A New Tool for Measuring Student Behavioral Engagement in Large University Classes

By Erin S. Lane and Sara E. Harris

We developed a classroom observation protocol for quantitatively measuring student engagement in large university classes. The Behavioral Engagement Related to Instruction (BERI) protocol can be used to provide timely feedback to instructors as to how they can improve student engagement in their classrooms. We tested BERI on seven courses with different instructors and pedagogy. BERI achieved excellent interrater agreement (>95%) with a one-hour training session with new observers. It also showed consistent patterns of variation in engagement with instructor actions and classroom activity. Most notably, it showed that there was substantially higher engagement among the same group of students when interactive teaching methods were used compared with more traditional didactic methods. The same general variations in student engagement with instructional methods were present in all parts of the room and for different instructors.

he main goal of this work was to develop a practical and objective observation protocol that could be used in large university classrooms to produce quantitative data identifying student behavioral engagement level at any given time. Engagement levels could then be used to determine the teaching practices that most effectively engage students and provide timely feedback to instructors. In achieving this goal, we wanted our observation protocol to use a simple coding system that would be quick and easy for observers to learn and that could be easily presented to the course instructor as timely feedback immediately after class. Here we present the Behavioral Engagement Related to Instruction (BERI) protocol, an observation tool that has achieved good interrater agreement with little training, that can provide timely and formative feedback to the course instructor in the form of a graphical representation of student engagement over time for a class, and has results that can be easily related to the pedagogy and curricular content of the lecture.

The theoretical basis for this work stems from constructivism, which describes learning as an active process through which individuals construct understanding of new information by linking it to previous experiences and knowledge (Bransford, Brown,

& Cocking, 2000). Students learn best when they are actively engaged and can therefore deeply encode material (Baddeley, Lewis, Eldridge, & Thomson, 1984; Hake, 1998). In large-lecture settings, however, students are more likely to experience anonymity and distraction leading to decreased engagement in class (Blatchford, Edmonds, & Marin, 2003; Fenollar, Roman, & Cuesta, 2007). Similarly, students may find it difficult to concentrate over an extended period of time in traditional lectures when they are in passive learning mode (Young, Robinson, & Alberts, 2009). When students stop paying attention and disengage with course content, they are unable to construct their own knowledge (Young et al., 2009). Therefore, it is important for instructors in large classes to examine what they might do to more deeply engage students to promote learning. But how can instructors measure the level of student engagement and effectiveness of their interventions to improve it?

Currently there are several tools to measure engagement in educational settings. A variety of self-report questionnaires have been used in research on student engagement (Chapman, 2003); however, the validity of the data yielded by these measures has been questioned because students' ability to accurately assess their own behaviors is highly variable (Assor &

Connell, 1992). Several studies have used checklists and rating scales for instructors to measure engagement in their own courses (Chapman, 2003). These measures are flawed by poor reliability and bias toward a positive result of the instructor's effectiveness.

Classroom observations have been widely used in education over the past couple of decades to measure effective teaching practices (Waxman & Padron, 2004). However, from Waxman and Padron's (2004) review of classroom observation studies, it is evident that most instruments focus on the instructor's behaviors as the unit of measurement rather than on the student's behaviors. Chapman (2003) specifically reviewed a number of third-party classroom observation instruments to measure student engagement. The majority of these protocols assess whether students are engaged during a specific moment or over a given time interval, and various behaviors are considered "engaged"—such as reading (aloud or silently), paying attention, or looking for materials (e.g., Greenwood & Delquadri, 1988).

However, all of the protocols reviewed by Chapman (2003) were designed for use in K-12 classes with 30-35 students. In these methods, a small representative sample of students are selected (typically five) and observations are rotated across students so that each student is observed continuously for about one minute at a time before moving on to the next student. Thus, in a typical observation session of 30 minutes, approximately 5 minutes of observation data can be collected on each student, with six to seven sessions being required to observe a full class of 30–35 students. Although these methods may be effective at identifying the engagement levels of individual students, because each student is observed individually it would be difficult to relate teaching practices to levels of engagement in the classroom. In addition, these methods are time-consuming and not practical for use in large university classes.

There are a handful of published classroom observation instruments that have been developed for postsecondary education settings with large classes: the Reformed Teaching Observation Protocol (RTOP; Piburn et al., 2000), VaNTH Observation System (VOS; Harris & Cox, 2003), STROBE (O'Malley et al., 2003), and the Classroom Observation Protocol for Undergraduate STEM (COPUS; Smith, Jones, Gilbert, & Wieman, 2013). RTOP does not explicitly deal with student engagement and is more focused on the teachers' implementation of the lesson plan. VOS and STROBE measure student engagement levels as a part of a larger series of items to characterize the classroom climate and only capture coarse measurements of student engagement by a global judgment of the proportion of the class as a whole that is on task. In addition, RTOP, VOS, and STROBE rely on complicated coding systems, scorings, and questionnaires to contextualize what is happening in the lecture at a given time. These protocols therefore require lengthy user manuals and training sessions, and there is a danger that the instructor will find feedback difficult to use because of the complexity of the coding system. COPUS includes documenting what students are doing in class but does not explicitly measure levels of engagement.

Various definitions of student engagement have appeared in the education literature. The term *student engagement* is used to describe a variety

of different things: from time-on-task studies to studies that investigate the quality of effort, interest, and willingness to participate in learning activities (Kuh, 2009). Student engagement levels have also been indexed by cognitive, affective, or behavioral criteria (Chapman, 2003). Cognitive criteria assess the extent to which students are expending mental effort in the learning tasks. Affective criteria assess the level of students' investment in, and emotional reactions to, the learning tasks. Behavioral criteria assess the extent to which students are actively participating in relevant learning tasks presented.

Here we present an observation protocol that measures university student behavioral engagement, which we define simply as on-task behavior in the classroom. The BERI protocol was developed after we noticed differences in student behavioral patterns in the classroom that seemed to be related to different instructional methods (i.e., didactic parts of a lesson versus more interactive parts). This research examines the following questions:

- 1. Can we quantify differences in student behavioral engagement with a simple, valid, and reliable observation protocol?
- 2. How does student behavioral engagement vary with classroom activities, instructional methods, and between instructors?

Methodology *Developing the protocol*

The first stage of developing BERI was to conduct unstructured observations of several large classes to note patterns of student behavior that would define engagement. First, all student in-class behaviors were recorded, described, and classified

in Table 1 (engaged behaviors) or Table 2 (disengaged behaviors). Although observation of one of these behaviors does not guarantee that the student is engaged/disengaged, we have limited the set to pronounced behaviors for which there is widespread agreement from observers as to what it represents. Also, although

BERI is able to identify on-task behaviors, it is unable to differentiate the depth at which the student is cognitively interacting with the material. For example, frantic writing may be a sign of active student paraphrasing or passive note copying.

The next stage was to determine the optimal number of students to simultaneously observe. We wanted the observations to be as unobtrusive as possible and decided that the observer would sit in one place during the entirety of the class. We also did not want the students to notice or be distracted by the observer, so we decided the observer would sit behind the group of students being observed.

TABLE 1

Descriptions of student in-class behaviors that indicate they are engaged.

	Engaged
Listening	Student is listening to lecture. Eye contact is focused on the instructor or activity and the student makes appropriate facial expressions, gestures, and posture shifts (i.e., smiling, nodding in agreement, leaning forward).
Writing	Student is taking notes on in-class material, the timing of which relates to the instructor's presentation or statements.
Reading	Student is reading material related to class. Eye contact is focused on and following the material presented in lecture or preprinted notes. When a question is posed in class, the student flips through their notes or textbook.
Engaged computer use	Student is following along with lecture on computer or taking class notes in a word processor or on the presentation. Screen content matches lecture content.
Engaged student interaction	Student discussion relates to class material. Student verbal and nonverbal behavior indicates he or she is listening or explaining lecture content. Student is using hand gestures or pointing at notes or screen.
Engaged interaction with instructor	Student is asking or answering a question or participating in an in-class discussion.

TABLE 2

Descriptions of student in-class behaviors that indicate they are disengaged.

	Disengaged
Settling in/ packing up	Student is unpacking, downloading class material, organizing notes, finding a seat, or packing up and leaving classroom.
Unresponsive	Student is not responsive to lecture. Eyes are closed or not focused on instructor or lecture material. Student is slouched or sleeping, and student's facial expressions are unresponsive to instructor's cues.
Off-task	Student is working on homework or studying for another course, playing with phone, listening to music, or reading non-class-related material.
Disengaged computer use	Student is surfing web, playing game, chatting online, checking e-mail.
Disengaged student interaction	Student discussion does not relate to class material.
Distracted by another student	Student is observing other student(s) and is distracted by an off-task conversation or by another student's computer or phone.

The observer still needed to have an unobstructed view of the students (could see their faces, computer screens, and notes) and was in close enough proximity to get the general context of student interactions. After several trials we chose 10 students as the optional number, with the observer sitting in the row directly behind the students being observed and at an angle so that the students were still within the observer's sight line. It is worth noting that all of our observations were conducted in large-lecture theatres with tiered seating, which allowed the observer to sit higher than the students being observed and thereby get a better view. For alternative classroom settings, the number of students or observers' position may have to be adjusted so that all students in the observation group can be adequately seen. Using the behaviors in Table 1 and 2 as a guide, we recorded how many of the 10 students were engaged ("E"), disengaged ("D"), or were of uncertain ("U") engagement. If the observer's view of a student was temporarily blocked or the student's behavior was unclear, he or she was entered into the uncertain category. It took approximately 3 to 10 seconds to gauge the level of engagement of each student, with a 10-student cycle taking approximately one minute to complete. We did not record the specific behavior of each individual; rather, for each 10-student cycle we recorded one observation point (e.g., "8/10 students engaged") that was time stamped at the start of the cycle.

Using the protocol

Prior to conducting a classroom observation, the observer receives a printed copy of the instructor's notes or lesson plan. The observer randomly chooses a place to sit during the class where 10 students can be clear-

FIGURE 1

Coversheet for the Behavioral Engagement Related to Instruction (BERI) protocol observations.

Observation Protocol Coversheet		
Date of Observation:		
Course Name, Number and Section:		
Instructor(s):		
Observer's Name:		
Classroom Number:		
Estimate of class attendance:		
Position in Class: (drawing a diagram of the class may be useful)		
Notes on classroom environment: (i.e., description of space and seating arrangement, abnormal temperature, use of technology).		
Brief description of instructional method: (i.e., traditional lecture mixed with clicker questions).		
Notes about group of students being observed: (i.e., 5/10 are using a computer).		

ly seen. At the beginning of the class, the observer fills out a coversheet that contains general information and notes about the class (Figure 1). Once the class starts, observation points are recorded directly onto the copy of the instructor's notes in the section corresponding to what is being covered. This ensures that the instructor will later be able to relate engagement with what was happening in their class at any specific time. An observation point is taken for every page of notes, for any major change in activity or content, or at 2-minute intervals depending on which time interval is shorter. Changes in the classroom activity (e.g., clicker question, in-class discussion, demonstration) or instructor behaviors (e.g., moving around the classroom,

using humor or real-world examples) are recorded under each observation point. Because classroom activities and instructor behaviors vary in different courses and with different instructors, this section is open-ended and determined by the observer. Instructor questions to the class and student questions to the instructor are also documented with the following information: the section of the room in which the question/answer originated and how the interaction is followed up (e.g., entire class, subgroup of students, one student). In most cases in this study, the classroom activity was clearly outlined in the class notes so that the observer needed only to record activities, behaviors, and questions that were not preplanned.

Training and interrater reliability

The next stage was to test BERI to ensure that we could train observers and get consistent results. Ten different observers-instructors, graduate student teaching assistants, and science education researchers-were trained to use BERI. Training consisted of a brief description of the methodology and purpose of the protocol followed by a supervised observation session during a routine 50-minute undergraduate class. The trained observer and trainee sat side by side and agreed to observe 10 students that they both felt confident that they could adequately see. To ensure that these observations were completed at exactly the same time, the trained observer indicated with unobtrusive hand signals when to start the observation cycle. Both observers independently cycled through each of the 10 students in sequence and for each student recorded "E" for engaged, "D" for disengaged, and "U" for uncertain. At the end of each cycle and after class, the trainer and trainee compared results and discussed any differences. The observer and the trainee then attended a second class together where they still sat together, agreed on the students they were going to observe, and used hand gestures to cue the start of an observation cycle; however, this time they did not share or discuss their results. The data from these sessions were then used to calculate the interrater reliability of observers, calculated as the percentage of identically marked Es, Ds and Us for individual students for each pair of observers for all sessions observers attended together.

After only one practice session, new observers were able to produce reliable results compared with their trained partner. Data from 2,154 judgments of individual student engagement from six pairs of observers

in three different educational settings with five different instructors were used to evaluate interrater reliability. The average interrater agreement was calculated to be 96.5%. Trained observers only put students in the uncertain category 3.3% of the time, and this typically happened when the observer's view of the student was obstructed. Observers in training placed students into the uncertain category 7.3% of the time. On the basis of this high percentage agreement between observers, it appears that one 50-minute practice class is adequate training for new observers to produce reliable results. This confirms that the behaviors in Tables 1 and 2 are obvious when observed. Debriefing after an observation cycle during the training sessions revealed sources of discrepancy between observers; in many of the cases, an observed student was displaying a subtle behavior. For example, a student may have appeared to be listening to the instructor, whereas in fact the student was sending a text message on his or her cell phone so discretely one of the observers did not notice. Another small source of discrepancy occurred when an observer was not familiar with the course material and was not able to determine if students were writing or reading notes for the class in progress or doing work for a different course. This was more common shortly before exam time and was usually recognized after a couple of observation cycles when the observer realized that the student was not changing his or her behavior in response to what was happening in class. Although we believe that anyone can be trained to use the classroom observation protocol, familiarity with the material being covered in class does help the observer in cases like this.

Testing validity

To test the validity of BERI, we compared observation data within single class periods, within a single course over multiple class periods, and across multiple courses.

First, we wanted to determine if 10 students could provide an accurate representation of engagement for the whole class. To do this, we conducted an observation where we had three trained observers take simultaneous observations of three groups of 10 students from different sections of the classroom.

Second, to determine the dependence of engagement on location in the classroom, we observed 25 classes in a single course over the semester and compared the average student engagement for classes where the observer sat in the front of the class (9 classes) versus the back (16 classes). The expectation was that students in the front of the class would be more engaged than students in the back of the class.

Third, we tested BERI on seven classes of varying sizes (50–300 students), in different subjects within science, on different course levels (first year to fourth year), and on 14 different instructors with varying instructional methods.

Linking engagement patterns to pedagogy

One of the main goals in developing BERI was to see if instructional methods could be associated with student engagement. To test this, we did the following:

1. Observed the same group of students and the same instructor on two different days. On the first day the instructor used a well-prepared traditional, didactic lecture and on the second day

- delivered an interactive lecture in which students completed a preclass reading and applied their knowledge by answering in-class questions using personal response systems (clickers).
- 2. We conducted 27 observations of a course where two instructors with different levels of experience using student-focused pedagogy taught the same group of students on different days using very similar teaching methods. Both instructors followed a similar format for each class where they presented learning goals for the lecture; mixed didactic lecture with inclass questions, discussions, and real-world examples; tested student knowledge using clicker questions that counted for grades; and ended with a summary of each lecture.

Results and discussion *Validity*

Our findings from simultaneous observations of students in different sections of the class by three trained observers suggest that although the overall level of engagement varies between different groups of students, the general trends over time in student engagement remain consistent. This result implies students are engaged or disengaged by the same in-class activities regardless of where they are sitting in the classroom. Therefore data from a single observer reporting on 10 students can serve as a proxy for trends in engagement for the full class and provide useful feedback to a course instructor.

On the basis of an independent *t*-test, we also determined that the average level of student engagement is significantly higher in the front of the class versus the back of the class,

with 8.4 ± 0.12 (front) and 7.5 ± 0.14 (back) students out of 10 engaged, respectively (p < .001). This difference may be because students who are more likely to be engaged choose to sit in the front or because the proximity of the instructor causes students to be more engaged (Benedict & Hoag, 2004).

BERI was tested and able to differentiate engagement levels in

classes of varying sizes, course levels, and subject matters within science. However, in classes where the instructor wrote notes on a blackboard and students did not receive a copy of the notes before or after class, engagement levels were found to be consistently high. This was because students were taking notes for the entirety of the class period and were classified as being engaged as they fit

FIGURE 2

Average student engagement (number of students engaged) grouped by instructional activity for each instructor in a first-year Oceanography course. Error bars represent standard error. The number of observation sets (n) used to calculate mean and standard error is indicated above each bar. For example, the bar representing "Instructor 2-Clicker" is the average of 82 instances in which 10 students were observed during a clicker question. These numbers also roughly show the distribution of activities during class time, with "Lecture" most common, followed by "Clicker."

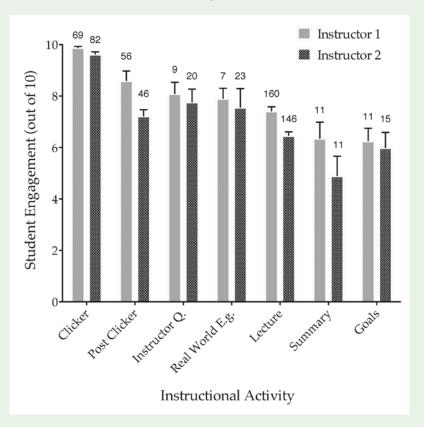
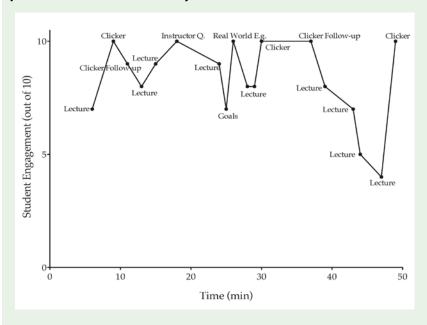


FIGURE 3

Sample of engagement data over a 50-minute class period, showing classroom activities that are more/less engaging. Data like these are provided to instructors shortly after observation.



into the writing category of Table 1. It is worth noting that although we have identified a list of overt student behaviors that indicate whether the student is engaged or disengaged, we cannot confirm how deeply the student is cognitively processing the material to be learned. Students who engage in rote note taking may not actually be cognitively engaged in the class material. We think that BERI is best used in large lecture theater classes where students are given the notes in advance or at the end of class and is best used to determine if the students are on-task or off-task in the classroom.

One concern with classroom observations is that students may realize that they are being observed and therefore alter their behavior. All indications are that the students were unaware that they were being observed. Students were not explicitly told that observers were recording their behaviors, and the observers always sat behind the observed students and dressed casually so as to blend in with the student population. Several observers were asked by students to participate in student activities, indicating that the students next to them did not realize they were not part of the regular student population. Cameras could potentially replace observers; however, human judgments about engagement, based on the footage, would still be needed. We have not explored the challenges cameras would present.

Linking engagement patterns to pedagogy

Having an instructor change the method of instruction had a large effect on student engagement. On the day the instructor used a well-prepared traditional didactic lecture, engagement for the 10 students averaged to 5.7 ± 0.09 over the class period. In contrast, when the same instructor gave an interactive lecture, engagement increased to 8.6 ± 0.14 for the same group of students.

We also found that the level of engagement related to various instructional methods is similar for different instructors. Figure 2 shows data on student engagement from a large firstyear class where teaching is split between two different instructors with different levels of experience with interactive teaching techniques. Both instructors used similar instructional methods, and average engagement for each instructor was grouped on the basis of the classroom activity. Based on average numbers of students engaged out of 10, we found that the most engaging activities are clicker questions and clicker question follow-up ("Post Clicker" in Figure 2); the least engaging activities are instructor lecture, summaries, and learning goals. These data also show that the most common activity during the observation semester was lecture. followed by clicker questions and follow-up (Figure 2). This type of information can help instructors reflect on their current practices and iterate toward more engaging methods.

Some differences between instructors for the same instructional practice are revealing. For example, although there is little difference for many practices, there is a notable difference in the average level of engagement for the clicker question follow-up ("Post Clicker" in Figure 2). These observations were conducted during the first term in which Instructor 2 used clickers in class, whereas Instructor 1 had used clickers for 3 years. As with any pedagogy, it takes

practice to improve one's skill both in creating clicker questions that will generate discussion and in responding in real time to clicker results in ways that encourage engagement. These observation data can serve to illuminate for Instructor 2 a specific instructional skill on which to focus and improve.

Providing formative feedback to instructors

Figure 3 shows a typical studentengagement-versus-time plot for a 50-minute lecture. This demonstrates how BERI allows an instructor to easily see how student engagement rose and fell in relation to what was happening in class, including the instructor's own behaviors. For example, using Figure 3 and observation data recorded in a copy of the instructor's class notes, the instructor can see that when he or she asked a clicker question in class, there were high levels of student engagement. Engagement dropped when the instructor discussed learning goals for the lecture at the 24-minute mark; however, when the instructor subsequently discussed the relevance of the class material and its application to the real world, engagement increased. When the instructor lectured for more than 10 consecutive minutes toward the end of class using equation-heavy slides, the engagement dropped down to 40%, but when a clicker question was used in the last minute of class, engagement rose significantly. Providing instructors with formative feedback such as this helps them see how they can improve engagement in their teaching.

Conclusion

BERI is the first systematic classroom observation instrument for large university classes that provides quantitative data identifying student behavioral engagement. It can be used with high reliability with only one hour of training. It provides a quantitative measure of how the level of engagement depends on an instructor and the teaching practices used by that instructor. Although our main purpose here is to introduce the instrument and examples of its use, BERI could supplement a wide variety of work from testing new pedagogies to informing professional development for faculty members to relating engagement to student performance. We believe this instrument will be a valuable tool for improving instructor's teaching practices to achieve greater student engagement and, ultimately, improved learning.

Acknowledgments

This work was supported at the University of British Columbia (UBC) through the Carl Wieman Science Education Initiative. Appreciation goes to the many Science Teaching and Learning Fellows who tested and used the protocol. Approval to observe classrooms and instruction at the UBC and publish results of that work is provided to the Carl Wieman Science Education Initiative by the UBC under the policy on institutional research.

References

Assor, A., & Connell, J. P. (1992). The validity of students' self-reports as measures of performance-affecting self-appraisals. In D. H. Schunk & J. Meece (Eds.), *Student perceptions in the classroom* (pp. 25–46). Hillsdale, NJ: Erlbaum.

Baddeley, A. D., Lewis, V., Eldridge,M., & Thomson, N. (1984).Attention and retrieval from long-term memory. *Journal of*

- Experimental Psychology: General, 13, 518–540.
- Benedict, E. M., & Hoag, J. (2004). Seating location in large lectures: Are seating preferences or location related to course performance? *Journal of Economic Education*, 35, 215–231.
- Blatchford, P., Edmonds, S., & Marin, C. (2003). Class size, pupil attentiveness and peer relations. *British Journal of Educational Psychology*, 73, 15–36.
- Bransford, J., Brown, A. L., & Cocking, R. R. (2000). How people learn: Brain, mind, experience, and school (expanded edition). Washington, DC: National Academies Press.
- Chapman, E. (2003). Alternative approaches to assessing student engagement rates. *Practical Assessment, Research & Evaluation, 8*(13). Retrieved from http://pareonline.net
- Fenollar, P., Roman, S., & Cuesta, P. J. (2007). University students' academic performance: An integrative conceptual framework and empirical analysis. *British Journal of Educational Psychology*, 77, 873–891.
- Greenwood, C. R., & Delquadri, J. (1988). Code for Instructional Structure and Student Academic Response (CISSAR). In M. Hersen & A. S. Bellack (Eds.), *Dictionary of behavioural assessment* (pp. 120–122). Elmsford, NY: Pergamon.
- Hake, R. R. (1998). Interactiveengagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66, 64–74.
- Harris, A. H., & Cox, M. F. (2003). Developing an observation system to capture instructional differences

- in engineering classrooms. *Journal* of Engineering Education, 92, 329–336.
- Kuh, G. D. (2009). The National Survey of Student Engagement: Conceptual and empirical foundations. *New Directions for Institutional Research*, 141, 5–21.
- O'Malley, K. J., Moran, B. J., Haidet, P., Schneider, V., Morgan, R. O., Kelly, P. A., . . . Richards, B. (2003). Validation of an observation instrument for measuring student engagement in health professions settings. *Evaluations and the Health Professions*, 26, 86–103.
- Piburn, M., Sawada, D., Falconer, K., Turley, J. Benford, R., & Bloom, I. (2000). *Reformed Teaching*

- Observation Protocol (RTOP) (ACEPT Technical Report No. IN-003). Tempe, AZ: Arizona State University.
- Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, 12, 618–627.
- Waxman, H. C., & Padron, Y. N. (2004). The uses of the Classroom Observation Schedule to improve classroom instruction. In H. C. Waxman, R. G. Tharp, & R. Soleste Hilberg (Eds.), *Observational research in U.S. classrooms: New*

- approaches for understanding cultural and linguistic diversity (pp. 72–96). Cambridge, England: Cambridge University Press.
- Young, M. S., Robinson, S., & Alberts, P. (2009). Students pay attention! Combating the vigilance decrement to improve learning during lectures. *Active Learning in Higher Education*, 10, 41–55.

Erin S. Lane is a former science teaching and learning fellow and Sara E. Harris (sharris@eos.ubc.ca) is a senior instructor, both in the Earth, Ocean and Atmospheric Sciences Department and the Carl Wieman Science Education Initiative at the University of British Columbia in Vancouver, British Columbia, Canada.