Theory and Practice of 
CLASSROOM 
OBSERVATION

Report researched and written by Sarah MacDonald, 
2016-2017 CFDE Dean’s Teaching Fellow

September 2016

Contents:

1. Classroom observation as a tool for summative and formative assessment
2. Methodological considerations
   a. Who
   b. What
   c. How
3. Follow-up feedback from classroom observations
1. Classroom observation as a tool for summative and formative assessment

Classroom observation is one of the key methods used to study and assess teaching, i.e. to measure and evaluate the quality and effectiveness of teaching. The data gathered from classroom observations are usually supplemented by evidence drawn from other sources, such as student evaluations or performance, review of teaching materials, administrative reviews, the instructor’s own self-assessment, etc. However, classroom observation is the assessment method that provides the most direct data about the instructor’s act of teaching.¹

Particularly when classroom observation is employed in research and/or summative evaluation that informs policy-making or personnel decisions, those doing the observation must necessarily be concerned with the validity and reliability of the data gathered through this method. Much of the research into classroom observation techniques is concerned with how to strengthen these measures, particularly reliability.² Hill et al. assert that using classroom observation effectively for assessment—that is, in a cost-efficient manner, while still producing reliable results—requires developing observational systems that combine quality instruments (i.e. the assessment tools or forms indicating the teaching performance components to be measured), capable raters, and carefully determined scoring designs.³

In their study of reliability criteria for teaching assessment and feedback, van der Lans et al. conclude that achieving the modest reliability needed for fruitful formative feedback requires at least three different classroom observation visits performed by three different observers. Meanwhile, the higher reliability needed for fair summative decisions (e.g. about tenure, salary, dismissal, etc.) demands at least ten different visits by different observers. Still, the researchers also note that one-time classroom observations, while not sufficient to offer an overall picture of the teacher’s general proficiency, can provide reliable and insightful feedback about the specific lesson and classroom activities observed.⁴

---

¹ Ronald Berk, “Survey of 12 Strategies to Measure Teaching Effectiveness,” in International Journal of
⁴ Van der Lans et al., “Once is not enough,” 93, 94.
At the CFDE, the classroom observations we do are less extensive and less formally structured, since they’re not serving as educational research or as the basis for personnel decisions. Rather, our goal is to provide instructors with some formative assessment to help them reflect on and strengthen their teaching performance. In addition, we’re working within certain constraints of time and human resources, given limited CFDE staff to serve all university faculty members who request teaching consultations.

Each classroom observation visit we conduct occurs within the context of a teaching consultation and is typically both preceded and followed by a meeting between the instructor and the one—or sometimes two—CFDE staff members who do the classroom observations. These initial and follow-up conversations provide space for the instructor to share her or his primary concerns about teaching and for the CFDE staff to communicate observations and assessment from the classroom visit. Occasionally a teaching consultation may include more than one classroom observation visit, but that is rare; usually the assessment is based on one visit. Especially in light of the findings of van der Lans et al. cited above, our CFDE teaching assessments should therefore not be construed as overall evaluations of the instructor’s general teaching practices or skill. It seems more accurate and helpful to think of these teaching consultations and the accompanying classroom observations as focused interventions designed to address an instructor’s particular concern(s) and to provide feedback (and suggest resources for further growth) on that aspect of the instructor’s pedagogy.

2. Methodological considerations

Weade and Evertson describe direct observation as “systematic, intentional action”—that is, a conscious, purposeful, and often formalized process for gathering evidence—and as a “multifaceted phenomenon,” involving various interacting components. The components these researchers highlight in their discussion can be encapsulated by their guiding question: “Who will be observing whom, doing what, when, where, under what conditions, for what purposes, with what instruments or tools, and with what intended outcomes?” In the paragraphs that follow, I’ll look more closely at what appear to be some of the most important considerations in determining the methodology one uses to conduct classroom observations. With each point, I’ll also briefly consider how this methodological component manifests in the CFDE process of classroom observation.

a. Who

Classroom observation (as a form of systemic study and assessment) may be conducted by various individuals: peers, administrators, external evaluators and/or researchers. Each of

---

6 Weade and Evertson, 37.
7 It’s fair to say that students also engage in classroom observation—though generally more tacitly than systematically—and their observations may then form part of the basis for their feedback on
these is positioned differently vis-à-vis the instructor and students, and in every case, the positioning of the observer(s) will have some impact on the observation process and its outcomes, including what gets noted or recorded and why.

Peer observations are becoming an increasingly popular and common form of classroom observation. While such observation may be conducted in an “evaluatory” style (whether formative or summative), Vidmar recommends instead what he calls “reflective peer coaching,” which is “a formative process that facilitates introspection and self-awareness prior to, during, and after teaching.” In the model he lays out, the instructor meets with a colleague (“peer coach”) for about 10 minutes prior to a class period, to discuss intended goals, objectives, and activities, as well as what should constitute evidence of successfully achieved learning outcomes. Immediately following the class, the instructor and peer coach meet again for 10 minutes, this time for the instructor to reflect on what actually happened and why, and especially to reflect on any “critical incidents” that may have occurred. Ideally, this process should be repeated several times through regular meetings. Notably, in Vidmar’s model the peer coach generally does not visit the classroom to observe. Rather, observation within the classroom is conducted by the instructor, and the role of the peer coach is to facilitate the instructor’s process of self-reflection in order to prompt deeper awareness, which can then inform future practice. Although this process of reflective peer coaching is not a standard example of classroom observation, it is another way of working with the evidence gained by paying attention to what’s occurring in the classroom. Further, Vidmar’s reasons for proposing this model (rather than peer observation via classroom visits) relate to the ways that classroom dynamics get affected in the presence of a third party (i.e. someone other than the instructor or students).

Teacher evaluations. One might also say that an instructor engages in a form of self-observation, when that instructor is carefully self-aware during teaching and self-reflective afterwards. Berk, in fact, describes student ratings, peer ratings, and self-evaluation as the “triangulation of the three direct observation sources of teaching performance” (“12 Strategies,” 52). However, classroom observation performed by one of the above-named “outsiders” (i.e. someone who doesn’t regularly participate in the class) often has the capacity to be more systemic, disinterested, and wider-angled in vision than the observations done by either students or the instructor herself—and therefore has the potential to fruitfully supplement student or instructor observations by capturing certain details that other observations may miss.

Peer observation of teaching (POT) gets used for a variety of purposes, both for professional development (i.e. learning from modeling by others) and for evaluating or monitoring teaching quality. See Deborah Peel, “Peer observation as a transformatory tool?” in Teaching in Higher Education 10.4 (October 2005): 489-504, for a conceptual and theoretical framework of POT, which includes some literature review and contextual overview of POT in higher education.


Vidmar argues that occasional observations done by a third party do not necessarily reflect the instructor’s usual teaching practice, since the instructor may prepare more carefully and thoroughly for a class that s/he knows will be observed. In addition, the higher-stakes nature of such occasional observation (which can’t entirely avoid the character of summative evaluation, even if it will not be used for personnel decisions) can have a quelling or intimidating effect on the classroom dynamics and/or on the instructor’s own learning process. As Vidmar notes, no matter how constructive the intent of critical feedback, negative comments are likely to be what stands out the most memorably for the instructor receiving the evaluation. See “Reflective peer coaching,” 137.
CFDE staff who conduct classroom observations are distinctively positioned. In a certain way, we might be considered “peer” observers, in that we are fellow teachers (with adjunct faculty status) and not occupying a supervisory role in relation to the instructor observed. Yet we’re also different from peer observers in that we are not members of the instructor’s department. This means that, although we may come with less disciplinary knowledge of course content, as external observers with considerable experience in conducting classroom observations, we’re well positioned to make disinterested evaluations and to notice particular key elements of pedagogical methodology.

While Vidmar’s model of reflective peer coaching describes a process different than what we in the CFDE engage in (and different, I believe, than our mandate, which is to offer more pedagogical expertise and direct mentoring than is presumed in reflective peer coaching), there may be elements of his model that could be helpful for us to consider and perhaps incorporate into our process. I was struck especially that his model involves reflective meetings before and after the class itself, and that he stresses how important it is for these meetings to be guided by the instructor’s own concerns and objectives—all of which somewhat resembles our teaching consultation process. So the fairly detailed suggestions he offers for prompting the instructor’s fruitful self-reflection might have relevance for us. Even more, it may be useful for us to bear in mind his emphasis on the key role the instructor should have in shaping the assessment and the strategies for improvement, as well as his hesitations regarding the challenges of third-party classroom observations and the limitations of what can be gained through them.

b. What

Weade and Evertson stress the necessity of having a defined focal point for the observation process. Although this need not absolutely limit what gets observed, without a guiding focus, the observer will only be able to gather random impressions, which will then be difficult to compare, contrast, or aggregate in any meaningful fashion. The authors also note that the purpose of the observation should determine what gets observed most deliberately and attentively. They further recommend that the focal point of classroom observation not be the teacher’s performance per se, but rather patterns of teaching and learning as revealed through teacher-student interactions, as well as through student engagement in response to the teacher’s actions. In other words, they recommend approaching the observation process with a wider angle of vision (and in the case of videotaping, literally to widen the angle of the camera lens) in order to more effectively capture the breadth of behaviors and interactions occurring in the classroom—thus to better understand, in particular, how student learning is impacted by the teacher’s decisions, speech, and actions.

In the CFDE process, our observations are guided by the items listed on the observation forms we use. The quantitative form includes three general areas for observation—organization and presentation (of lesson material), and interaction (between instructor and students)—with various specific actions or qualities to rate within each of these categories.

The qualitative form does not include such specific items to be ranked, but rather lists several areas for observation—communication; questions/discussion; engaging students; feedback to students; flexibility in lecture, discussion, and activities; respect & rapport; clarity; organization; presentation skills; and use of technologies—with space for notes in each of these areas. Even when the observer does not completely follow or fill out either form, the categories and teaching components listed do channel how attention gets directed in the classroom and what will most likely be noticed and commented upon. Thus, the forms can be viewed as a reflection (whether intentional or implicit) of what we consider the most important or valuable activities and qualities in teaching.

It might be useful for us to reflect further on the construction of these forms (and perhaps on our observation methods more broadly) in light of Weade and Evertson’s recommendations about what the focal point of classroom observation should be. If we believe that fruitful student learning is indeed the desired outcome of teaching—and that evidence of learning reveals that effective teaching is occurring—then how might this shape what we attend to and make notes about during classroom observations? Certainly, learning is a complex process that won’t be simply or solely demonstrated during classroom interactions—yet how might we shape our observation methods (including forms and frames of reference) in ways that highlight not simply teacher performance, but more deeply, patterns of teaching and learning as, in the words of Weade and Evertson, “interdependent and co-occurring processes”

In addition, while the observation forms serve as our general focusing guide to observations, each particular classroom observation we do is also guided by what the instructor has identified as her or his most pressing concerns and questions. This relates to my above point that it’s probably most helpful and accurate for us to perceive our classroom observations as a process of gathering concrete data intended to support focused interventions within the instructor’s professional and pedagogical development (more than as an effort to assess the instructor’s overall teaching proficiency).

c. How

Weade and Everston describe the observer herself as “the first instrument of observation.” In other words, the observer’s frame of reference (like the “defined focal point”) will shape what gets noticed, recorded, and/or assessed during the observation process. Beyond this, most observers also use one or more additional instruments—which may range from a category system or method of taking field notes, to audio- or video-recording the class—and these instruments similarly both “support and constrain what will be observed and what can be learned by observing.”

For classroom observations to lead to fruitful assessment and learning, it is necessary not only to make observations in the moment, but also to have some form of documenting or recording what is observed. This is important to make possible future review and analysis of

12 Ibid.
13 Weade and Evertson, 40.
the data, including by the instructor whose teaching was observed. Even when the purpose of
the observation is to derive formative feedback (as opposed to, say, gathering data for
educational research), some kind of written and/or audio-visual documentation is important
for providing specific details in which to ground the feedback.

There is a wealth of possible observational instruments and recording tools, each with its own
benefits and limitations. Evertson and Green categorize these tools into four general (though
not mutually exclusive) classes:14

1) **Category systems:** These systems involve checklists of preset categories (generally,
specific behaviors of teachers and/or students) to be tallied and/or ranked during the
course of the observation period. These observational tools are the most
predetermined and highly structured, which makes for more streamlined and efficient
recording procedures. This also offers the advantage that the instrument can be used
to make comparisons across classrooms, with higher levels of agreement among
observers and hence of reliability. The main disadvantages of category systems,
however, are their lack of flexibility and responsiveness to context-specific variables,
as well as their limited explanatory power. In particular, they don’t offer much insight
into the relationship(s) between the teacher’s actions and students’ responses.

2) **Descriptive systems:** These systems are designed to observe target behaviors
occurring within their natural context (in this case, the classroom). Like the category
systems, the descriptive systems also may have preset categories to guide or focus the
observation process. However, these systems are more open: instead of simply
tallying or ranking the occurrence of behaviors, the observer will record descriptions of
the behaviors, with attention to context-specific variables and the impact of context on
the behaviors. Using this tool makes for somewhat less streamlined observation and
recording, but it also offers potential for deeper analysis and understanding.

3) **Narrative systems:** Similar to the descriptive systems, these systems are also designed
to generate detailed and context-specific written descriptions of behaviors or other
phenomena occurring the classroom. However, the narrative systems are even more
open, having no preset categories, so they tend to record broad segments of events and
may tell more of the overall story of what is getting observed. A possible disadvantage
of using this observational tool or methodology is that it may be a more labor-
intensive way of gathering data. However, this style of observing and recording may
be especially pertinent when one wishes to understand specific cases in depth and/or
make comparisons across cases.

4) **Technological records:** These systems gather “raw footage” of what’s occurring in the
classroom, typically through video-recording (or sometimes audio-recording). They
provide the most open and unfiltered record of what occurs (although, notably, what
they pick up will depend on the placement of the camera and/or microphone, as well
as on the timing of the recording periods, so they shouldn’t be misconstrued as

---

14 An abridged version of their schema appears in Weade and Evertson, 42-43. For the fuller version
and discussion, see C.M. Evertson and J.L. Green, “Observation as inquiry and method,” in
Handbook of research on teaching, 3rd ed., edited by M.C. Wittrock (New York: Macmillan, 2001): 162-
213.
recording *everything*). These observational tools collect a large amount of concrete data, creating a record that can be reviewed by multiple evaluators. Also, the recordings only capture raw data, without evaluation; evaluation and assessment of the data must necessarily come as a follow-up step, as one watches/listens to the recording afterwards—all of which may be an advantage or disadvantage, depending on the underlying purposes for conducting the classroom observation.

In general, of course, which instrument or system is most appropriate for the observation task at hand will depend on the observer’s targeted focus and larger purposes. One should also note that all these observational tools, as well as Evertson and Green’s classification system, reflect research methods in the social sciences. And since, as I’ve already remarked, our CFDE classroom observations are not intended to serve as educational research, our observation methodology does not exactly conform to any of the above systems (nor does it need to). The instruments we use (see above descriptions of our observation forms) are probably best understood as a kind of “descriptive system.” Certainly our qualitative form is in this style. And while our quantitative form does resemble a category system, even when this form gets used, the rankings alone don’t constitute all the notes we take, nor are the rankings themselves necessarily the main focus of the feedback we share with the instructor. (And we’re certainly not using these forms to do any kind of comparisons across classrooms or teachers.)

Overall, the **goal for CFDE classroom observations is to generate some knowledge, based in concrete examples, of an individual instructor’s pedagogical methods and then to use this knowledge to inform the teaching consultation.** So our concern is less with the data itself recorded on our forms or in our notes, and more with the feedback we’re able to give the instructor. For this reason, while a narrative system of observation would likely generate more extensive data, I don’t believe this is necessary for our purposes, and the additional labor involved would likely be a hindrance to us, given our needs for efficiency and more immediate analysis. Similarly, while we do occasionally videotape a class if the instructor requests that, this is typically for the purpose of creating a record that the instructor can then view and assess, rather than for the purposes of our own assessment work.

### 3. Follow-up feedback from classroom observations

Finally, given that the primary purpose of CFDE classroom observations is to provide instructors with formative feedback, I close this report with a brief consideration of the kinds of feedback that are most helpful in supporting performance improvement.

Shute has conducted one of the more recent and extensive literature reviews of the vast body of educational research studies published on feedback and its relation to learning.\(^{15}\) She

---

\(^{15}\) Valerie J. Shute, “Focus on Formative Feedback,” in *Review of Educational Research* 78.1 (March 2008): 153-189. Shute’s review focuses on students as the primary recipients of formative feedback (and most of the research on feedback appears to share this focus). However, Shute’s conclusions about what makes feedback most effectively formative (or not) appear equally relevant to situations in...
defines formative feedback as “information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning.” Strikingly, she notes that while feedback is widely considered a crucial and powerful facilitator of learning, several studies have found that it can also have a negative effect or no effect at all on learning—and so, for feedback to be genuinely formative and a positive influence on learning, it should be valid, objective, focused, clear, and timely. Shute concludes her piece with several feedback guidelines aimed toward enhancing learning—things to do, things to avoid, and how to determine the kind of feedback (immediate or delayed, directive or facilitative, scaffolding or verification) that will be most appropriate to the learner, the task, and the desired outcomes.

While much of the research into formative feedback in educational contexts has studied students as the recipients, there are a few texts that attend to teachers as the recipients of feedback. Especially relevant to this report is Khachatryan’s study of qualitative performance feedback based on observations of teaching. Khachatryan grounds her study in Kluger and DeNisi’s Feedback Intervention Theory (FIT), which categorizes feedback interventions based on how they direct the recipient’s attention: whether to the self (meta-task processes), to the accomplishment of the task (task motivation), or to details of the task (task learning). For purposes of her study, Khachatryan identifies these three levels as: “self-feedback,” or commentary about the teacher’s character and skills; “product-feedback,” or assessment of how well the teacher did or did not accomplish the task; and “process-feedback,” which describes the particular steps taken toward achieving the task.

Research shows that “self-feedback” tends to impede learning and change—positive comments, by inducing complacency and decreasing motivation to grow; and negative comments, by evoking defensiveness, discouragement, or disengagement. Khachatryan which classroom teachers are the “learners,” i.e. receiving formative feedback for the sake of improving their teaching performance. Shute’s piece is an excellent resource. In addition to her overall literature review, she looks more in-depth at four influential studies that offer preliminary theories or conceptual models of formative feedback, and she concludes her piece with several specific guidelines.

16 Shute, 154.
17 Shute, 177-181.
18 Edit Khachatryan, “Feedback on Teaching From Observations of Teaching: What Do Administrators Say and What Do Teachers Think About It?” in NASSP Bulletin 99.2 (2015): 164-188. Khachatryan’s study is of secondary school teachers (observed by their administrator); however, her conclusions about how to make formative feedback most useful and likely to promote positive change seem applicable also to teaching observation and assessment in higher education.
19 Feedback Intervention Theory is used in the field of psychology and related disciplines and was developed from Kluger and DeNisi’s 1996 meta-analysis of the effects of feedback interventions. In this report, I rely on summaries of FIT from Khachatryan and Shute. For further details or deeper understanding of this model, see A. N. Kluger and A. DeNisi, “The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory,” in Psychological Bulletin 119 (1996): 254-284; and A. N. Kluger and A. S. DeNisi, “Feedback interventions: Towards the understanding of a double-edged sword,” in Current Directions in Psychological Science 7.3 (1998): 67-72.
20 Khachatryan, 170.
further explains that “additional cognitive resources are necessary to make the jump from comments about the ego, something innate to the person, to the task particulars or teaching moves.”

In contrast, “product-feedback” is likely to increase the teacher’s motivation, and “process-feedback” (according to FIT, the most effective of all) helps the teacher learn about her teaching moves, which she can then replicate or adjust as needed. In her study, Khachatryan found that the feedback which most elicited teachers’ plans to try to improve were either process comments or those which combined process and product messages. In other words, “feedback that focuses attention on the details of instructional moves… prompts teachers’ learning processes and plans for changes in teaching. Breaking down the steps in instructional moves and communicating them to teachers would make clear which components of their practice may need attending, refining, and improvement.”

Khachatryan’s conclusions also note the importance of engaging teachers in reflection around the feedback, as an important part of encouraging learning and positive change.

The content of feedback—what kind of information is conveyed, and where it directs the recipient’s attention—is obviously an important factor in its efficacy. Also significant is the feedback process, or how the information is conveyed to the recipient. In a literature review of theoretical and empirical discussions of feedback—supplemented with insights gained from her own experiences as an instructional consultant—Brinko offers several suggestions for how to frame and communicate feedback most effectively. A notable theme in Brinko’s recommendations is that the most effective feedback process is client-centered:

Rather than assuming the role of expert or problem solver, the consultant acts as facilitator, helping the client identify problem areas, set priorities, set goals, brainstorm for alternative behaviors and strategies, and so forth. This collaborator/facilitator role ensures that all authority, as well as responsibility, lies with the client rather than the consultant.

This point seems in line with her research-based observation that self-generated feedback is, generally, “more valued, better recalled, and more credible than feedback from other sources”—although individuals with lower self-esteem and/or an external locus of control...

---

21 Khachatryan, 177.
22 Khachatryan, 183.
23 Self-reflection as an essential—perhaps the essential—component of growth and transformation in one’s practice is a prevalent note also in Vidmar, “Reflective Peer Coaching,” and in Peel, “Peer observation as a transformatory tool?”
24 Kathleen T. Brinko, “The Practice of Giving Feedback to Improve Teaching: What Is Effective?” in The Journal of Higher Education 64.5 (September/October 1993): 574-593. While Brinko’s 1993 study is hardly recent, it still feels pertinent, as well as unusual for being focused on the context of higher education. I found relatively few studies or discussions that are focused on teachers (rather than students) as the recipients of feedback in educational contexts, and fewer still that attend to the particular context of higher education.
25 Brinko, 578. Again, Brinko’s recommendation here seems very in line with other pieces cited above that also emphasize the importance of the instructor’s own self-reflection and engagement in the assessment process.
may benefit more from feedback derived from others. Likewise, she emphasizes that feedback is more effective when it is part of a process the instructor has voluntarily engaged in, when that process allows for the instructor to respond to and interact with the observer/giver of feedback, when the instructor has some choice in how observations are conducted and feedback is delivered, and when that feedback relates to problems or goals the instructor has identified.

In addition, I was struck by Brinko’s assertion that creating “a moderate amount of cognitive dissonance” makes for more effective feedback. In other words, if the feedback reveals discrepancies between the instructor’s ideals and self-perceptions and the evidence of phenomena observed in the classroom, then this creates a psychological climate that encourages change. Other points from Brinko that are significant—and by now familiar, given studies cited above—include her descriptions of the most effective feedback as: focused and specific, grounded in accurate and concrete data, and directing attention toward behavior, rather than personality traits.

In conclusion, our teaching consultation process in the CFDE seems to already reflect many of the conclusions that Shute, Khachatryan, and Brinko highlight. In particular, our process is “client centered” in that each consultation begins with an inquiry into the instructor’s presenting concerns and/or goals. Beyond this initial inquiry, it’s important for us to stay attentive to how we can keep the process fruitfully collaborative—so that we function not simply as “experts” or “problem solvers” but as facilitators of the instructor’s own discoveries and growth. Further, it should serve us well—both as we conduct classroom observations and as we convey follow-up feedback—to keep attention (our own and the instructor’s) directed toward concrete and specific steps of the teaching and learning process, both instructional moves and student responses.

---

26 Brinko, 577, 583. See also pp 587-588 for her recommendations on how to adapt the feedback process depending on whether the instructor has an internal or external locus of control and high or low self-esteem.
27 Brinko, 581, 582, 584.
28 Brinko, 580, 584. Note that the dissonance should be moderate, rather than very large or small, in order to most effectively foster change.
29 Brinko, 579-580.