REPORT ON THE FINDINGS OF THE PEER REVIEW OF TEACHING PROJECT

Aleksandra Kasztalska

CTL-CAS Writing Program Exchange

Boston University

2021

TABLE OF CONTENTS

The Center for Teaching & Learning (CTL) at Boston University established the Classroom Observation Troupe and Learning Community (generally referred to as the Observation Troupe) in Fall 2019. Former Assistant Director Brad Wheeler facilitated the group in its opening meetings, and I transitioned into this facilitation role early in 2020. The original goals of the group were to "engage faculty with classroom observation techniques helpful for student learning or teaching improvement, explore and apply classroom video recording technology for observation, and share the observation process with each other and the BU community."

The members of the Observation Troupe are Amy Bennett-Zendzian, Lecturer, CAS Writing Program; Kyna Hamill, Director of the College of Arts & Sciences Core Curriculum and Master Lecturer affiliated with BU's School of Theatre and African American Studies; Nathan Jones, Associate Professor of Special Education; Aleksandra Kasztalska, Lecturer, CAS Writing Program; and Christina Michaud, Master Lecturer and Interim Director of ELL Writing, CAS Writing Program.

The group met regularly through Spring 2020 and Fall 2021 to review the scholarly literature on peer review of teaching, investigate techniques and technology associated with peer observation, collate resources such as rubrics and guidelines produced by other universities, and formulate guidelines and supporting materials that would be helpful to the BU community.

In Spring 2021, Dr. Aleksandra (Aleks) Kasztalska consolidated the work of the Observation Troupe while working part-time as a faculty consultant with CTL. This report and the accompanying website guide are the product of her fruitful time with CTL. We in CTL are very grateful to Aleks and to all members of the Observation Troupe for their work. Their commitment to enriching the teaching experience of their colleagues through peer review will, in turn, enrich the learning experience of BU's students.

The next stage of the group's work is to consider how to support the growth of peer review of teaching at Boston University. Group members are experimenting with different approaches in their programs and we look forward to consolidating this work and supporting other faculty who are interested in reflecting on their teaching in this way. Stay tuned for more news about these efforts! And, in the meantime, please dive into the report for more information on peer observation and its role in supporting the development of our teaching and learning.

Deborah Breen

Director, Boston University Center for Teaching & Learning

1. INTRODUCTION TO PEER REVIEW OF TEACHING (PRT)

1.1. WHAT IS PRT?

Peer review of teaching (PRT) refers to various procedures and instruments used to observe and offer feedback to instructors on their work. PRT can be conducted with the goal of helping an instructor grow professionally and/or to assess an instructor's pedagogical practices, content knowledge, and other expertise, qualities, or skills. As a result, PRT can be used in many situations and for different purposes, including annual reviews and contract renewals, hiring decisions, coaching and mentoring of junior faculty, or for professional development of current faculty.

Most commonly, PRT involves a faculty member observing another instructor's class and offering feedback and/or evaluating the instructor's effectiveness. However, it should be noted here that classroom observations are only one source of data that can be collected to understand and assess an instructor's work. A comprehensive PRT process should collect various types of evidence and indicators of the instructor's development and/or effectiveness across multiple contexts and situations (Chism, 2007).

When done properly, PRT can be an invaluable tool that supports faculty in their professional development and emphasizes teaching as a skill that needs to be regularly re-examined, contextualized, and updated. And because PRT is a process that involves two or more faculty members working together and learning about one another's professional practices, it also promotes a more open and collaborative model of teaching.

When teaching is studied, shared, and discussed in a supportive and constructive manner, faculty in the program can benefit from these conversations and may better meet their students' needs. Additionally, PRT can help stimulate conversations about best teaching practices and principles valued in the program. Thus, by making teaching more transparent, PRT has the potential to not only improve individual instructors' professional practice but also to encourage important conversations about excellence in teaching.

1.2. A NOTE ON TERMINOLOGY

PRT is an umbrella term that can refer to either teaching observations or teaching evaluations.

- o **Teaching observations** are usually conducted by the instructor's peers (that is, faculty of similar professional rank or status), who provide the instructor with constructive feedback on their teaching and help them improve and grow professionally.

- o **Teaching evaluations** tend to be conducted by a more senior or experienced instructor, whose task is to assess another instructor's overall effectiveness vis-a-vis a set of professional standards or general expectations for teaching in the program.

Because the term "peer review of teaching" is widely used to refer to both observations and evaluations, this guide will also use "PRT" as the overarching term to refer to any situation in which teaching is observed. The term "reviewer" will likewise be used to refer to any observer

and/or evaluator. When necessary, this guide will distinguish between teaching *observations* and teaching *evaluations* to highlight the fundamental differences between the two types of PRT.

2. TYPES OF PRT

2.1. FORMATIVE PRT

The goal of **formative PRT**--which is often referred to as **peer observation**--is to offer constructive feedback on an instructor's teaching that can lead to adjustments and improvements in their future work (Chism, 2007). Formative PRT is often more informal and more private than summative PRT in order to encourage risk-taking and honest discussions of teaching practices. As such, the results of formative PRT tend to be confidential and excluded from annual reviews, external assessments, and other high-stakes personnel decisions.

Many argue that formative PRT should not be mandated, but rather encouraged and flexible in its implementation (Robertson, 2006). Imposed and rigid formative PRT protocols can create the impression among faculty that PRT is an unwelcome inconvenience and an encroachment on their teaching. Instead, seeking formative feedback should be regarded as part of the larger departmental or institutional culture, and instructors should engage in it not because of fear of penalty, but because they intrinsically and genuinely want to improve their teaching.

There are different approaches to conducting formative PRT and a variety of factors that should be considered in selecting peers for this type of review. Some of the most common approaches are outlined in Robertson (2006) and summarized below. For example, certain PRT models resemble a **buddy system**, in which two instructors observe each other over a longer period of time and reflect on their practices as well as share feedback with one another. In other cases, instructors can form **teams or clusters**, where they regularly observe each other's classes, discuss and critique each other's teaching, and learn from one another.

In these and other models, the observer and the observed might be colleagues from the same department, or they may be instructors from different academic programs, depending on the goals of the PRT and the needs of the instructors. Moreover, peers or groups could be matched based on their prior experience and seniority. For example, a new hire can be matched with a more senior instructor who serves the role of a mentor, or instructors can pair up with others they already know and trust. Alternatively, peers can be matched based on the courses they teach and their familiarity with the program curriculum. In sum, formative PRT can take many forms to meet the needs of the faculty and the program.

2.2. SUMMATIVE PRT

In contrast to formative PRT, **summative PRT** has a more evaluative component and may be referred to as **faculty evaluations or assessment.** It is typically used in personnel decisions, such as hiring, promotion, or tenure and is intended to assess an instructor's overall teaching effectiveness, as well as whether they are meeting program goals. As a result, summative PRT is conducted in reference to explicit and transparent departmental values and best practices that are recognized as the foundations of effective teaching in the given program (Robertson, 2006). In other words, summative PRT evaluates whether an instructor meets certain minimum criteria and benchmarks and assesses their effectiveness vis-à-vis general expectations for teaching in the program.

Summative PRT must be grounded in agreed-upon and observable professional standards & best practices.

Summative PRT tends to be less nuanced than formative review, because the goal of summative PRT is assessing the overall effectiveness of an instructor's teaching vis-a-vis certain standards and thresholds (Chism, 2007). Moreover, summative PRT may include quantitative components, such as when different teaching activities are counted, coded, and/or assessed using numerical scales (e.g., Likert scale). At the same time, effective summative PRT should use multiple types of evidence of teaching effectiveness that are not limited to classroom observations, but also include a review of course materials and assessment tools, review of work produced by students, and/or instructor's self-reports and self-assessments.

Because summative PRT is used in high-stakes personnel decisions, great care should be taken in designing protocols for this type of faculty review. Above all, summative PRT must be transparent and grounded in agreed-upon professional standards and best practices, and such standards and practices should in turn be tied to observable behaviors that can be collected, documented, and assessed (Chism, 2007; Robertson, 2006). Protocols should also be in place to ensure that the reviewer is impartial as well as qualified to evaluate teaching, and reviewers should have the option to recuse themselves if there are personal or professional conflicts of interest between them and the observed instructor. Moreover, reviewers must have an in-depth understanding of the professional standards underlying summative PRT, and they should receive prior training in how these standards are to be measured and assessed.

Comparing Formative and Summative PRT

	Formative PRT	Summative PRT
Purpose	Feedback on teaching with the goal of informing future instruction, identifying areas for improvement, and demonstrating growth	Assessment of overall teaching effectiveness and achievement of professional benchmarks, used in making hiring, merit increase, tenure, and other personnel decisions
Reviewer	Usually faculty of similar rank and experience	Senior/experienced faculty or department head
Format	Informal, semi-structured, optional, confidential	Formal, structured, mandated, shared with department head or dean
Features	- Low-stakes - Flexible and informal - Easier to implement - Encourages honest discussions of teaching - Encourages reciprocal learning and collaboration between instructors	- High-stakes - Standardized and formal - Requires careful planning - Involves qualitative and/or quantitative assessment - Encourages learning from more experienced faculty
Product	Informal written report, narrative log, and/or rubric that encourages discussions on teaching and documents growth over time	Formal written report and/or rubric that offers a snapshot of instructor's effectiveness and achievements at a given point in time
Audience	Content of observation is confidential and not included in annual reviews or external assessments (unless instructor chooses to include it in their teaching portfolio)	Formal report is shared with departmental head or dean and used in annual review tenure, merit increase, and other high-stakes decisions

2.3. FORMATIVE AND SUMMATIVE PRT WORKING TOGETHER

Because formative and summative PRT have very different goals and stakeholders involved, separate protocols and instruments should be developed for each type of review (Robertson, 2006). Moreover, the two types of PRT may necessitate different kinds of reviewers, who may use different methods and criteria to observe and/or evaluate teaching. Finally, each type of PRT will lead to a different outcome or material product: Formative PRT is likely to result in confidential feedback aimed at helping the instructor improve their teaching, while summative PRT usually ends with a formal written report that is highly evaluative and is shared with program directors or other administrators, who may use the report to make high-stakes personnel decisions.

However, despite some key differences between formative PRT and summative PRT, the two types of review should be developed together and inform one another. In particular, both types of PRT should be grounded in the professional standards and best practices that underlie effective teaching in the program. Thus, the feedback solicited through peer observations should help the instructor move closer toward the same goals and benchmarks that are used in the summative evaluations.

And because formative PRT should lead to improvements in teaching, it should ideally "feed into" summative PRT, in which these improvements are measured and assessed (Chism, 2007). In other words, formative PRT is essential not only for self-improvement but also for preparing instructors for summative PRT, while summative PRT sets professional standards that instructors are continuously working towards and that they can reach more easily and systematically with the support of their peers. Consequently, even though each type of PRT has distinct features and functions, formative and summative review often work together within a larger, cohesive system (Robertson, 2006).

3. BENEFITS OF PRT

3.1. DEVELOPMENT OF TEACHING EXPERTISE

Teaching involves the interaction between a complex set of cognitive and interpersonal skills as well as pedagogical expertise and an in-depth understanding of course content. Such skills and expertise can be honed through practice, but research (e.g., summarized in Zeng, 2020) suggests that teaching experience alone does not guarantee improvement and innovation. PRT, and formative review in particular, can encourage the development of teaching expertise among

faculty in a program and institution. In continuously reviewing their own and others' teaching practices and exchanging contextualized feedback, instructors engaged in systematic PRT see teaching as an ongoing activity that needs to be studied and upgraded. Thus, a robust system of PRT can promote the pursuit of pedagogical excellence and a more explicit definition of teaching excellence (Chism, 2007).

3.2. DEEPER INSIGHT INTO TEACHING

PRT can help an instructor gain insight into their own pedagogical practices (Zeng, 2020). Formative PRT in particular encourages a close examination of one's teaching and a reflection on the various components that constitute effective teaching. Engaging in such metacognitive activities (which research suggests teachers need more experience with) can offer instructors a deeper understanding of their own practices as well as an opportunity to review and adjust their teaching. Moreover, when observing or evaluating others, reviewers are exposed to other pedagogical approaches and tools, thus enriching their own pedagogical expertise and repertoire. Thus, PRT benefits not only the teacher being observed, but also the reviewer (Chism, 2007).

3.3. EVIDENCE OF TEACHING EFFECTIVENESS

Given that much of teaching happens "behind closed doors" (Massy, Wilger, & Colbeck, 1994, as cited in Chism, 2007), PRT gives instructors a chance to share their teaching with others and to demonstrate their abilities. Formative PRT, in particular, has the potential to highlight and document an instructor's effectiveness and innovation in the classroom (Chism, 2007). Although the content of peer observations is generally kept confidential, an instructor may include the observation report and other documentation in their teaching portfolio or use such materials when applying for contract renewal, merit increase, teaching awards, etc. In sum, the process of PRT can enhance or complement an instructor's teaching portfolio as well as document their contributions to the pedagogical mission of the program.

3.4. COUNTERMEASURE AGAINST BIASED EVALUATIONS

In helping an instructor document their success and demonstrate growth over time, systematic PRT can act as a countermeasure against limited, narrow, or potentially biased forms of summative assessment. In particular, PRT can supplement end-of-semester course evaluations submitted by students who are not qualified to judge the pedagogical soundness of certain choices and methods implemented by the instructor (Chism, 2007). Similarly, administrators with limited experience in teaching in a given field may not be suited to evaluate an instructor's overall performance. Thus, peers who are experts in the subject area and relevant pedagogy and who have been trained to evaluate teaching effectiveness can potentially offer a more objective and nuanced description and assessment of an instructor's work.

3.5. ACCOUNTABILITY AND MAINTAINING STANDARDS

PRT is an important programmatic tool to ensure accountability and maintain high teaching standards within the academic unit and the larger institutional context. When instructors are regularly observed and when they are provided with feedback on their work, teaching becomes more open and transparent, stimulating improvement and innovation in the classroom. As such, PRT can encourage ongoing professional development among faculty of all ranks and prevent pedagogical stagnation. Additionally, systematic PRT ensures that teaching standards are continuously discussed, maintained, and revised as needed (Chism, 2007).

3.6. CULTURE OF CONVERSATION AND COLLABORATION

In encouraging open discussions of teaching, PRT helps instructors see teaching as a shared activity and a collaborative process. PRT can thus foster a culture of reciprocity and reflection, in which collegiality and collaboration are the foundations of effective teaching in the program. Moreover, honest and informative conversations on teaching can promote a departmental climate in which the pursuit of pedagogical excellence and high standards is an ongoing and worthwhile effort (Chism, 2007; Dawson & Hocker, 2020).

4. CHALLENGES TO PRT

4.1. LACK OF TRAINING

One of the greatest challenges in developing a robust system of PRT is ensuring that the reviewers have appropriate training in observing and, for summative PRT, evaluating teaching activities. All too often, faculty are expected to engage in PRT without prior training or experience in identifying or judging teaching activities. This can lead to unreliable, inconsistent, and potentially biased reviews (Chism, 2007; Dawson & Hocker, 2020). Consequently, it is crucial that reviewers receive sufficient and context-appropriate training, especially in cases when the results of PRT are used in high-stakes personnel decisions. Reviewers should also be made aware of different types of observational biases (Tobin, 2015) and ways to minimize them.

Lack of training

Fairness and equity concerns

Fear of criticism

Distrust of observations

Practical constraints

Limited research

Monitoring

4.2. FAIRNESS AND EQUITY CONCERNS

A reasonable concern about PRT, and summative PRT in particular, is the fear that the reviewer's assessment of another instructor's teaching may be influenced by personal preferences and biases. For example, a reviewer may prefer certain teaching methods and look unfavorably at methods that do not align with their own beliefs about effective pedagogy. Alternatively, reviewers may negatively evaluate instructors based on their own preconception about race, gender, religious affiliation, and so on. These concerns must be taken seriously and mitigated through multiple means (Chism, 2007). A key safeguard against biased judgments is requiring reviewers to undergo training. Another provision is to develop assessment tools that above all measure observable teaching behaviors and are directly tied to teaching best practices (Chism, 2007; Robertson, 2006). Thus, thoughtful development and implementation of PRT is the most important way of minimizing bias in faculty observations and evaluations.

4.3. FEAR OF CRITICISM

Because summative PRT is heavily evaluative, some instructors may perceive it as a criticism of their teaching practice or react to criticism with anxiety (Lasagabaster & Sierra, 2011). Others may regard teaching as a highly personal and subjective activity, and may thus feel uncomfortable sharing it with others (Robertson, 2006). To mitigate these concerns, feedback and assessment must be an integral part of a larger and ongoing conversation on teaching excellence within the program and institution. When teaching is regularly shared with others and examined with the goal of improvement and innovation, rather than criticism or personal put-downs, faculty are more likely to find it valuable and worthwhile.

4.4. DISTRUST OF OBSERVATIONS

Instructors who are not used to being observed or evaluated by others are likely to feel nervous about PRT and to distrust the process. In particular, negative prior experiences with PRT can lead to mistrust and frustration: For example, faculty may have been observed by untrained reviewers and may thus have the impression that PRT is unhelpful or unfair (see literature review in Zeng, 2020). Moreover, when faculty are not involved in the development of PRT instruments and protocols, they may be reluctant to embrace PRT and may regard this process as a top-down imposition or an attempt to "standardize" teaching and stifle pedagogical freedom (Lasagabaster & Sierra, 2011; Thomas et al., 2014). It is therefore of utmost importance that all stakeholders have a say in the development of PRT, that the PRT procedures and tools are developed with care, and that the process is transparent as well as continuously monitored.

4.5. PRACTICAL CONSTRAINTS

PRT may be viewed as yet another requirement that instructors have to fulfill. Especially for faculty with heavy teaching loads or on career tracks that emphasize research, seeking external funding, and so on, PRT may seem like an unwelcome and untimely burden (Chism, 2007). Moreover, when PRT is not planned or scheduled in advance, it might indeed be a stressful and

tiring experience for both the reviewer and the observed instructor. Because effective PRT, and especially summative PRT, requires careful preparation and training (Robertson, 2006), it should be conducted only as often as is necessary. Finally, PRT may even be recognized as service to the department, such as when conducting PRT might be the purview of a rotating committee.

4.6. LIMITED RESEARCH

Another limitation of PRT is the scarce empirical research on this topic in the higher education context (Thomas et al., 2014). Although PRT has been extensively studied in K-12 settings, the protocols and instruments that are used to observe and evaluate university instructors have not been systematically or empirically tested. Consequently, there are some doubts about the reliability and validity of the available instruments. In particular, it is not well understood to what extent and in what ways formative PRT at the tertiary level influences teaching practice or which summative PRT instruments are the most objective and reliable. Another gap in the literature is in understanding the potential effects that PRT has on university student outcomes. Because of the potential of PRT to improve both teaching and learning, there is a great need for a more systematic examination of the available PRT approaches and tools in higher education.

5. SEVEN PRINCIPLES OF EFFECTIVE PRT

5.1. PRINCIPLES OF EFFECTIVE PRT

Effective PRT, whether formative or summative, should be developed and implemented with several key principles and best practices in mind. The following list summarizes some key PRT characteristics and guidelines and is informed primarily by the recommendations outlined in Nancy Schism's foundational work, *Peer Review of Teaching: A Sourcebook.*

PRINCIPLE 1: APPROPRIATENESS

Above all, PRT must be designed with the broader institutional context and mission in mind, and ideally also adjusted to fit the structure, needs, and resources of the specific academic unit. Additionally, PRT should be grounded in and directly tied to the professional values and practices that are recognized within the institution and/or program as the foundations of effective teaching.

PRINCIPLE 2: TRANSPARENCY

The procedures and instruments used in both formative and summative PRT should be transparent. The guidelines should be written down and publicly available, so that all stakeholders can familiarize themselves with them and understand not only the procedures but also the motivations and pedagogical principles behind PRT. Ideally, all stakeholders should also have a say in the development and improvement of the PRT procedures and instruments. This will ensure faculty buy-in and trust in PRT procedures and outcomes.

PRINCIPLE 3: THOUGHTFULNESS

PRT procedures should be carefully researched and thoughtfully planned with the ultimate goal of promoting excellence in teaching and with efforts taken to mitigate potential limitations. PRT should also be conducted in a systematic manner, so that faculty can prepare for it in advance and see PRT as reasonable and fair. In sum, teaching observations and evaluations benefit faculty the most when they are seen as a fundamental and informative component of professional development.

PRINCIPLE 4: TRAINING

When PRT is regarded as fair and useful, it can promote wider departmental and institutional conversations about teaching and encourage collaboration among faculty. PRT should therefore be tied to other initiatives that frame teaching as a collaborative process. To sum up, PRT is part of an ongoing professional dialogue on the pursuit of pedagogical excellence.

PRINCIPLE 5: EXPLICITNESS

All stakeholders should feel prepared to participate in PRT. This means that both the observed instructors and the reviewers must understand the procedures and instruments used in PRT. In particular, reviewers should have access to step-by-step guides and other PRT tools, such as rubrics and sample written reports. Additionally, it is important to clearly define the reviewer's role and positionality during the classroom observation and/or evaluation.

PRINCIPLE 6: COHESION

An effective system of PRT should ideally include both formative and summative review. The two types of review are therefore often designed as part of a larger system of observation and evaluation. Although each type of review follows different procedures and has different goals, summative PRT should inform PRT, and formative PRT should help prepare instructors for summative assessment. In other words, the two types of review should work together, so that each feeds into the other and each mitigates the others' limitations.

PRINCIPLE 7: MONITORING

Finally, PRT instruments and protocols should be continually monitored and reassessed. They should be updated or adjusted as needed, so that they are appropriate, timely, and useful for all stakeholders. Regular monitoring and assessment of PRT can be conducted by a faculty committee or in consultation with external experts.

5.2. KEY QUESTIONS IN ESTABLISHING PRT

When designing and planning PRT, there are several important questions that need to be answered. The answers to these questions can profoundly shape PRT instruments and protocols, so they need to be thoughtfully considered, with input from all stakeholders.

WHAT IS THE ROLE OF PRT IN OVERALL FACULTY ASSESSMENT?

PRT should be situated within a larger system of documenting an instructor's professional development and evaluating their effectiveness (Robertson, 2006). The role of PRT in such a system must be clearly defined, especially when instructors are evaluated for the purpose of merit increase, promotion, and other high-stakes personnel decisions. Importantly, the results of summative evaluations should be carefully interpreted according to previously established and agreed-upon guidelines and used in combination with other appropriate measures of teaching effectiveness, such as instructor self-reflection and student course evaluations. In sum, when PRT is used for high-stakes purposes, its role and relative weight in overall faculty assessment must be carefully thought out.

WHO SHOULD CONDUCT FORMAL FACULTY EVALUATIONS?

The characteristics of the reviewer and their relationship to the observed instructor must be clearly defined, especially when instructors are being formally evaluated. One factor to consider is the reviewer's teaching experience: Summative PRT is generally performed by a more senior or experienced instructor, or in some cases by a department head. Another factor is faculty rank, appointment type, and career trajectory of both the reviewer and the observed instructor. Tenure-track faculty, lecturers, and adjuncts may all need to be observed and evaluated according to different procedures and criteria, so PRT designers must determine what constitutes fair and reasonable assessment of faculty at various ranks and with different teaching appointments. Some (e.g., Tobin, 2015) even recommend that high-stakes teaching evaluations be conducted by a group or a committee made up of various types of stakeholders.

HOW SHOULD REVIEWERS PREPARE FOR PRT?

Especially when PRT is used in high-stakes situations, reviewers need to be trained in properly using the tools and in mitigating potential biases and limitations of PRT (Chism, 2007). Reviewers should also have a chance to conduct mock observations and/or evaluations as well as to calibrate or norm assessment rubrics. Such training should be offered to reviewers on a regular basis, for example once a year or once every other year. As a result, training reviewers can be a time consuming and labor-intensive process, which is why a program may decide that it is appropriate to create a rotating PRT committee and to recognize the reviewers' work as service to the department.

HOW OFTEN SHOULD PRT BE CONDUCTED?

Frequency of PRT depends on faculty rank, appointment type, and other considerations. While informal observations may be conducted once or more per semester, summative evaluations are more labor intensive and stressful, so they should be conducted only as frequently as necessary and as part of regularly scheduled faculty review and promotion processes. An important consideration is the number of class sessions to be observed and evaluated. In general, the more class meetings a reviewer is able to observe, the more nuanced and in-depth their observations will be, which can in turn encourage further professional development and offer more evidence in evaluating teaching (Chism, 2007; Robertson, 2006). However, this affordance needs to be leveraged against the increased time and labor of multiple observations.

HOW SHOULD PRT BE CONDUCTED ACROSS PEDAGOGICAL CONTEXTS?

PRT instruments and protocols should be adapted to each pedagogical setting. For example, the differences between large lectures, clinical labs, and small discussion-style seminars necessitate somewhat different observation and evaluation methods. Moreover teaching observations and evaluations may vary by course modality, due to the differences between teaching face-to-face, online, and hybrid courses (Tobin, 2015). Nonetheless, because it is often impractical to develop unique tools and procedures for a wide range of pedagogical settings and modalities, PRT instruments can be flexible, so that they can be more easily adapted to various course types and modalities (Chism, 2007).

HOW SHOULD TEACHING BE ASSESSED IN PRT?

The criteria used to assess teaching should reflect the best pedagogical practices and core values of the institution as well as the best practices in the given program. As Chism (2007) points out, these criteria can involve a lower or higher degree of inference: **Low-inference criteria** are more descriptive in nature and may result in more consistent and more objective assessments, while **high-inference criteria** reflect judgments about the quality of teaching and can lead to more useful, but also potentially more biased, evaluations. Although both types of criteria can be insightful, low-inference criteria tend to lead to more reliable and neutral assessments.

WHO SHOULD HAVE ACCESS TO PRT DATA?

Some important questions in developing high-stakes summative PRT are: Which kinds of evidence of teaching effectiveness will be collected and assessed? Where will this data be stored and for how long? And finally, who will have access to this data? Generally, only the predetermined and relevant teaching activities and materials should be documented, and all data collected during PRT should be shared only with the appropriate stakeholders. All data should be stored in a predetermined, secure location, and the confidentiality of the data must be carefully maintained.

6. STEP-BY-STEP GUIDE TO ESTABLISHING PRT

STEP 1: ASSESS PROGRAM READINESS

The first step in developing and implementing PRT is to gauge readiness at the program (and sometimes also institutional) level. In particular, it is important to survey and synthesize stakeholders' prior experiences with PRT, their beliefs about and expectations of PRT, as well as their concerns or reservations about the process. If PRT is used across the university but is not standardized or universally mandated, it is also helpful to survey different colleges and departments about the tools and procedures for PRT they have used in the past.

Chism (2007) recommends that a survey of PRT readiness should include two types of questions: Those that identify **perceived rationales** for PRT, and those that highlight **potential reservations** about PRT.

- o **Perceived rationales** are designed to understand the arguments for and against PRT that are the most salient among stakeholders in the program. Such questions can reveal, for instance, if faculty find PRT useful primarily because of the increased accountability associated with observing and/or evaluating teaching or because of the possibility of external reward when effective teaching is documented.

- o Similarly, the survey results can reveal **potential reservations** and limitations of PRT--such as time and resource constraints or lack of program emphasis on honing its faculty's teaching skills--that are of greatest concern and should be explicitly addressed by PRT developers.

In sum, the results of the survey can paint a broad perspective of the departmental beliefs and attitudes toward PRT and help guide conversations on establishing pedagogical consensus, which constitutes the next step in establishing PRT protocols.
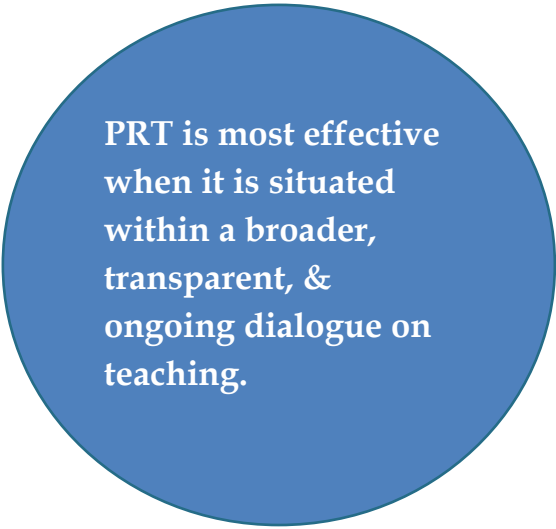
STEP 2: REACH PEDAGOGICAL CONSENSUS

Before establishing PRT protocols and instruments, it is also necessary to clearly define the pedagogical mission and teaching philosophy of the program, as well as identify the key characteristics of effective teaching and benchmarks for professional development. These should be embedded within and align with the larger mission of the institution, and then refined to reflect the specific goals and ethos of each academic unit. In other words, PRT depends on and necessities pedagogical consensus among faculty and other stakeholders.

Such pedagogical consensus can best be achieved through extensive and structured departmental discussions that include the voices of all stakeholders. Some (e.g., Chism, 2007; Robertson, 2006) recommend that these discussions be moderated by an external facilitator or a more neutral party, such as an expert from an institutional center for teaching and learning. An external, experienced moderator may be more successful at facilitating these complex conversations, especially in academic programs where teaching is not regularly shared and

discussed. In these contexts, there may be initial resistance to identifying a common set of standards and pedagogical practices, as they may be viewed as a threat to instructor privacy and autonomy. Moreover, in departments where prior PRT was conducted inconsistently or incompetently, there may be reasonable distrust of the process or misunderstandings about its purpose and usefulness.

As a result, establishing new PRT protocols may require a "culture change" (Chism, 2007, p. 19) within the program or the institution as a whole. This is because PRT is most effective and useful when it is situated within a broader, transparent, and ongoing dialogue on teaching. If such a dialogue is not common practice, this new way of thinking and talking about teaching may take some getting used to. Importantly, stakeholders should not be coerced or forced to embrace PRT, as this may result in resentment and distrust of the process.

> **PRT is most effective when it is situated within a broader, transparent, & ongoing dialogue on teaching.**

Instead, the program must present sound justifications for PRT and a list of the many ways in which it can support faculty, as well as offer thoughtful and honest responses to any concerns raised about PRT. And ultimately, the program should also explain the pedagogical and professional utility of embracing a more collaborative perspective on teaching and of reframing teaching as a practice to be regularly re-examined through self-reflection and external review.

Once reasonable pedagogical consensus is reached, the agreed-upon pedagogical values and principles of effective teaching should be summarized in writing. The written statement of teaching values should be made available to all stakeholders and should serve as the basis of PRT protocols and instruments (Chism, 2007).

STEP 3: DEVELOP PRT PROTOCOLS AND INSTRUMENTS

PRT instruments and protocols should be developed carefully and with input from key stakeholders. In some cases, a centralized institutional body may be put in charge of developing a new PRT process or revising an older one. The biggest advantage of convening a cross-departmental or other centralized institutional body to develop PRT tools is that it can save tremendous time and effort at the college and program level. However, for PRT tools to be applicable across various programs and contexts, they need to be broad and flexible, so that they can be easily adapted by each academic unit (Chism, 2007). It is then the responsibility of each department to ensure that PRT instruments and protocols align with the best practices and expectations in that program.

Whether PRT protocols and instruments are developed at the institutional or departmental level, one of the first decisions to be made is whether the system will include formative PRT, summative PRT, or both. The choice ultimately depends on the intended purposes of the PRT: Is the program primarily interested in encouraging faculty to reflect on and monitor their own teaching, so that they can improve over time? Or is the program aiming to evaluate instructors' overall teaching effectiveness and their alignment with program goals and mission? Formative PRT is appropriate in the former cases, while the latter situations call for summative PRT.

Since summative PRT is generally high-stakes and mandated, the instruments through which faculty will be evaluated need to be carefully thought out. In general, summative assessments of teaching necessitate the development of appropriate and consistent evaluation criteria. These criteria should serve as the foundations of effective teaching in the program and be mapped directly onto student learning outcomes (Robertson, 2006). Moreover, the evaluation criteria should be connected to observable and measurable teaching behaviors to increase validity and reliability of the assessments.

In addition to formal summative PRT, it is also important to provide faculty with opportunities to engage in informal peer observations that are opt-in and less structured. Setting in place a system of informal peer observations can help normalize teaching observations and honest conversations about teaching, as well as prepare instructors for more high-stakes, structured formal evaluations.

According to Robertson (2006), informal peer observations of teaching can take the form of a buddy, team, or mentor system, as summarized in the table below.

Buddy System
An instructor is observed by another faculty member of similar professional rank or experience. Buddies are self-selected based on shared interests or experiences. The peer collaboration can continue long-term and include sharing of teaching materials, resources, and research. The system can be opt-in or mandated, and faculty can be asked to document their collaboration in annual self-evaluations or teaching portfolios (however, the content of the peer discussions and observations should not be disclosed).

ADVANTAGES	RISKS AND DRAWBACKS
o Can lead to honest discussions and ongoing collaborations o Can help instructors build a closer professional relationship o Can lead to honest and nuanced discussions of teaching	o Risk of getting poor advice if buddy is not chosen wisely o Risk of overly positive feedback if buddies are close colleagues or friends o Potential difficulty in finding buddies if program is small or instructor is new

Team System

Instructors form small groups or clusters to conduct observations and discuss their teaching. Teams can be formed based on shared experiences and interests or with the intent of including different voices. Team diversity can lead to multiplicity of perspectives and experiences, thus encouraging richer and more complex discussions of teaching. As in the buddy system, peer self-select teams they want to join or create their own teams.

Advantages	Risks and drawbacks
o Can expose instructors to different perspectives and teaching methods o Can help instructors connect and find a support system larger number of team members can facilitate scheduling of observations Can lead to honest and nuanced discussions of teaching	o Risk of scheduling difficulties, as more team members means it's harder to find a time for all to meet o Potential difficulty in creating teams if program is small o Potential difficulty in finding buddies if program is small or instructor is new

Mentor System

A less experienced or new instructor is matched with a more experienced or senior instructor, who observes one or more classes and offers ongoing feedback and support. The goal of the system is for the veteran instructor to serve as a mentor and guide to the newer or less experienced teacher. Usually, the mentor is matched with a mentee, and the system may be mandated rather than opt-in.

Advantages	Risks and drawbacks
o Can help new instructors acclimate to the program can help less experienced instructors to learn from more senior faculty works well in larger programs with big pools of instructors	o Power imbalance between instructors can prevent reciprocal learning and hinder collaboration o Risk of being perceived as a threat or burden to instructors, if system is mandated o Risk of interpersonal conflicts if mentors are imposed

Once PRT protocols and instruments are agreed upon, they should be summarized in writing. The written summary of PRT tools and procedures should tie the criteria directly to the teaching values and principles espoused by the program and should be shared with all stakeholders.

STEP 4: IMPLEMENT AND MONITOR PRT

Once PRT protocols and instruments are finalized, they must be implemented thoughtfully and in a way that minimizes their potential drawbacks or limitations. Above all, to increase the validity and reliability of PRT tools, all reviewers should undergo specialized training that familiarizes them with the relevant protocols and instruments as well as offers them an opportunity to conduct a mock PRT (Chism, 2007). For example, reviewers can observe a live or pre-recorded class and perform a mock evaluation of the observed instructor. Afterwards, the reviewers should have an opportunity to debrief and share their evaluations with one another. The purpose of this activity is to allow for norming and calibration of the assessment tools (Robertson, 2006). Moreover, reviewers can also benefit from observing instructors who have been recognized for excellence in teaching. Observing and assessing such a class interaction can aid reviewers in identifying the types of teaching strategies and behaviors that are considered particularly effective in the given program or field.

Finally, PRT should be regularly monitored, re-evaluated, and revised whenever necessary (Chism, 2007). Continuous oversight of the process may be the purview of a department chair, an institutional center of teaching and learning, or a special committee responsible for conducting PRT. Moreover, the protocols and instruments should be carefully evaluated on a regular basis (e.g., annually or every other year) by an appropriate body qualified to assess the effectiveness, appropriateness, and fairness of PRT tools. In the event that such a body reveals any problems or evidence of bias, the PRT system should be accordingly revised.
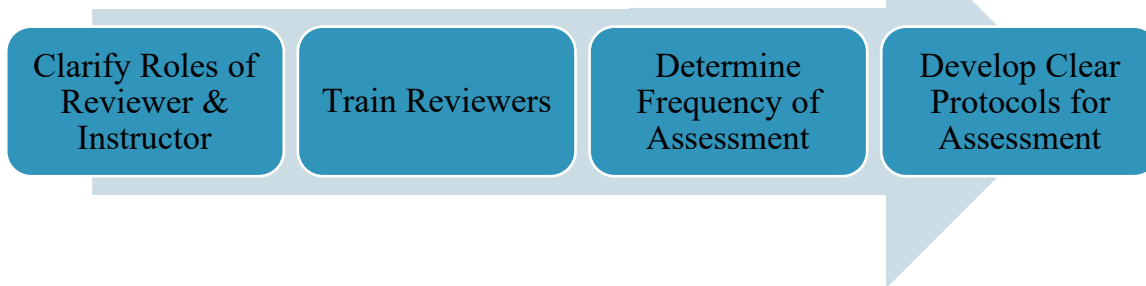
7. CONSIDERATIONS IN DEVELOPING PRT PROTOCOLS AND TOOLS

7.1. DEVELOPING PRT PROTOCOLS

In developing PRT protocols, there are several key decisions to be made. Firstly, the role of the reviewer and their relationship to the observed instructor must be clearly defined for both formative and summative PRT. For example, while formative review is generally performed by peer instructors, summative PRT is most commonly conducted by a more senior or experienced faculty (sometimes at a higher rank) or even by a program chair or other administrator. Some also recommend that PRT, especially when tied to high-stakes decisions, be conducted by a group or a panel of reviewers.

Secondly, PRT developers should consider the type of preparation and training that is required of reviewers. Especially in high-stakes situations, such as when the results of summative assessments are used in merit or promotion decisions, reviewers must receive appropriate training and practice with evaluating teaching (Chism, 2006). Such training should be offered on a regular basis, for example once a year or once every other year. Because of the time commitment involved in preparing for and conducting PRT, a program or college may create a rotating or intact committee tasked with observing and evaluating teaching, so that the participating faculty members' work can be recognized as service to the department.

Key Decisions in PRT Protocol Design

| Clarify Roles of Reviewer & Instructor | Train Reviewers | Determine Frequency of Assessment | Develop Clear Protocols for Assessment |

Besides specifying the role of the reviewer and the frequency of review, another important question to consider at this stage is how often faculty will be observed and/or evaluated. Frequency of PRT might depend on faculty rank, appointment type, and other considerations, and information about frequency of review should be communicated in faculty contracts and/or union contracts.

Finally, PRT developers must take great care in developing the protocols for observations and evaluations of teaching. In particular, clear instructions should be provided to both the reviewer and the observed/evaluated instructor for each of the different stages of the PRT process (Chism, 2007). For example, before a class is observed, the reviewer and observed instructor may need to first meet and set a broader context for the observation, as well as arrange a debriefing after the observation. It is also important to describe in detail the specific procedures used for formative and summative review as well as to provide a clear and comprehensive list of criteria that observers and reviewers should focus on (Robertson, 2006).

It is important that the criteria be mapped onto specific, observable teaching behaviors reviewers can collect and, in some cases, measure.

7.2. DEVELOPING PRT INSTRUMENTS

The criteria and standards that serve as the basis of observing teaching and evaluating teaching effectiveness need to be carefully thought out and described for the purpose of PRT.

Criteria are the elements of teaching that are observed and that may also be assessed. They are important to define, because they serve as a guide to the observer or evaluator during PRT. Chism (2007) suggests that teaching criteria can be further divided into broad categories—such design of course materials or class management—and these categories can then be linked with more specific, desirable characteristics or qualities, such as clarity and delivery of lecture. In presenting criteria as a list of key categories associated with individual qualities, PRT developers must identify elements of teaching that are the most important and most highly valued.

Next, in breaking the criteria down into concrete qualities, developers guide reviewers to narrow in on more specific aspects of teaching. An important decision to be made at this stage is whether these qualities should include a lower or higher degree of inference.

- o **Low-inference criteria** are more descriptive in nature and may thus be more consistent and neutral.

- o **High-inference criteria** ask reviewers to make a judgment about the quality of the observed teaching activity, so they may not be as useful in assessing teaching.

Although both types of criteria can be insightful, low-inference criteria tend to lead to more reliable and neutral assessments.

While the criteria describe the elements of teaching that are the focus of PRT, these expectations of effective teaching may be met in various ways by the instructor. Thus, it is important that the criteria be mapped onto *evidence*, or the specific, observable teaching behaviors that the reviewer can collect and, in some cases, measure. Chism points out that evidence of teaching effectiveness can be found not only inside of the classroom, but also in other tools and materials created by the instructor, such as the course syllabus, selected texts, feedback on students' work, and so on.

It is therefore important that PRT, especially summative review, make use of multiple sources of evidence to holistically assess an instructor's performance (Chism, 2007; Tobin, 2015). Finally, these criteria must be appropriate for the given course, pedagogical setting, and teaching modality, and should also align with program goals and values. As a result, the PRT instruments should be broad or flexible enough, so that they can be easily adapted to various situations.

8. FURTHER READING

Chism, N. V. N. (2007). Peer Review of Teaching: A Sourcebook. Anker Pub. Co.

Dawson, S.M, & Hocker, A. D. (2020). "An evidence-based framework for peer review of teaching."Advances in Physiology Education, 44, 26–31.

Robertson, W. (2006). Peer Observation and Assessment of Teaching: A Resource Book. The Center for Effective Teaching and Learning & Instructional Support Services at University of Texas at El Paso. https://www.utep.edu/faculty-development/_Files/docs/utep_peer_observation_booklet.pdf

Thomas, S., Chie, Q. T., Abraham, M., Raj, So., & Beh, L. (2014). "A Qualitative Review of Literature on Peer Review of Teaching in Higher Education: An Application of the SWOT Framework." Review of Educational Research, 84(1), 112–159.

Tobin, Thomas J. (2015). Evaluating Online Teaching: Implementing Best Practices. Jossey Bass.

Zeng, Lily Min. (2020). "Peer review of teaching in higher education: A systematic review of its impact on the professional development of university teachers from the teaching expertise perspective." Educational Research Review, 31, 1-16.