look like include Teaching Philosophies and Dossiers Guide (University of Calgary, 2018) and
Developing and Assessing Teaching Dossiers (University of Toronto, 2017).

Teaching Dossiers

1. Develop a Teaching Dossier Guidelines (or Toolkit) that outlines the purpose and benefits of using a
dossier and provides guidelines for developing and evaluating a dossier, which might include guiding
questions, rubric, checklists, and an example of a completed TD in the appendix. While this may
result in a lengthy document, it addresses the main concerns about lack of guidelines, criteria, and
tools as well as clarifies questions about the purpose and benefits of using a TD. On that note, the
guide should highlight in the literature, for example, provides a framework for reflective practices and a scholarly approach to teaching, documents academic
teaching successes, increases the validity of the evaluation process, and provides opportunities for
self-evaluation and the contextualization of SET. Two examples from the scan that illustrate what this
might look like include Teaching Philosophies and Dossiers Guide (University of Calgary, 2018) and
Developing and Assessing Teaching Dossiers (University of Toronto, 2017).

2. Develop a “TD in Brief Guide” that includes a TD template based on Holmes “focused dossier”
proposal (CTE). The focused dossier model consists of the essential components of a TD identified in
the literature while offering a more streamlined and time efficient approach. Consider and indicate
where information or data already being documented for existing forms (such as performance review
or annual activity forms?) currently used at the University of Waterloo may be incorporated into a TD.
Both the guide and template should be useable campus wide but flexible enough to accommodate
Faculty priorities and different stages in a faculty member’s career.

3. Recommend only one format of teaching dossier (not both a reflective and evaluation format) that is
integrative of formative and summative evidence; however, carefully consider how summative and
formative processes and/or data can be linked.

Peer Review of Teaching

The key components of an effective PRT process are outlined on page 12 of this report (see also Smith
2014). Based on these findings, the following are recommended.

1. Develop a clear PRT process based on the basic 3 step model of pre-observation meeting (with
review of course materials), observation, and post-observation meeting, highlight the importance of a
timely feedback cycle (see Gromally et al 2014 and Mager et al. 2014), and develop a Guide with
tools. Peer Observation of Teaching: Effective Practices (University of Toronto, 207) illustrates what
this guide might entail; however, the University of Waterloo PRT model should attempt to build on existing Waterloo PRT models, as applicable, and be flexible enough to be adapted for different modes of teaching and “faculties values.”

2. The literature recommends multiple visits and cycles, and a minimum of two observations should be conducted for formative evaluation. Consider how a multi-observation PRT cycle can include both formative and summative evaluation to maximize time efforts and benefits (see Mager et al. 2014). Training and tools/instruments must support this process in order to ensure that feedback and outcomes are “useful” and beneficial. (Again, instruments/tools should be flexible and adaptable to faculty values and different modalities).

3. Piloting a formative PRT process and model could help support uptake of PRT in general; this may also facilitate shifting PRT towards a more collegial or reciprocal approach (peer review model) that fosters mutual self-reflection and benefits. The “peer review model” that focuses on formative feedback and mutual reflection described by Gosling 2002 (see also Hendry 2014 and Hubball and Clarke 2011) is prevalent in the literate, and adapting such a model presents an opportunity for CTAPT to develop a hybrid approach that potentially could be adopted for different purposes – formative or evaluative.

4. Think beyond the pairs and consider “support triangles” when talking about peer review of teaching. This could involve facilitation or support by an “educational expert” for summative evaluation, or facilitation and monitoring for formative evaluation. In this “support triangle,” observer, observee, and educational expert, are “critical friends” who encourage reflection and set actionable goals of improvement or growth within a collegial framework (see for example model described by Bell and Cooper 2013). Should CTE facilitate and monitor a peer review program? Other PRT models in the literature involve trained reviewers from another faculty, pairing of new with experienced faculty, or voluntary reviewer-review dyads with alternating roles (e.g. Bell 2001). There is no one ‘right’ model, but roles must be clear and protocols established (e.g. confidentiality, who observes, how observers are selected, etc.) to maintain transparency, trust, and validity.

5. Recommend that time and opportunities (professional development, CTE workshops) for growth be available to support a PRT program (and MFA approach to evaluation in general) and promote an improvement or growth orientated approach to evaluation by highlighting the benefits of formative PRT (such as opportunities for dialogue and reflection about teaching, fostering confidence and collegiality, lead to changes or improvements, and ultimately strengthening validity of evaluation in the case of summative).
Bibliography


## Appendix A: CTAPT Overview of Environmental Scan Findings

<table>
<thead>
<tr>
<th>University</th>
<th>Teaching &amp; Learning Centres</th>
<th>Practices: Guides and Tools Found</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>University of Alberta</strong></td>
<td>Centre for Teaching and Learning</td>
<td>Peer Review of Teaching (PRT) (online and doc) Peer Evaluation-evidence based (Guide and PPT); and Resources (PRT Lit Review Table); Appendix O - Tool for evaluating TD; Multifaceted Evaluation Guide</td>
</tr>
<tr>
<td><strong>University of Calgary</strong></td>
<td>Taylor Institute for Teaching and Learning</td>
<td>Teaching Dossier (TD) Guide &amp; Report; Teaching Awards - instructor and for Grad Supervision (criteria list, application uses TD &amp; Letters); Formative Evaluation (Initiative; Guides, reports and resources); Teaching Squares guide</td>
</tr>
<tr>
<td><strong>University of British Columbia</strong></td>
<td>Centre for Teaching, Learning, and Technology</td>
<td>Teaching Dossier Guide (online and PDF); Peer Evaluation - Formative (Guide and resources, questions to consider) Peer Evaluation - Summative PRT (Rubric and Guidance notes); Definitions Effective Teaching; Wieman link - resources, tools</td>
</tr>
<tr>
<td><strong>University of Manitoba</strong></td>
<td></td>
<td>Evaluation Guides are specific to Faculty / School; examples downloaded 1. Dentistry Prom and TT Ref Manual, 2. Architecture</td>
</tr>
<tr>
<td><strong>University of Saskatchewan</strong></td>
<td>Teaching and Learning</td>
<td>Framework for PRT best practices doc; TD Guide; Teaching Awards (criteria) - online</td>
</tr>
<tr>
<td><strong>University of Ottawa</strong></td>
<td>TLSS - Teaching and Learning Support Services</td>
<td>N/A--</td>
</tr>
<tr>
<td><strong>University of Waterloo</strong></td>
<td>Centre for Teaching Excellence</td>
<td>From CTE: TD Guidelines; Developing a Teaching Philosophy; Tools for reflecting on Teaching PRT an holistic approach From CTE sharepoint, Chairs folder 2012: PRT Categories and Items document (checklists and Rubrics); PRT Examples (UW and other)</td>
</tr>
<tr>
<td><strong>Western University</strong></td>
<td>Centre for Teaching and Learning</td>
<td>Teaching Dossier Guide (online) Policy Guidelines; Promo and Tenure Art.</td>
</tr>
<tr>
<td>University</td>
<td>Centre/Resource</td>
<td>Summary</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>University of Toronto</td>
<td>Centre for Teaching Support and Innovation</td>
<td>Teaching Dossier Guide (Definition, evidence, resources, evaluation guide for chairs) -online and PDF; Peer Observation of Teaching Practices (guide, tools, resources, questions); Formative Evaluation for mid course (examples and workshop), Def good teaching; Graduate Supervision Guide; Reports - Faculty Mentoring for Teaching (lit review, qual study); and P2P Pilot Report</td>
</tr>
<tr>
<td>McMaster University</td>
<td>MacPherson Institute for Leadership, Innovation, and Excellence in Teaching</td>
<td>Teaching Portfolio policy and structure outlined SPSB12; Preparing a Teaching Dossier Guide; Education Dossier Guide; Report - Recommendation to Improve Course and Teacher Evaluations; SoLT Guide; Peer Observation of Teaching (POT) Project and TOOLKIT, TP Guide for Health Sciences; Peer Evaluation Health Sci Folder (9 docs)</td>
</tr>
<tr>
<td>Queen's University</td>
<td>Centre for Teaching and Learning</td>
<td>Teaching Dossier Guide - basic elements doc, and online Preparing a TD PPT doc – definition, elements, etc.</td>
</tr>
<tr>
<td>McGill University</td>
<td>Teaching and Learning Services (TLS)</td>
<td>Teaching Portfolio Guidelines (brief, TLS) - online and guiding questions Full guidelines - Employment policy; PPT from career and planning services</td>
</tr>
<tr>
<td>Universite de Montreal</td>
<td>N/A--</td>
<td>N/A--</td>
</tr>
<tr>
<td>University Laval</td>
<td>N/A--</td>
<td>N/A--</td>
</tr>
<tr>
<td>Dalhousie University</td>
<td>Centre for Teaching and Learning</td>
<td>Quick Guide - TD step by step creation link; TD template (word doc)</td>
</tr>
<tr>
<td>STLHE</td>
<td>Teaching and Learning Centres in Canada</td>
<td></td>
</tr>
<tr>
<td>CAUT</td>
<td>TD Guide</td>
<td>What is Fair? Q&amp;A on PRT</td>
</tr>
<tr>
<td>N Dakota U</td>
<td>PRT for Online Course - checklist (pdf)</td>
<td></td>
</tr>
<tr>
<td>Piedmont College</td>
<td>Observations of online and hybrid courses (checklist)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B: U15 Environmental Scan Summary Report - Best Practices

Added April 2018

An environmental scan of current practices of teaching evaluation was conducted in the winter of 2018 based on an examination of publicly available documents found on U15 websites. A searchable excel table called CTAPT Overview compiles and documents over 65 representative documents and resource links (including description of items). In addition to this table, the following Resource Folders on the CTAPT SharePoint site contains PDF/word/ppt documents (when available for download) from the environmental scan:

1. Teaching Dossier Resources from scan
2. Peer Review Resources from scan
3. UWaterloo Resources from scan
4. Reports about Evaluation
5. Other-TA-Grad from scan (other miscellaneous, teaching awards, graduate supervision)

Examples of Best Practice from U15 Environmental Scan:
Informing findings in the literature and the objectives of CTAPT, this report highlights some of the “best" or “best practice" examples of teaching dossier and peer review of teaching guides, guidelines, tools, and instruments from the scan based on a comparison of content. The purpose of this report is to recommend relevant and useful examples and explain the value of each of these resources to facilitate the process of developing complementary assessment of teaching methods for CTAPT. These examples also offer insight on what CTAPT models and end products of guides or tools could look like. Links to documents in the CTAPT sharepoint file, or links to webpages where downloadable documents are unavailable, are provided.

Teaching Dossier, Best…

❖ **Comprehensive Evidence-based Guide**: It’s a tie! Both examples have strengths in different areas:

❖ **University of Calgary**
   ➢ Guide includes evidence presented in the document with outline of principles of effective teaching. Includes descriptions, guiding questions, and examples for each TD section; a section on evaluation of a TD as well as a guide for providing evidence in the appendix (includes facets of teaching expertise, examples and sources/examples of evidence); and a bibliography. Has everything needed to develop (and evaluate) a TD if one can dedicate the time to read it. Although it is long, such an organized guide allows a person to work through sections of the TD at their convenience using the Guide as needed. Missing: an evaluation instrument – rubric or checklist.
   ➢ Guide for Providing Evidence of Teaching

❖ **University of Toronto**
   ➢ Not only does this guide describe the components of a TD (contents of a TD) and provide examples of material and evidence sources, it also outlines descriptive criteria of teaching effectiveness from competent to excellent. Provides examples, description, and further references for each section of a TD.
   ➢ Includes section on developing teaching philosophy statement with guiding questions and additional sources. Sections 9 and 10 provide examples of evidence to include that demonstrates “Leadership in Teaching” and “Professional Development,” respectively.
   ➢ FAQ section and “Evaluating Teaching Dossiers for Department Chairs and Tenure, Continuing Status and Promotion Committee Members” in the Appendix

❖ **Brief TD Guide**: Queen’s University
   ➢ Concise, clear, straightforward. Outlines and describes the essential components of a TD in a way that does not seem onerous.
- **University of Toronto** is also a good example because it includes examples of what evidence to include and how to evaluate a TD; however, it falls short in that it offers less explanation than the Queen’s version on the different sections of a TD. It is also a longer ‘brief’ document overall.
- Best overall that combines developing and evaluating TD - Holmes focused dossier

- **Resources for electronic portfolios:** McGill University (online), links to various sources and articles.

- **TD template:** Dalhousie University
  - Fillable and adaptable word document template makes first time construction quicker and easier.

- **Explanation of purpose and use of a TD:** McMaster University
  - Guide describes why you should use a TD and the benefits (for you, Dean, Chair, Faculty…): a TD serves to clarify goals and commitment to teaching, track educational contributions, and document teaching and learning accomplishments for reviews. The Guide also mentions how existing processes and documents (such as CV and APR) tie in, and describes the value of TDs for facilitating discussions at reviews and setting goals.
  - Limit – could go further to also tie in PRT and emphasize how using a TD enhances the validity of evaluation. There are additional benefits identified in the literature that are not mentioned in this Guide; the point is that it is important to highlight purpose, use, and benefits of TD in a Guide.
  - Guide also includes what they call a TD template at the end, which provides guiding questions for developing each section of a TD (rather than a fillable document like the Dalhousie example above).

- **Tool for evaluating contents of TD:** University of Alberta (Appendix O)
  - A rubric adapted from Seldin 2010 with 2 scales (meets standards / below standards) with weighting for different sections of TD allocated as such: materials from self 40-50%; from others 30-40%; products of good teaching 10%; Appendix (and completion and organization of TD) 10%. Offers an idea of how to use evidence and potentially distribute weightings in a tangible example.
  - Other tools are available from the scan; most U15 refer to other sources or adapt existing tools.

- **Online only guide:** UBC
  - Provides good examples of what to include in each section of a TD, not just description of what the sections should consist of (cf. Western online guide, although Western expands on developing a Philosophy statement and provides ‘real’ examples of completed TDs).

**Peer Review of Teaching (PRT), Best…**

- **Comprehensive evidence-based Guide:** University of Toronto
  - Clearly describes a formative PRT model that can be used for in-person and online courses but also explains how it can be adapted for summative evaluation. Well-structured and detailed but concise Guide informed by a University of Toronto study and supported by literature with description of benefits of PRT clearly stated.
  - Recommends 3 step model, with detailed guidelines for each step, but also provides examples of other peer observation protocols such as teaching squares. Also highlights the importance of feedback and outlines best practices for choosing observers and observing.
  - Offers an abundance of tools and instruments for observation in Part III, including guiding questions and tools for each step, for example, pre-observation template, observation templates (open ended form, narrative log sample template, and observation form checklist), and an online observation sample template. An appendix with suggested best practices for visits and questions to avoid complete this Guide.

- **Differentiated Formative and Summative Guidelines:** UBC (online)
  - Formative: Outlines PRT program and purpose (what to review, aspects of PRT, and how I will benefit), which uses the 3 step model. Includes a FAQ tab that provides useful tips on what,
when, who, confidentiality, training, biases, time/effort, as well as a tab with a list of available Peer Reviewers from different Faculties (with short bio for each person).
- NB: Saved to SharePoint folder the following supporting tools: 1. PRT Pre-Observation Questions (goals for review and what to review, etc); 2. PRT Classroom Observation Questions for Peer Reviewers; 3a. Guidelines for the Written Report for the Peer Reviewer; and 3b. PRT: Post-Observation Questions
- **Summative:** Provides a summative PRT Rubric using Likert scale (with 6 levels exceptional to poor) with descriptors and examples for each level of 3 categories – Learning Design, Teaching Activities, Educator Development – along with Guidance Notes for using the rubric. Similarly provides an action button / link to Faculty representatives of summative PRT and documentation from the Faculties (applied science, arts, and business) where PRT have been adapted at faculty level in policies and guidelines.
- NB: Rubric links to document in SharePoint

- **Procedural Document:** University of Alberta
  - Outlines procedures for developing a summative PRT protocol in a concise, one page figure.
  - Similar to CTE University of Waterloo “Steps for designing a scholarly peer review of teaching system” on the Peer Review of Teaching: A holistic approach to the Review of Teaching on their webpage (also referenced in Report LR2), which was used as one of the sources to develop this figure!

- **Website Toolkit and infographics:** McMaster (online), with some documents/infographics examples saved and uploaded on CTAPT SharePoint
  - POT Toolkit includes: Introduction to what is POT; FAQ; 3 models infographic; an implementation infographic, 5) opportunities for POT (direct, recorded, online); related forms; feedback grids (clinical and lecture); and Guide (1page)
  - Note: PPT deck provides good details of formative evaluation - fears, academic culture, relationships/selection/feedback, and process (3 stages and GROW approach)

- **Bank of PRT Tools and Instruments:** University of Alberta
  - Appendix from “University of Alberta Peer Review of Teaching” document includes examples from multiple sources all in one spot, such as criteria examples (from Hubball and Clarke 2011 and UBC Faculty of Arts), observation tools (Likert and checklist), sample materials evaluation tool (Checklist), sample rubric, tool for Review of Laboratory Instruction, tool for Peer Review of Clinical Teaching, tool for online teaching, tool for reviewing a TD (recommended in TD section above), sample report and menu.
  - Will also note they produced a (workshop?) powerpoint, which clearly outlines the differences between formative and summative PRT, purpose, use, and process.

Other:
- **Formative Feedback Guide:** University of Calgary
  - More broadly about using feedback from 4 lenses perspective – self, peers, students, scholarship – to contextualize feedback, using the CARRA Model (Curiosity, Ask, Receive, Reflect, Act) as a framework for collecting formative feedback with guiding questions (provided). Tips on how to give and receive feedback with sample activities (1 minute paper, stop-start-continue, peer consultation, literature reviews, and professional development).
Appendix C: Preliminary Email Consultation Summary Report

Added April 2019

This brief report summarizes results from the preliminary consultation email sent to all Faculties at the University of Waterloo in April 2018. The scoping email sought to gain insight on how units define and measure ‘good teaching’ and assess graduate supervision. Each respondent/response represents one Department or School “unit.” Individual names, departments and schools are not identified in this report or data analysis files.

Q1. Do you have definitions of “good teaching” to use in summative assessments (i.e. merit review) and/or teaching awards? If so, can you forward them to me?

NO (Total n=9)
None (n=5)
- Five (5) of the nine respondents explicitly said they have no formal definition of good teaching: “The definition of good teaching is a challenge.” One of the five specified that tenure and promotion guidelines are also used for APR, which outlines what counts for assessment but does not define what is “good.”

No, based on SET (n=4)
- These respondents said they have no definition but specified they rely heavily on teaching scores in evaluation, suggesting that SET results are used as indicators of “good teaching”:
  “Teaching scores are leaned on heavily (i.e., a good teaching score = good teaching)”
  “…departmental Discipline Standards for Tenure and Promotion indicate that ‘A 'strong performance’ in teaching would be indicated by teaching evaluations generally above the Faculty of Arts average…”

Yes (Total n=2)
- Two (2) stated they have guidelines or criteria they use to define good teaching based on Policy 77, Bain 2014, and criteria such as “…engaging students, grading fairly, organizing courses well, following Departmental policies for teaching, and generally treating students with respect.”

Q2a. Do you use a method other than course evaluation data to help assign teaching scores during annual performance review? Q2b. Have you found any alternative methods (e.g. peer review, teaching portfolios) to be helpful?

- SETS are heavily relied on however many noted that peer evaluation “should” or “could” be used to contextualize SET scores or balance assessment. Many units indicated they use additional evidence or criteria for evaluation, three mentioned this is documented through APR reports.
- Most respondents (n=15) said that SETs are the main source of evidence used. Many stated that SETs are heavily relied on or one of the primary sources (n=8); however, just as many (n=7) said that SETs need to be used in context, and things like class course, size, etc. are also considered.
- Thirteen (13) units use or ‘should use’ Peer Evaluation. Note that this number includes information from published guidelines used as a ‘unit response’ (i.e. where units did not respond or from respondents who quoted/referred to guidelines that state peer observation could be used) and therefore does not necessarily reflect actual practice.
- Some respondents did specify frequency or use:
  - Five (5) said they use PRT on occasion or periodically (also “some,” “limited,”) in conjunction with promotion and tenure.
Two (2) responded they use PRT, one of them stated they have a “fair amount of weight” in evaluation.

Two (2) do not find peer review useful while one (1) said PRT is very useful to support probationary faculty; this respondent also stated they did not know of a single faculty member who has asked for a peer review.

Barriers and concerns mentioned: reviews are always positive, hesitant to criticize colleagues (1); unclear guidelines and process (should it be part of promotion process or not? will it be valued?) (1); debate about process – suggested senior faculty to observe but resistant due to time commitment (1).

One (1) said all faculty are encouraged to create and maintain a Teaching Dossier while another (n=1) said that they used to use dossiers but do not anymore because they are not sent to externals.

One (1) said all faculty are encouraged to create and maintain a Teaching Dossier while another (n=1) said that they used to use dossiers but do not anymore because they are not sent to externals.

Most (n=19) said that they did not use other ‘methods’ but listed additional evidence or criteria that may be used in evaluation such as teaching awards, contribution to teaching excellence, successful innovations, professional development, curriculum development, etc.

In this category of “additional evidence or criteria,” the number of teaching awards faculty has been nominated for or won and professional development activities were mentioned most often (6 and 14 mentions, respectively).

Professional development includes activities as outlined in CTAPT’s conception of effective teaching, for example: unique approaches to teaching/learning, engaged in mentorship, evidence of pedagogical techniques, new course development or course improvement, participates in workshops/conferences, and evidence of self-evaluation.

Three (3) mentioned this evidence can be documented on the APR forms / activity reports.

Table 1: Hierarchy Tree Visualizing Number of Coded Responses to Questions 1 & 2

![Hierarchy Tree Visualizing Number of Coded Responses to Questions 1 & 2]
Q3. What approaches do you use to assess graduate supervision as teaching?

- The main approaches to assess graduate supervision are quantitative measures (number of...)
  and engagement in ‘supervisory activities’ such as training and mentorship.
- Three (3) units stated there is no process, or none that they are aware of, for assessing quality of
  graduate supervision.
- Seventeen (17) units said they use quantitative data, including number of graduate students
  supervised (12 mentions); number of supervisory committees served on (5 mentions); number of
  students to complete graduate degree (7 mentions) within standard timeframe, or progressing in
  a timely fashion. One (1) also mentioned graduate student attrition rates are considered.
- Some units (n=7) stated supervisors (presumably ‘good’ supervisors) are expected to support
  graduate students by involving or supporting them in research related activities or opportunities
  (i.e. ‘training’), which can be measured by outcomes / outputs such as graduate students
  publishing or co-publishing research (6), presenting at conferences (4), obtaining awards (3), and
  finding appropriate placements or jobs (2). (Expectation to provide students with research
  funding, if available, mentioned by one (1) unit.
- Only four (4) units mentioned mentorship; one (1) specified activities such as meeting regularly
  with student, being available and responsive, and providing feedback and advice.

Table 2: Hierarchy Tree Visualizing Number of Coded Responses to All Questions*

*Also includes coding related to Question 2 - units who expressed interest in alternative methods.

5 2 N/A, 1 none for evaluating quality
Email Text

I am a member of the university’s Complementary Teaching Assessment Project Team (CTAPT), and we held our first meeting last week. Once action item that arose was an environmental scan to identify how the units in our Faculty define good teaching and how teaching is assessed using methods other than course evaluation data.

Could you let me know the following in brief or forward this on to someone else in your department/school who is better-suited to answer:

1. Do you have definitions of “good teaching” to use in summative assessments (i.e. merit review) and/or teaching awards? If so, can you forward them to me?
2. Do you use a method other than course evaluation data to help assign teaching scores during annual performance review? Have you found any alternative methods (e.g. peer review, teaching portfolios) to be helpful?
3. What approaches do you use to assess graduate supervision as teaching?

Please note that the identity of units that provide this information will be protected when discussed by the project team.

If easier to have a quick phone call rather than correspond by email, that is no problem at all. It would be helpful to receive information by Friday 13 April for discussion at CTAPT’s next meeting.

Thanks!
ian VanderBurgh