Formative Assessment Through Protocol-Based Observations

Andrew Martin, Faculty Advisor, ASSETT, and Professor of Ecology and Evolutionary Biology
Shane Schwikert, Educational Technology Researcher, ASSETT
Mark Werner, Associate Director of Academic Technology Strategy and Support, OIT
Noah Finkelstein, co-Director of the Center for STEM Learning and Professor of Physics

October 30, 2017

Introduction

Peer review is a pillar of assessing teaching quality and effectiveness. An important aspect of peer review is peer classroom observation, which can assist our faculty in generating relevant and authentic results for formative assessment, and for data-driven course and teaching revision.

Nationally renowned organizations and communities are increasingly calling for a renewed focus on teaching quality and instructional strategies that improve the learning and success of all students. For example, the Association of American Universities (AAU) has launched a major initiative to improve undergraduate STEM education. This initiative calls for faculty members to adopt evidence-backed best teaching practices, and to better evaluate and recognize faculty who demonstrate efforts to improve their teaching.

Adopting authentic and valid methods of peer review of teaching are important steps toward improving formative assessments of teaching. They can be extremely valuable for aligning faculty personnel evaluations with best practices in education. The University should institutionalize best practices for peer assessment of teaching and connect them to professional development so that it is clear the evaluation is not the data *per se*, but rather what happens with the data as a consequence of the review.

The Arts & Sciences Support of Education through Technology (ASSETT) team has developed a service called the Visualization of Instructional Practices (VIP) that uses a versatile platform for gaining authentic and valid data useful for formative assessment of teaching and that can be the basis of evaluating teaching excellence during personnel evaluations.

The purpose of this paper is to 1) emphasize the value using valid and authentic methods of peer review of teaching, 2) encourage data-driven professional development and revision of teaching, and 3) encourage adopting technology that enables collection of teaching and student engagement data in the classroom.
We emphasize that the data are not be used in a comparative manner for personnel evaluation; but instead the data should form the basis of productive data-driven revision of teaching in a

manner that promotes adoption of research-supported best practices. Furthermore, it is important that all data are owned by the faculty member who was observed. When the faculty member owns the data, they can choose to use it as they see fit, and they are more likely to engage in productive formative assessment. Conversely, if the data are not owned by faculty, they are likely to question whether they should allow the review to happen at all, and if pressured to allow it, they are likely to treat it defensively, rather than as a basis for formative guidance. We envision that faculty and individual units would determine their own uses of these key data, such that it becomes a normative cultural practice.

Towards improving teaching quality: Classroom observations

A major component of the AAU's ongoing initiative is a call for the development and use of more effective ways to evaluate teaching in the faculty rewards structure[1]. An important part of evaluation relies on peer review, and what an educator does with peer formative assessment data. Additionally, one of the most relevant steps that a faculty member can take to advance their teaching is to solicit qualitative and quantitative feedback from a peer. It is important to recognize that while peer assessments can be subjective or biased (e.g., via professional reciprocity), they can be done in a way so as to minimize bias by using one or more of a variety of published tools available for valid, authentic, and data-rich peer observation. These tools are referred to as classroom observation protocols. Each protocol contains a set of codes that typically correspond to classroom activities or behaviors. An observer who is well-trained on the chosen protocol will attend a class session (or watch a video of class), and record various behaviors and activities at set time intervals throughout the duration of the class. Ideally, the peer reviewer collects the data, or is present in the classroom when some or all of the data are collected. The collected information can then form the basis of a data-rich conversation about teaching.

Observation protocols are designed to measure particular aspects of teaching, and the variety of existing protocols provides instructors with one or more tools that can best capture the scope of what happens in the classroom. Some protocols focus on whether instructors use scientific teaching practices, track student and teacher activities during class, assess whether instructors use research-supported best practices, or track student engagement over time (**Table 1**).

Table 1. Abbreviated list of research-validated observation protocols and their primary focus and disciplinary relevance.

Example protocol	Focus	Primary discipline/focus	Citation
Classroom Observation	Tracks student & instructor behaviors and	STEM / large lectures	Smith et al. (2013)

[1] See: hyperlink

Example protocol	Focus	Primary discipline/focus	Citation
Protocol for Undergraduate STEM (COPUS)	activities across a class period		
Methods Instrument for Scientific Teaching (MIST)	Use of scientific teaching practices	Scientific teaching	Couch et al. (2015); Handelsman et al. (2007)
Teaching Dimensions Observation Protocol (TDOP)	Use of research supported best practices in teaching	All disciplines	Hora et al. (2013); Hora (2015)
Behavioral Engagement Related to Instruction (BERI)	Tracks student behavioral engagement across a class period	Large lectures	Lane & Harris (2015)

Support Through ASSETT

Despite the myriad of benefits that classroom observations provide, there are also challenges (e.g., determining the most appropriate protocol to use, identifying and training observers, and analyzing and interpreting the data). To address this shortcoming, members from the Center for STEM Learning (CSL) and the Academic Technology Design Team (ATDT) unit of the Office of Information Technology (OIT) initiated collaborations with national leaders from the Bay View Alliance, University of Wisconsin, and the Science Education Initiative. Over the following two years protocols for observation, faculty engagement and reporting were developed (see OIT's description, the Tools for Evidence-based Action (TEA) site, the 2016 COLTT site, and the AAU Scaffolding site.

In 2016, as ASSETT became a part of OIT, this observational protocol effort was transferred into ASSETT's Visualizing Instructional Practices (VIP) service. Behind a team of well-trained students, researchers, and pedagogy experts, ASSETT strives to support faculty who are interested in using classroom observations, while removing any challenges and barriers to the process. Faculty who participate receive a customized report of their data (see Appendices A, B, C, and D or examples), and may choose to consult with a staff member to help interpret their data.

Value for Personnel Evaluations

As we mentioned earlier, peer review of teaching is one of the pillars of teaching evaluation. Validated observation protocols enable collection of relevant data that can be used for formative assessment and revision of teaching, and demonstrate a faculty member's commitment to professional development. The fact that a faculty member would elect to use a valid observation protocol for peer review, and use the data in ways that advance teaching effectiveness is an example of teaching excellence. We do not advocate for a particular observation protocol or for using the data for comparative purposes during personnel evaluation associated with promotion and tenure. Nonetheless, the use of observation protocols—such as COPUS, MIST, BERI or more tailored approaches—demonstrates a commitment of faculty to gaining valuable feedback that enables data-driven revision of teaching practices and should, in our opinion, become a normal part of evaluations of teaching.

Conclusion

As the University revises its practices and guidelines to take a more scholarly and evidence-based approach in the evaluation of teaching, the VIP service can serve as a key tool. VIP provides rich evaluation of teaching performance, allows instructors to make adjustments to teaching based on those data, and to seek advice from teaching experts on how to interpret formative data on teaching effectiveness. Observational protocols have been developed, studied, and widely adopted among AAU peers. As the university makes the use of this best-practice in teaching evaluation more widespread, it should do so in a sustainable, scalable manner by providing appropriate resources, incentives and recognition.

Acknowledgements

The authors wish to thank Sarah Wise, OIT and CSL, for her work on the VIP service and her comments on this draft. We also wish to acknowledge the work of Viktoriya Oliynyk, Digital Learning Initiatives, and Elias Euler, Center for STEM Learning, for their important work on the OPLE protocol.

References

Couch, B. A., Brown, T. L., Schelpat, T. J., Graham, M. J., & Knight, J. K. (2015). Scientific teaching: defining a taxonomy of observable practices. *CBE-Life Sciences Education*, *14*(1), ar9.

Fabrice, H. (2010). *Learning our lesson review of quality teaching in higher education: Review of quality teaching in higher education*(Vol. 2010, No. 2). OECD Publishing.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415.

Handelsman, J., Miller, S., Pfund, C. 2007. *Scientific Teaching*. W. H. Freeman and Company, New York, NY. 184 pp.
Hora, M. T. (2015). Toward a descriptive science of teaching: How the TDOP illuminates the multidimensional nature of active learning in postsecondary classrooms. *Science Education*, *99*(5), 783-818.
Hora MT, Ferrare JJ (2013). Instructional systems of practice: A multidimensional analysis of math and science undergraduate course planning and classroom teaching. Journal of the Learning Sciences 22: 212-257.
Hora MT, Oleson A, Ferrare JJ (2013). Teaching Dimensions Observation Protocol (TDOP): User's manual. hyperlink.
Lane, E. S., & Harris, S. E. (2015). A new tool for measuring student behavioral engagement in large university classes. Journal of College Science Teaching, 44(6), 83-91.
Smith MK, Jones FHM, Gilbert S, Wieman C (2013). The classroom observation protocol for undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. CBE Life Sci Educ 12: 618-627.

Appendix A: Example COPUS data

Figures 1 shows two example visuals that faculty receive in their report when they request an observation using the Classroom Observation Protocol for Undergraduate STEM (COPUS; Smith et al., 2013). The COPUS protocol is ideal for large lecture STEM classes, and tracks numerous behaviors that have been identified as good teaching practices.

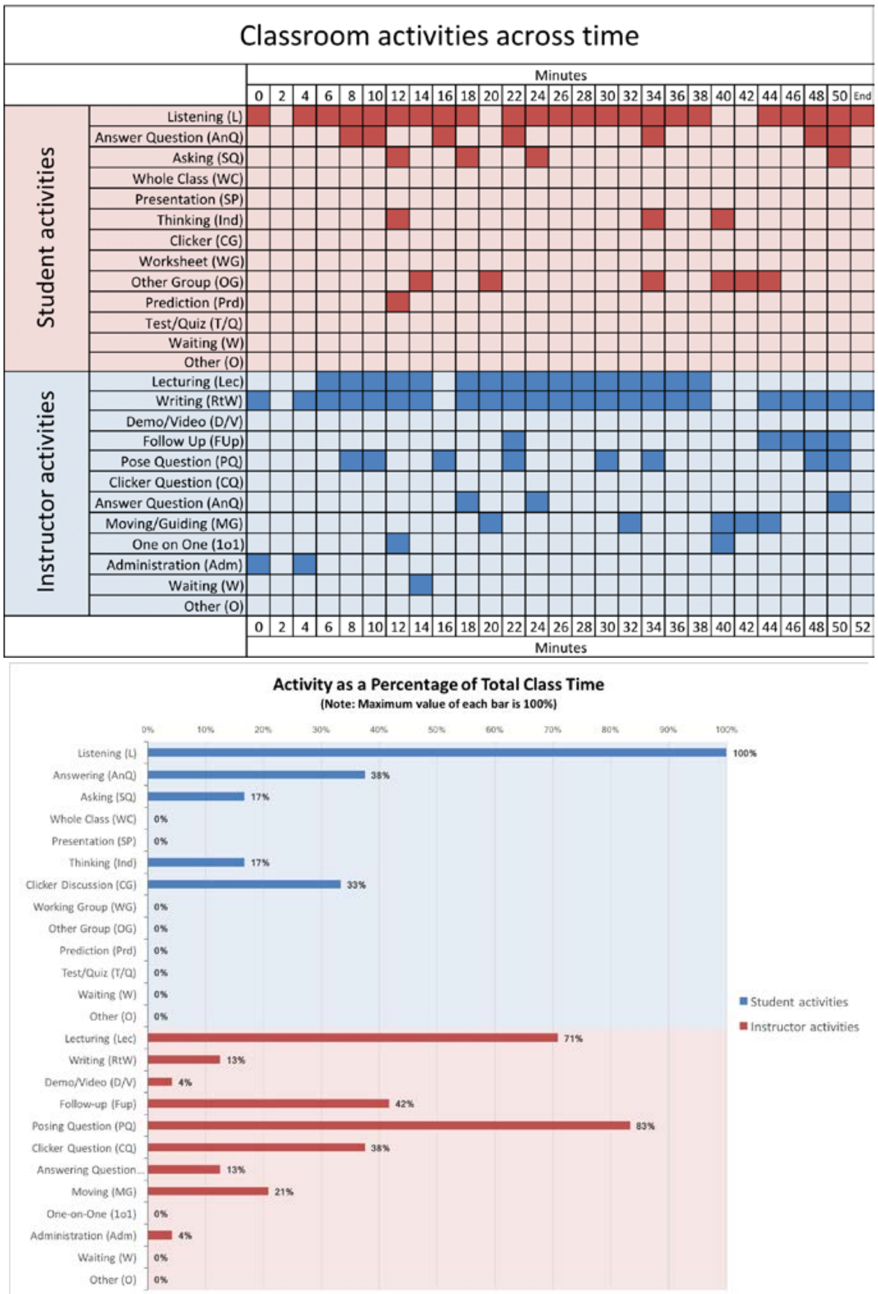

Figure 1. *Upper graph.* Sample COPUS protocol visual showing various instructor and student activities occurring across the course of a single class period. Shaded boxes represent when activities occurred in time. *Lower graph*. Sample COPUS protocol visual showing various instructor and student activities that occurred as a percentage of total class time.

Appendix B: Example BERI data

Figure 2 shows an example visual that faculty receive when they request an observation using the Behavioral Engagement Related to Instruction (BERI; Lane & Harris, 2015) protocol. The BERI protocol can be used in any class or discipline, and tracks student engagement across a class period.

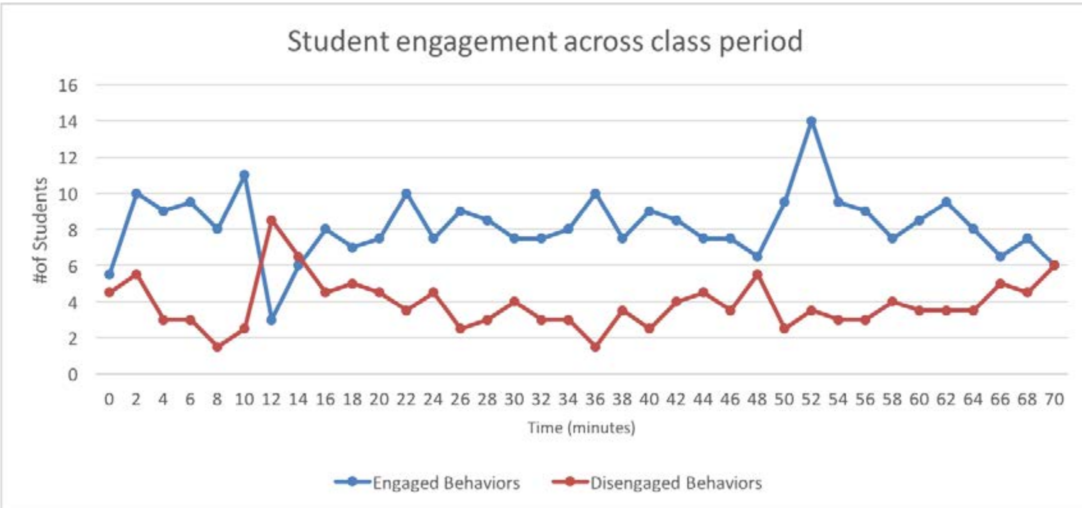

Figure 2. Sample BERI protocol data showing counts of students exhibiting either engaged or disengaged behaviors across a single class period.

Appendix C: Example of MIST data

Figure 3 shows an example of data generated by implementing the MIST protocol (Couch et al., 2015). This observation protocol focuses on recording scientific teaching practices (see Handelsman et al., 2007). Individual instructors or professors are provided observation scores across a variety of categories of teaching (not shown) and a composite rank score based on a number of instructors within a particular discipline.

Table 2. Example of a portion of the results from the MIST analysis (done by the University of Nebraska-Lincoln) on one professor at CU Boulder. Percentile indicates the rank of the focal course relative to all courses in the MIST analysis database.

MIST scale type	Score	Percentile
MIST composite	67.60	98th
MIST subcategories:		
Active Learning Strategies	0.636	94th
Learning Goal Use & Implementation	0.786	94th
Inclusivity	0.823	95th
Instructor Awareness	0.808	90th
Experimental Design	0.463	83rd
Data Analysis and Interpretation	0.748	99th
Cognitive Skills	0.677	95th
Reflection	0.616	97th

Appendix D: Example of a tailored observation protocol

Figure 4 shows an example of data generated by implementing an observation protocol developed by the instructor and the observer. In this case, the instructor was interested in what was happening in class (the lower 5 categories below the horizontal dashed line) and the amount of time dedicated to particular types of learning goals (the upper 3 categories above the horizontal dashed line). The ability to tailor observation protocols for the particular goals of an instructor is important because what is most important is that instructor seek data that can serve as a basis for the formative assessment and data-driven revision of teaching.

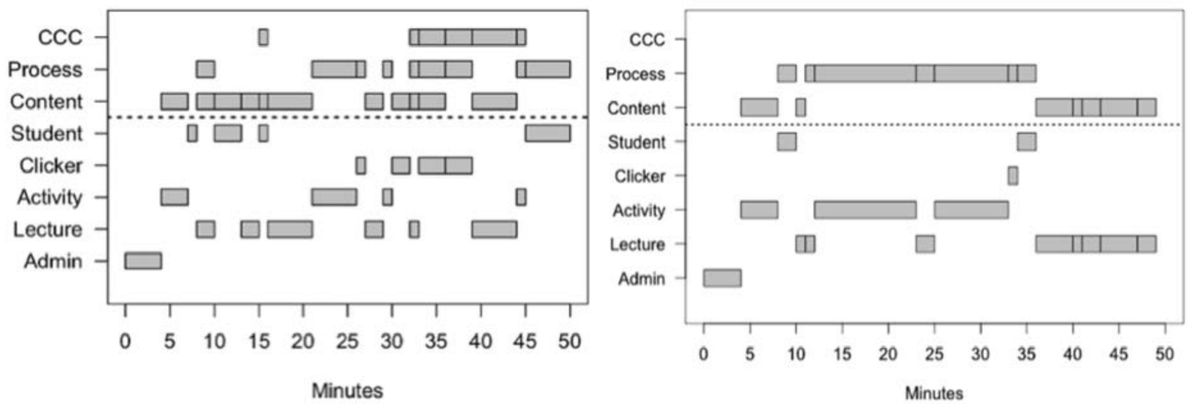

Figure 4. Visualization of an example custom observation protocol in which the professor was interested in two things: (1) when and how often she emphasized particular types of learning goals (CCC = cross-cutting concepts, Process = science process skills, and Content = science content goals), and (2) when she engaged in particular types of teaching strategies (Student = student-centered approaches; Clicker = use of peer instruction using clickers; Activity = students working on an activity in a group; Lecture = professor lectures, and Admin = professor "takes care of business"). The two figures shows the results from two separate observations of the same instructor.