CHAPTER ONE

CHALLENGES IN EVALUATION OF ONLINE TEACHING

The growth of online learning has created an opportunity to reexamine teaching practices through a scholarly lens. The review and evaluation of teaching practices in general are sometimes performed in a pro forma fashion or only for summative reasons such as promotion and tenure decisions. Donna Ellis at the University of Waterloo in Waterloo, Ontario, sees teaching evaluation as a holistic enterprise: “teaching and its assessment should…be seen as scholarly activities. The review of teaching is an intentional process—one that is carefully designed, situated in context, and leads to interpreting teaching effectiveness based on multiple sources and types of evidence” (2012). Online courses offer us a rich variety of information sources from which to study and improve our teaching practices.

Evaluating Online Teaching is an attempt to bring together the disparate strands of research and best practices for evaluating online teaching practices in higher education. This book targets three key audiences in the higher-education community:

- Administrators who wish to create and adopt consistent standards for evaluating the teaching practices in online courses
- Faculty members who wish to know what criteria are evaluated as best practices in online teaching

COPYRIGHTED MATERIAL
Evaluating Online Teaching

- Support staff and teaching center personnel who work closely with faculty members and administrators to implement effective pedagogical skills across campus

The instructors, technologists, and administrators who read this book will walk away with a clearer understanding of four phases of the evaluation process for online teaching:

1. Developing evaluation skill sets
2. Creating and applying evaluation methods for administrative, peer, student, and self-reviews
3. Preparing faculty members for the evaluation process
4. Sustaining the online-teaching-evaluation life cycle

A Theoretical Foundation

Our approach in this book relies on the theoretical frameworks seen in Chickering and Gamson (1987) and, later, Chickering and Ehrmann (1996), in which best practices for teaching with technology are outlined in seven core principles:

- Encourage student-faculty contact.
- Develop reciprocity and cooperation among students.
- Use active learning techniques.
- Give prompt feedback.
- Emphasize time on task.
- Communicate high expectations.
- Respect diverse talents and ways of learning.

There are also several best-practice guidelines for online instructors that rely on large-scale studies performed during the 2000s and early 2010s, especially Savery (2005) and McCullom (2010), in which the best online teaching relies on several areas of focus:

- Make learning goals and paths clear to students.
- Use deliberate practice and mastery learning strategies.
- Provide prompt, constructive feedback.
- Provide a balance of challenge and support.
- Elicit active, critical reflection.
- Link inquiries to genuine issues of high interest to the learners.
• Develop learners’ effectiveness as learners.
• Create an institutional environment that supports and encourages inquiry.

Two studies in particular help to undergird the strategies advocated in this book. The Hanover Research Council (2009) established three categories of online instructional behavior: planning and development, teaching in action, and student assessment and data evaluation. The work of the Penn State Online Faculty Engagement Subcommittee (2011) established measurable faculty competencies for online teaching:

• Attend to unique challenges of distance learning.
• Be familiar with unique learning needs.
• Achieve mastery of course content, structure, and organization.
• Respond to student inquiries.
• Provide detailed feedback.
• Communicate effectively.
• Promote a safe learning environment.
• Monitor student progress.
• Communicate course goals.
• Provide evidence of teaching presence.

Based on all of these studies and approaches, this book presents a variety of real-world case studies to support various configurations of assessment techniques, including variables such as institution type, size, and union status. In order to set the stage for what readers will encounter in the book, it is useful to think back, for a moment, to the early days of online teaching.

The same core attributes go into good online teaching as go into good teaching in other modalities. All good faculty members teach their courses using clear expectations for themselves and for their students. Good teaching relies on a theoretical and logical framework. Best practices for teaching in any medium include the following:

• Making interactions challenging yet supportive for students
• Asking learners to be active participants in the learning process
• Acknowledging variety in the ways that students learn best
• Providing timely and constructive feedback

Online and face-to-face teaching share core teaching behaviors. Throughout this book, we will show that the differences between online
and face-to-face teaching have to do with the specific techniques employed to demonstrate those behaviors. Our emphasis will be to showcase ways of evaluating online teaching that serve two distinct ends: first, comparing the performance of online instructors against themselves and their peers, regardless of the modality in which they teach, and second, identifying practices unique to online teaching.

The evaluation of teaching in general has undergone several similar cycles between a single dominant teaching method and a plurality of teaching methods. Some examples highlight how the growth of online teaching can be seen as the latest iteration of that cycle.

In the 1970s and 1980s, the model for evaluating teaching was its contribution to “faculty vitality.” The dominant mode of teaching in higher education was the synchronous, classroom-based course. Faculty members were judged on how well their students were able to perform on examinations and in the workplace. In a 1978 issue of *New Directions for Institutional Research* called “Evaluating Faculty Performance and Vitality,” Bogen upholds the “vigor” of faculty members’ interactions with students—the “fit” between students’ expectations and faculty members’ interests—as the key predictor of their teaching effectiveness. He says that “we could learn much from the coach, who in large measure owes his career not to his technical competence but to the success of his securing a match between the student-athlete’s needs and the institution’s ability to satisfy those needs” (Bogen, p. 62). Similar to the climate today, in 1978 Bogen saw higher education as “entering an era of declining enrollments and relatively reduced public expenditures” (p. 64). The best teachers were seen as those who could hold and keep the interests of their students through the process of “fitting” the goals of the classroom to align best with the students’ expectations.

The evaluation principles outlined in Seldin’s *Successful Faculty Evaluation Programs: A Practical Guide to Improve Faculty Performance and Promotion/Tenure* (1980) are focused exclusively on face-to-face teaching, yet, despite language about voice volume and eye contact, the principles themselves are robust and can be applied to online teaching with only slight modification.

- **Instructor knowledge.** Does the instructor exhibit content mastery, breadth, depth? Has he or she religiously kept abreast of the discipline? Is his or her subject expertise appropriately demonstrated in the classroom? Is his material appropriate to the level of the course? And appropriate to the level of student preparation?
• **Method of instruction.** Does the classroom presentation exhibit clear signs of planning and organization? Is the material clearly presented? Is class time used efficiently? Does the instructor adapt methods to meet new situations? Is special or supplementary material (e.g., handouts) effectively handled by the instructor? How much critical thinking and analysis by students in the class does he or she elicit?

• **Instructor-student rapport.** Does the instructor demonstrate fair and equitable treatment of the students in the class? Are questions answered in a direct and understandable manner? Does the instructor betray any sarcasm in dealing with students? Does he or she encourage student involvement when relevant? In dealing with student viewpoints contrary to his or her own, does the instructor encourage full and fair class discussion? Does he or she appear receptive to student suggestions? What is the best description of the instructor-student relationship as exhibited in the classroom?

• **Teaching behaviors.** Is the instructor’s oral delivery too rapid, too slow? Can he or she be heard easily? Is his or her choice of language understandable to the students? Is the classroom activity level too high, too low? Does the instructor at times express himself or herself nonverbally? Does he or she exhibit any distracting mannerisms? Does he or she maintain eye contact with students? Is he or she blind to any part of the classroom and fail to call on students in that area?

• **Enthusiastic teaching.** Does the instructor exude enthusiasm for the subject? Does he or she show signs that he or she enjoys teaching? How hard does he or she try to stimulate students to master the subject? Does he or she encourage informal discussions with students before or after class?

• **Concern for teaching.** Does the instructor show interest in improving his or her teaching by an analysis of classroom performance and by innovating new teaching techniques? Does the instructor make the subject relevant to students and tie it to recent developments? Does the instructor seek out colleagues for discussions on teaching improvement?

• **Overall.** What parts of the teaching seemed particularly to enhance the learning process? What suggestions are needed to improve the teaching performance? Does the instructor merit recommendation to the students you advise? Was the classroom observation under circumstances that permitted accurate judgment on the efficacy of the teaching-learning process? How would you rate this instructor against others in the department? In the institution?
By the early 1990s, the personal computer was hailed as a means of advancing knowledge and decried as a source of extra work for already-busy faculty members. Computers were seen as adding to existing faculty-member workloads, especially with the advent of e-mail systems, which created a new and demanding communication stream outside of the traditional classroom-and-office-hours scenario. Several theorists began to discriminate among the categories of teaching behaviors and the *techne* (τεχνη) of the specific methods used to achieve those broader results, methods that were in flux with the advent of new time-shifting technologies.

Braskamp and Ory’s seminal work, *Assessing Faculty Work* (1994), defines “assessment” in light of its etymological roots as a process of “sitting beside,” with the evaluator “judging and providing feedback about the other’s performance,” in a “dialogue and discourse, with one person trying to understand the other’s perspective before giving value judgments” (p. 13). This squares well with their examination of what goes into “the work of teaching,” which is rooted in the precomputer practices prevalent up until the late 1980s. Although the authors include items such as “developing computer exercises” (p. 41) in their catalog of faculty teaching work, the assumption is that computers are stand-alone machines that can be used like other development tools to supplement face-to-face instructional methods. It is also worth noting that Braskamp and Ory advocate the holistic evaluation of teaching—all faculty-produced teaching tasks should be assessed via methods as diverse as written appraisals, rating scales, interviews, observations, videotaping, external indicators (e.g., teaching awards), achievement measures, and teaching portfolios (pp. 165–236). All of these methods continue to be best practices for incorporating teaching evaluation into a larger institution-wide scheme of evaluation and planning.

The American Association for Higher Education’s (AAHE) “9 Principles of Good Practice for Assessing Student Learning” from 1996 hold equally in the face-to-face and online environments:

1. The assessment of student learning begins with educational values.
2. Assessment is most effective when it reflects an understanding of learning as multidimensional, integrated, and revealed in performance over time.
3. Assessment works best when the programs it seeks to improve have clear, explicitly stated purposes.
4. Assessment requires attention to outcomes but also and equally to the experiences that lead to those outcomes.
5. Assessment works best when it is ongoing, not episodic.
6. Assessment fosters wider improvement when representatives from across the educational community are involved.
7. Assessment makes a difference when it begins with issues of use and illuminates questions that people really care about.
8. Assessment is most likely to lead to improvement when it is part of a larger set of conditions that promote change.
9. Through assessment, educators meet responsibilities to students and to the public. (Astin et al., 1996, pp. 1–3)

Indeed, AAHE’s principles help online teachers to avoid the common mistake of facilitating online courses as “read and respond” analogs to postal correspondence courses (albeit at a faster pace). These values preface the current focus areas in online teaching on authentic assessment (getting learners involved in real-world problems and issues) and engaged discovery (in which teaching happens within a sense of shared inquiry).

The pendulum is now swinging back toward seeing “teaching as teaching” regardless of the delivery medium. Although indications that students are learning, such as those advocated by the New Leadership Alliance for Student Learning and Accountability (2012), remain fairly constant, faculty members who teach online and the administrators who evaluate them are today able to take advantage of deeper and more detailed records of their teaching. Gardiner, Corbitt, and Adams (2010) advocate folding online teaching into institutions’ overall program assessment model, and several recent theorists have called for “crosswalks” to map online teaching methods against cognate face-to-face techniques.

For example, we find that most of the online teaching evaluation instruments in use today are created to evaluate content design rather than teaching practices—or, at best, include the evaluation of teaching practices alongside the evaluation of design concerns. The assumption that the instructor of an online course is also its designer is a common belief that must be challenged in these times of master-course development, in which one person designs the content and environment and many others teach from it. Arreola (2007) echoes this assumption when he defines teaching as occurring within five broad skill dimensions: “content expertise, instructional design skills, instructional delivery skills, instructional assessment skills, [and] course management skills” (p. 19). Piña and Bohn (2013), conversely, identify nine representative predictors of faculty teaching quality, all of which center on interaction between faculty members and learners.
Even when researchers are purposely looking to create lists of online-only teaching behaviors, such as the frequency of instructor log-ins to the course environment, the list (including our own) inevitably includes general best practices that could apply in various teaching modalities. This is cause for hope as well as reason to continue to examine what is unique about teaching online.

### The “Traditional” Teaching-Evaluation Process

In 2004, Tom Tobin was hired by a for-profit university to teach business English and communication courses at its newly opened location in Pittsburgh, Pennsylvania. Although taught in a face-to-face classroom, all courses were to be supported by online elements. Lecture notes, preparatory discussions, feedback on student work, and assignment grades were all expected to be handled within the university’s learning management system (LMS).

As part of the university’s procedures, Tom was slated to have his teaching observed by the chair of the communication department. The process was familiar to him, based on the many years’ worth of observations he had already been part of, both as a graduate teaching fellow and then as an early-career professor.

The setup was fairly routine: the department chair selected a class meeting date and time and notified Tom that his class would be observed. With the notification e-mail message, the department chair attached a copy of the evaluation instrument that would be used and asked Tom to reply with a copy of his syllabus. The message ended with a request for Tom not to make any special changes to the content or style of teaching: “Please forget that I am even there, and just teach a normal class.”

On the evening of the observation, the department chair entered the classroom a few minutes early, sat in the back of the room, and readied the observation form. A few students asked Tom who the visitor was, and he let them know that part of the university’s commitment to quality education for students entailed an observation of the teaching methods of all instructors. Tom taught the class, the department chair wrote on the form, and the department chair left the classroom at the mid-point break in the ninety-minute class period.

A week after the observation, the department chair invited Tom to meet to review the results of the observation. During that conversation, the chair provided Tom with a copy of the class observation form (see Figure 1.1). The observation form seems to indicate that Tom was
FIGURE 1.1. CLASS OBSERVATION FORM, 2004

Class Observation Form – iOptimize Courses

Instructor: [Name]
Course: [Course Name]
Center: [Center Name]
Room: [Room]

No. of Students: [Number]
Observer: [Name]
Date: [Date]
Time: [Time]

Top(s) Covered: [Topics]

I. Classroom Environment:
Temperature: [Temperature]
Lighting: [Lighting]
HVAC: [HVAC]
White Board [Blank]
Overhead [Blank]
Tables/Chairs [Blank]
Noise [Blank]
Other [Blank]

II. Instructions: Please circle one number. Rating Scale: 4=excellent, 3=good, 2=fair, 1=poor

<table>
<thead>
<tr>
<th>Onsite Component of Course</th>
<th>Characteristic</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction: (questioning strategy, probing for understanding, use of student names, involves all students, thought-provoking)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation: (organization, clarity, summarizes development of theories and concepts, reference to prior and future weeks' topics, concrete examples)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homework Review: (practically relevant, teaching vs. answer-giving, calls on students)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom Management: (in control, use of board, other audio-visual, comfortable pace, handles student questions well)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitality: (energy, enthusiasm, movement, volume, eye contact, humor, connectedness, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Participation: (involuntary, active vs. passive, interested vs. bored, voluntarily, prepared, enthusiastic)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionalism: (appearance, demeanor, respect for students, fair, role-model, responsive, degree of preparation)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical Linkages: (relevant examples, practical insights, &quot;tricks of the trade,&quot; appropriate applications, integration with other functional areas)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

[Observation Details]

---
doing very well, with high marks in all categories. The department chair offered praise for Tom’s teaching, using words that echoed the suggested criteria on the form itself.

Tom was pleased that he had been able to start his teaching at this new university in such a strong manner. When he got home, he looked again at the form, and noted that it continued onto the other side of the paper. The back of the sheet held categories for the online components of the course:

- User activity
- Online component
- Discussion topics
- Feedback
- Visual appeal
- Communication style
- Posted items
- Course balance

There were also write-in spaces for strengths, opportunities for improvement, progress from previous observations, follow-up items, and an overall class rating.

The back of the form was completely untouched (see Figure 1.2). The department chair eventually confessed that because he had not himself taught using the institution’s LMS, he didn’t feel qualified to rate Tom’s use of its tools. The chair also mentioned that he would welcome training or a “walk through” so he could get familiar with the expectations for the use of the LMS.

Deans, department chairs, faculty members, and students rate and evaluate teaching at their institutions mostly through home-grown processes and forms. Although these are often constructed to help observers and raters to provide meaningful information, it is often the case that even now, many years after the example provided in Figure 1.2, little training is provided for those using the evaluation instruments. With online teaching, observers have many more streams of information they can observe for evidence of teaching effectiveness, and there aren’t clear boundaries about what counts as teaching behaviors—or even how much time an observer should spend performing an evaluation.

Further, it is common practice for state and provincial governments to require K–12 educators to receive training in not only the discipline(s) in which they are going to teach but also in pedagogy, educational psychology, and classroom management. In higher education, however, institutions typically require one to master only a specific discipline’s content in order
FIGURE 1.2. CLASS OBSERVATION FORM, REVERSE, 2004

<table>
<thead>
<tr>
<th>Component</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rigor</strong>: (appropriate level of rigor: focus on how and why (not just what); focus on understanding, decision-making, implications, problems, and analysis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Online Component of Course</strong>:  (use one set of numbers for each evaluation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Activity: (2.5 hours/week for faculty; 2 hours/week for students)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Online Component: (other than the threaded discussion, e.g., web research, report posting, in document sharing, weblog postings)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Discussion Topics: (TCO-related, uses broad-based questions that generate good discussion, including follow-up questions, rather than discrete answers)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Feedback: (timely; encouraging and supportive; responds to questions asked)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Visual Appeal: (*curb appeal, aesthetically pleasing)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Communication Style: (professional, grammatically correct, positive, proper &quot;etiquette&quot;)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Posted Items: (syllabus, assignments, other postings consistent with quality standards)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Course Balance: (appropriate balance between online and online activities)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Strengths:____________________________________________________

Opportunities for Improvement:________________________________

Progress Noted from Previous Observations:______________________

Follow-Up:___________________________________________________

Overall Class Observation Rating: ____________ (i.e., 3.15 . . . 4 excellent; 1 poor)

------------------------------------------------------------------
to teach in that field. Although the trend is starting to reverse, it remains true, even today, that higher-education faculty members rarely begin their teaching careers with any hands-on training in how to teach, and few avail themselves of such opportunities when they are offered on campus. This becomes a twofold concern for evaluation of online teaching: instructors who may or may not have been trained in pedagogical techniques are evaluated by peers and administrators who may or may not have a scholarly vocabulary (or the technical background) to be able to evaluate their online teaching practices.

Effective college and university teaching practices are typically learned on the job, often by trial and error, and are supported by a variety of evaluation methods that range from student feedback on end-of-course surveys to formal administrative observations in the classroom. Numerous tools and evaluation instruments have been developed and put into use to support these efforts, but the use of these tools varies widely (as evinced by Tom’s former department chair). Institutions apply the results of various evaluations methods to such disparate ends as the promotion-and-tenure process, future-course staffing decisions, instructor remediation, and awards and recognition.

Add to this situation the variable of modality (teaching online via asynchronous or technology-enhanced means) and the conversation about defining best practices becomes especially diffuse. April Bellafiore of Bristol Community College recently put the problem in sharp focus:

I think we’ve tried for many years to say online and face-to-face are the same—no significant difference. It’s the same as face-to-face; don’t treat it any differently. We’re going to go set up a classroom face-to-face, and we’ll also set up a “classroom” online. Where we’re struggling is that we know it is different. It’s been framed for our faculty that it’s a different modality, but we can’t pretend that they really are the same. Now, workload and effort are really coming into play on the front end, and evaluation effort and measurements on the back end. We have a very strong union as well, and we’re definitely heading down that road to have conversations about workload and expectations. (Baird & Piña, 2014).

As online education continues to grow in popularity and becomes a mainstream practice at many institutions, faculty members and administrators face the challenge of effectively measuring the quality of online teaching. Strategies and tools designed for face-to-face classrooms are too often applied in a one-size-fits-all manner to the online teaching
Challenges in Evaluation of Online Teaching

environment, leaving both reviewer and instructor with an ill-fitted and incomplete analysis. Designing and implementing strategies to evaluate online teaching necessarily differs from designing and implementing those used in the face-to-face environment; they vary as a function of the nature, purpose, and focus of the evaluation. Comprehensive evaluation models must balance the summative data available via formalized assessments with the formative feedback available via informal processes.

Challenges for the Evaluation of Online Teaching

Institutions share a few overarching concerns when it comes to establishing and implementing various evaluation methods for online teaching:

- What is the online equivalent to visiting a colleague’s classroom for ninety minutes?
- Which elements in an online class count as “design” and which as “teaching”?
- How much time should peer and administrative evaluators spend observing online teaching?
- Should observation of online teaching take place during or after the course?
- To what purposes will student ratings, self-evaluation, peer evaluation, and administrative observation be used?

There are also some common red herrings, concerns voiced by faculty members who have not yet experienced evaluation of their online teaching:

- If the LMS suffers technical glitches, does that mean lower student-rating scores for the instructor?
- Who is qualified to perform various aspects of evaluation?
- Shouldn’t my peer reviewer be from my own department?

Although these concerns can be addressed easily, they offer insight into the mind-set challenges for adoption of online-teaching evaluation in a given unit or across campus. In Evaluating Online Teaching, we will examine these and several additional areas in which evaluating online teaching must differ from evaluating face-to-face teaching.
Scope

In the typical evaluation of the face-to-face classroom, feedback from students, peers, and administrators is based on in-class performance only. Often, this occurs in response to a one-time observation designed to represent typical teaching behavior. Pre- and postwork (by the students and the instructor) is not observed or evaluated. For example, communications that take place outside the face-to-face classroom (e.g., office hours, e-mail messages) are not observed.

For online teaching, however, the lines between in class and out of class are blurred. Interactions among students, instructors, and materials take place in a variety of locations. Observers need to look at core course content (e.g., lecture notes, videos, readings, assignments) in order to provide context for the class, but they also need access to course communication systems (e.g., discussion forums or even e-mail) in order fully to observe student-student and student-instructor interactions.

Time Equivalence

Because the limiting factors for face-to-face classes are time and place, evaluation instruments often use time and place as the outside boundary of all observation activities. For instance, a peer evaluator might observe one ninety-minute class session in a specific room. Online teaching is not restricted to a specific geography or time frame, and it is thus more challenging to find an equivalent for evaluation.

There is a broader issue inherent in this concern as well. Because face-to-face class observations are time based (the reviewer sits in on a single class period), the assumption may be to find the equivalent of that time-based observation for online teaching as well. Class period is not a meaningful term within an online course, and observers must determine how much content to observe (e.g., one week or unit worth of materials and interactions). This is actually an opportunity to broaden the scope of online-teaching evaluation and to detach it from time-based criteria: the observation becomes more encompassing and includes a greater number of dimensions.

Separating Teaching from Design

In a face-to-face classroom, the teaching practices that are observed are easily separated from the tools used in the process of teaching. For example, speech patterns, body language, voice tone, eye contact, and conversational engagement are all factors that are identified as part of
the observable teaching practices of the instructor, whereas textbooks, whiteboards, and PowerPoint presentations are seen as tools used in the process of teaching. These two components—instructor behaviors and course materials—can be evaluated in a face-to-face scenario, but it is clear what is under the control of the instructor and what is not when we are looking at face-to-face teaching.

Conversely, in an online environment, it is more challenging to separate elements that represent teaching behaviors from those that represent foundational or design-related parts of the course. For instance, in an asynchronous discussion tool, are the instructor’s initial prompts to be considered as design elements or evidence of teaching taking place? How much instructor interaction is deemed to be “conversational engagement”? Cheryl Hayek (2012) advocates some “dinner party” suggestions for gauging optimal faculty discussion participation, whereas the rule of thumb at many for-profit institutions is to look for between 10 and 15 percent of the total posts coming from the instructor (DeVry University, 2009). Further, lecture notes, multimedia (e.g., audio and video podcasts), and narrated PowerPoint slide presentations could be seen either as tools placed within the design of the course or as evidence of teaching behaviors.

Another aspect that muddies the waters is the question of who created the content that is being reviewed. In many institutions, the faculty author who creates the content for an online course (including, say, the initial prompts for weekly discussion) may not be the same faculty member who teaches the course each term. There is already a well-established literature on creating and evaluating quality online-course materials and designing quality online-course environments, making it necessary to distinguish between “design” elements and “teaching” behaviors.

For instance, as recently as 2004, researchers have assumed that the designer of the content of a course is also the instructor. Walvoord (2004) lists three foundational questions in her “guidelines for the evaluation of teaching”: “1. Are the learning objectives of the course being met? 2. Are the course material, concepts, and activities rigorous, current, relevant for students’ needs? 3. Do students perceive themselves to be well taught?” (p. 94).

**Measurement Equivalence**

Chickering and Gamson’s “Seven Principles for Good Practice in Undergraduate Education” can be observed in any course, no matter the delivery method. In many evaluation instruments, such common elements as the seven principles make for valid comparisons between face-to-face
Evaluating Online Teaching

and online sections of the same course. Teaching to the course objectives, level of engagement with students, and responsiveness to learner concerns are all teaching behaviors that evaluators can observe regardless of the course modality. For example, Buller (2012) establishes six general principles for evaluating faculty teaching, regardless of delivery mode:

**Practice 1:** Follow your institution’s established procedures to the letter.

**Practice 2:** Documented evidence should focus on observable behaviors and verifiable results, not on general impressions of the person’s attitudes, opinions, or personality.

**Practice 3:** Use statistical measurements appropriately.

**Practice 4:** Make actual judgments when conducting summative evaluations; don’t merely summarize data.

**Practice 5:** Remember that you already know more than you think you do.

**Practice 6:** Don’t overlook the value of 360-degree review processes.

( pp. 23–38)

However, evaluation instruments often contain elements that are specific to a given delivery method, such as evaluating instructor presence via tone of voice, speed of verbal delivery, and body language. Many institutions are currently using evaluation instruments that ask observers to rate instructors’ use of classroom space, availability outside of class meetings, and use of class time—all of which are difficult or impossible to apply to online courses.

In this book, we will suggest some ways to redesign existing evaluation instruments to work around these measurement-equivalence challenges, such as focusing on the spontaneous interactions in online course environments, evaluating the quality of materials created by instructors in response to student questions (i.e., noncanned materials), observing the frequency and quality of instructors’ interactions with students, measuring the degree of conceptual scaffolding within course discussions, and comparing the level of course inquiry at the higher levels of the cognitive domain of Bloom’s taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956).

As mentioned previously, instructors also have doubts about evaluation of their online teaching, which we will explore. Will technical glitches experienced in the online-course environment translate to lower end-of-semester student ratings? How well can a department chairperson who has never taught online courses evaluate the teaching practices that
are happening in online courses? We will examine ways that various measurement techniques can help to answer these kinds of questions as well.

Instrument Applicability

Despite significant progress in the field in the past several years, there is still a challenge in creating, adapting, or adopting evaluation instruments that are broadly applicable to instruction in online courses. This book will examine a number of existing evaluation instruments in order to establish best practices for formal and informal self-, peer, and administrative review of online teaching. Many higher-education institutions continue to choose to create their own instruments because of the perceived unique features of their circumstances. Further, many existing evaluation instruments were originally designed for face-to-face course applications. Adapting existing instruments for use in evaluating online courses presents several forms of potential measurement error and unintentional bias, and we will offer ways to keep these factors in check.

Scalability

Online learning programs are more scalable than their face-to-face counterparts. There are no limits on needing to create more campus space for teaching and residential student populations. Indeed, online courses offer higher-education opportunities to students regardless of schedules and geography. Likewise, online programs attract instructors from beyond the traditional pool of campus-based faculty members. Many large online programs depend heavily on contingent faculty members. The implementation and administration of rating and evaluation methods becomes more challenging for larger and more-distributed online programs. Traditional face-to-face measures, such as real-time peer-evaluation review meetings, are not often designed to be scalable or to target contingent faculty populations. Throughout this book, we will offer case studies and best practices for scaling evaluation programs and for targeting specific instructor populations.

Quality Improvement versus Record Keeping

Why do we undertake evaluation programs at all? This question is not unique to online-teaching evaluation. However, online teaching offers a way to move beyond the formal evaluations that are typically conducted
for reasons of hiring, continuance, promotion, and tenure. Too often, the real reason we perform evaluations is that our accrediting bodies require us to provide data based on course outcomes, and we keep the records accordingly.

We will examine ways to create and nurture a culture of evaluation that includes the use of evaluations for instructor use only, such as those provided at midterm and those used as ad hoc “how am I doing?” checks. Informal and self-assessments enable campuses to build systems that get us back to the foundational reason to perform evaluation: to make sure we are providing the best quality educational experiences to our learners.

Looking Ahead

Evaluating Online Teaching offers readers a four-stage process for establishing a culture of evaluation on their campuses, with online learning as a key component of that culture. This book provides actionable steps for campus faculty members, administrative leaders, and support staff to take in order to plan for, adopt, adapt, and institutionalize the evaluation of online teaching.

In part 1, we explore the planning process for adopting evaluation of online teaching at an institution. Chapter 1 (this chapter) introduces several challenges specific to online-teaching evaluation that campus leaders need to understand, acknowledge, and plan to address. Chapter 2 introduces the general context in which to evaluate online teaching practices. We provide readers with ways to define campus needs, important questions to ask about the institution itself and the implementation and potential impact of an evaluation program, and case studies to help demonstrate how other institutions have carried out the planning process. Chapter 3 offers tips and strategies for developing an evaluation structure amenable to your institutional context. We examine several faculty and administrator skill-building initiatives that helped to lay solid foundations for successful online-teaching evaluation programs.

Part 2 focuses on formative evaluation mechanisms that help to broaden the impact of evaluation beyond collecting data for accrediting agencies and the promotion-and-tenure process. Formative evaluations help faculty members to improve their teaching practices. Chapter 4 explores how instructors’ personal use of frequent small-scope student ratings of online teaching demonstrates an institution’s culture of assessment and improvement. Faculty members who engage in an
informal student-review process typically see their student-retention and student-satisfaction rates increase by the time the institution’s formal end-of-semester surveys roll around.

We also examine the other side of the informal-evaluation coin, offering practical advice for setting up a program of peer and self-review for online teaching practices that faculty members will want to adopt because it makes their online-classroom experiences smoother and more rewarding; because it doesn’t count toward formal decisions about hiring, faculty retention, promotion, or tenure; and because it offers faculty members greater academic freedom by providing insight into how their teaching techniques are actually having an impact on their students’ learning.

In part 3, we focus on data gathering for summative evaluation processes. Summative evaluation helps institutions to make personnel decisions such as promotion, tenure, continuation of contingent faculty members, awards, and remediation. Chapter 5 reviews several tools for performing the most commonly used summative-decision data-collection method: end-of-course student ratings. Chapter 6 outlines challenges and solutions specific to formal administrative evaluation of online teaching, especially when department chairs and faculty committees are tasked with deciding whom to rehire, promote, or give tenure to. Chapter 7 provides concrete tools and techniques to address the administrative-evaluation challenges from chapter 6 by examining several case studies and suggesting a set of action steps and policy statements that can help readers’ campuses to implement administrative review processes smoothly. Chapter 8 examines the promise—and the hype—surrounding the use of data analytics for evaluative purposes.

Part 4 of this text offers concrete steps that readers can take in order to customize the evaluation strategies to fit their own campuses. Chapter 9 offers strategies for incorporating student, self, peer, and administrative review methods into the institution’s overall assessment program of online-teaching practices. We examine several existing evaluation instruments and explore the debate between using one instrument to collect data for all course-offering modalities and using separate instruments, each specific to a different way of offering courses. We present a concise program for setting up and running a program for collecting student-rating data about online teaching, often part of the largest effort to collect evaluation data on campus.

Once institutions have created a culture of evaluation, the final stages of the process center on sustaining and strengthening that culture. Perhaps the most challenging part of this book is chapter 10, where we will
offer strategies for implementing an online-teaching evaluation process on campus. Institutions that are creating brand-new evaluation programs may even have an advantage over places where there is already an evaluation process in place for face-to-face teaching practices because it enables such campuses to align their evaluation outcomes across delivery modalities. We will help readers to plan and initiate an online-teaching-evaluation program on their campuses, examine and offer advice on avoiding common missteps, and suggest how to recruit the skeptics (converts make the most convincing advocates).

Evaluation is a cyclical process, and the ease of data collection with online teaching makes a compelling case for building the entire life cycle of online-teaching evaluation into the institutional business as usual—from feedback to data gathering to reporting to interpretation to experimentation, change, and adaptation. Chapter 11 defines the various stages of the online-teaching-evaluation life cycle, and provides concrete steps that readers can take in order to move beyond bare-bones data collection for accreditation purposes into establishing a true culture of evaluation on campus.

Readers should use this book as a way to build a solid foundation for evaluation on their campuses. Once an institution has a robust evaluation culture in place, what resources and paths does that free up for accomplishing other goals related to their evaluation and quality improvement efforts? Chapter 12 and the companion website (www.wiley.com/go/tobin) for the book offer some possibilities for next steps, including a ten-minute take-away version of the book’s lessons, worksheets for readers to complete in order to craft their efforts toward their specific institutions, and links to resources for sharing knowledge with the growing community of online-teaching professionals, support staff, and administrators.

Conclusion

The differences between face-to-face teaching and online teaching reflect modality and scope—but not quality. It has been well established that faculty members can be just as effective teaching online as in the classroom (Russell, 2001), and best practices for teaching online have been evolving since the mid 1990s. Online teaching has existed for long enough that we can now differentiate between traditional online teaching (with its defined course duration, number of students, and asynchronous focus)
**FIGURE 1.3. CHECKLIST: CAMPUS READINESS FOR ONLINE-TEACHING EVALUATION**

Please indicate your level of agreement with the following statements. “My campus today has ...”

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a process for evaluating face-to-face teaching.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a method for communicating evaluation results to faculty members.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a method for communicating general student-rating results to students.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>specific tools for collecting information about teaching.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>faculty or staff members who are skilled in data analysis.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a culture that prizes teaching effectiveness.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a system for self-evaluation of teaching practices.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a system for informal peer evaluation of teaching practices (i.e., results are not used to make employment decisions).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a culture in which faculty members make teaching changes based on student-rating and evaluating feedback.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a process for evaluating online teaching.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>formal policies that define the teaching-evaluation process.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>faculty members who have taught online courses previously.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>administrators who have taught online courses previously.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and newer forms such as massive open online courses (MOOCs), in which instructor-student interaction is more distributed and students group together to help each other work through the course materials and challenges in a self-directed fashion. Whether your campus offers traditional online courses, MOOCs, or other online-course models, the interactions that define online teaching remain the same, and this book aims to help you to evaluate those behaviors consistently, fairly, and productively.

Just as it has taken time for higher-education administrators, staff, and faculty members to accept that online teaching can be as effective as face-to-face instruction, the adoption of formal and informal evaluation of online teaching has lagged behind evaluation of face-to-face teaching on many campuses. This book offers concrete steps for readers who want to move their campuses beyond data collection into establishing programs that help us all to accomplish the highest goal of higher education: providing our faculty members with meaningful feedback about their teaching practices and providing learners with the best education we can provide them.

A Thought Exercise

Take a few minutes to assess the status of your own campus regarding the evaluation of online teaching. Figure 1.3 provides a process model for identifying opportunities and pathways toward online-teaching evaluation competence on campus.

You will find a series of thought exercises similar to this one throughout the book to help you apply its research and ideas to your own campus. Refer back to your responses to previous thought exercises as you complete each chapter. Doing so can help you to identify areas of current strength, new questions to pose, and strategies for customizing your efforts for the best reception and results for your institution.