

Biosensors Show Promise as a Measure of Student Engagement in a Large Introductory Biology Course

Karen S. McNeal,^{†*} Min Zhong,[‡] Nick A. Soltis,[†] Lindsay Doukopoulos,[§] Elijah T. Johnson,[†] Stephanie Courtney,[†] Akilah Alwan,[†] and Mallory Porch[†]

[†]Department of Geosciences, [‡]Department of Biological Sciences, [§]Biggio Center for the Enhancement of Teaching and Learning, and [†]Department of English, Auburn University, Auburn, AL 36849

ABSTRACT

This study measured student engagement in real time through the use of skin biosensors, specifically galvanic skin response (GSR), in a large undergraduate lecture classroom. The study was conducted during an intervention in an introductory-level biology course ($N = 420$) in which one section of the course was taught with active-learning approaches and the other with traditional didactic teaching. GSR results were aligned and correlated with the Classroom Observation Protocol for Undergraduate STEM, or COPUS, and student self-reflections on their own engagement. Results showed that the active-learning section spent more time working in groups, resulting in GSR measures that trended higher and self-reported engagement, while showing indications of higher content learning gains compared with the traditional lecture section. Comparisons between COPUS scores and GSR readings indicate that engagement increased during group work and decreased during listening activities. Throughout a class period, GSR activity of the active-learning group showed increased trends compared with baseline measures, while the traditional lecture group showed decreased trends. Results indicate that GSR is a promising measure of real-time student engagement in the undergraduate classroom, bringing a new technique to discipline-based education researchers who aim to better measure student engagement; however, some limitations exist for broad-scale implementation.

INTRODUCTION

Research in various science, technology, engineering, and mathematics (STEM) disciplines over the last few decades has revealed a suite of high-impact teaching strategies (e.g., reformed-based teaching and learning, active learning, or student-centered learning) that contribute to improvements in several aspects of performance among college students (Singer *et al.*, 2012), including improvements in student learning (e.g., Derting and Ebert-May, 2010; Deslauriers *et al.*, 2011; Freeman *et al.*, 2014), increased retention rates (Russell *et al.*, 2007; Haak *et al.*, 2011; Eddy and Hogan, 2014), and reductions in the achievement gap among student populations (Haak *et al.*, 2011; Eddy and Hogan, 2014; Theobald *et al.*, 2020). A fundamental assumption of these reformed-based teaching practices is that the students are more engaged during the learning process when these teaching techniques are employed. Foundational to constructivist approaches is the principle that engagement is an essential part of the process in student learning (Vygotsky, 1978; Coates, 2005). We approach our theoretical framework similarly to Soltis *et al.* (2020), relying on the work of Fredricks *et al.* (2004), Malmivou and Plonsey (1995), and Bloom (1956). These researchers divide engagement into three dimensions: behavioral, emotional, and cognitive. Behavioral engagement describes how involved students are with the task at hand. The emotional dimension relates to student interest and emotional response. Finally, cognitive engagement relates to the level of rigor and challenge of the learning

C. Gary Reiness, *Monitoring Editor*

Submitted Aug 22, 2019; Revised Jul 13, 2020; Accepted Jul 21, 2020

CBE Life Sci Educ December 1, 2020 19:ar50
DOI:10.1187/cbe.19-08-0158

*Address correspondence to: Karen S. McNeal (ksm0041@auburn.edu).

© 2020 K. S. McNeal *et al.* CBE—Life Sciences Education © 2020 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

activities in which students are engaged. There has been much work to show that all three—emotional, cognitive, and behavioral engagement—are connected and correlated, often linking to student performance (Pekrun *et al.*, 2002; Linnenbrink and Pintrich, 2004; Chaouachi *et al.*, 2010; Goetz *et al.*, 2010). However, direct measurement of student engagement in classroom settings can be challenging.

Attempts to measure engagement in the undergraduate STEM classroom, including biology courses, have come in many forms. Several classroom observation tools have been designed to measure student (and instructor) participation and behavior in the classroom (Sawada *et al.*, 2002; Hora and Ferrare, 2010; Smith *et al.*, 2013; Eddy *et al.*, 2015; Lane and Harris, 2015). However, these tools are limited, because they only capture overt behaviors and do not measure the internal values students place on an activity. Consequently, other tools have been developed to obtain students' reflections on their own levels of cognitive and emotional engagement, providing a more detailed picture of a student's level of involvement (Pritchard, 2008; Chi and Wylie, 2014; Wiggins *et al.*, 2017). However, even these assessments can be limited, as they rely on student self-report as the primary data source, which although informative, may still fall short in obtaining a reliable and unbiased engagement data set, because they are documented postactivity and students must recall how they felt when the activity occurred. While engagement can be measured through observation and surveys, these results can be hard to validate and difficult to interpret.

The limitations for documenting instructor engagement are similar to those for student engagement. For instance, many classroom observation protocols document what the instructor's overt activities and behaviors are (e.g., Owens *et al.*, 2017; Smith *et al.*, 2013; Sawada *et al.*, 2002), but they rarely measure how the chosen activity may affect the instructor, as the focus is usually on the student learning impacts. In fact, there do not seem to be any documented cases that specifically show how the instructor physiologically responded as a result of the chosen classroom teaching behaviors.

Galvanic skin response (GSR) has been introduced to address the need for unbiased physiological data that provide information about subject engagement as a result of external stimuli (Poh *et al.*, 2010) and has recently been applied to the undergraduate classroom (McNeal *et al.*, 2014; Soltis *et al.*, 2020). GSR captures electrodermal activity (EDA), a measure of subtle electrical changes across the skin's surface due to changes in the amount of perspiration or moisture levels of the skin, which are controlled by sweat gland activity. As the sweat glands are regulated by the sympathetic nervous system, it has been postulated that increased sweat gland activity signifies both psychological and physiological arousal (Malmivuo and Plonsey, 1995; Dragon *et al.*, 2008). Changes in skin conductance, or the electrical response when low voltage is conducted by the skin, correspond to activation of the sympathetic nervous system, the same system that controls the body's flight-or-fight response. The sympathetic nervous system is activated by excitement or stressors. These stressors may be physical, emotional, or cognitive (Malmivuo and Plonsey, 1995)—the same three dimensions of engagement described by Fredricks *et al.*, 2004. Thus, skin conductance is a useful proxy for engagement (McNeal *et al.*, 2014; Soltis *et al.*, 2020).

Skin conductance is measured with a pair of electrodes touching the skin. Here, sweat that is generated by the wearer of the GSR device will increase the flow of electrons between the electrodes, and the EDA reading will increase. Traditionally, skin conductance has been measured on the palm; however, more recent research has demonstrated that the EDA signal is also discernible and usable when the biosensor is worn on the wrist (Poh *et al.*, 2010; Van Dooren *et al.*, 2012; Soltis *et al.*, 2020). Some issues exist with collecting GSR data, such as improper placement of the skin-sensing unit on the research subject or excessive sweat production making it difficult to identify meaningful skin conductance responses (Dawson *et al.*, 2007). However, such obstacles can commonly be controlled through placing the skin sensor on the wrist instead of the palm, as the wrist often produces less sweat and the wrist placement eliminates the possibility that hand movements by wearers could interfere with contact between the skin and the GSR device (Potter and Bolls, 2011). Skin conductance can be used to examine the level of engagement over sustained periods (hours) or during short-lived stimuli and bursts of activity (Benedek and Kaernbach, 2010; McNeal *et al.*, 2014).

Skin conductance methods have become common in psychological research and often form an important part of a larger set of instrumentation used to monitor biophysical responses, such as heart rate, muscle activity, skin temperature, blood-volume pulse, cardiac activity, and respiration (Haag *et al.*, 2004; Shen *et al.*, 2009); skin sensors are often used in place of some of these other psychophysiological measures due to the comparative ease of measurement and analysis of GSR data (Andreassi, 2007; Potter and Bolls, 2011; Kim, 2018). Studies have measured stress among autistic students during social interactions using skin biosensors (O'Haire *et al.*, 2015; Goodwin, 2016). Yet others have examined cognitive engagement during problem solving (Pecchinenda, 1996). Some studies have determined emotional engagement while playing violent video games (Ivory and Kalyanaraman, 2007; Ravaja *et al.*, 2008), watching a television show (Gregersen *et al.*, 2017), or interacting in a virtual environment (Beeli *et al.*, 2008). A number of studies have detected engagement during physical activity like dancing, painting, using scooters, jumping into a ball pit, and riding a zipline while using skin biosensors (Latulipe *et al.*, 2011; Hedman *et al.*, 2012).

Several studies have explored the role of affect in learning by using skin conductance (Dragon *et al.*, 2008; Arroyo *et al.*, 2009; Shen *et al.*, 2009; Woolf *et al.*, 2009; Hardy *et al.*, 2013; Kim, 2018). In the undergraduate classroom, skin biosensors have been used to measure engagement during various interventions, including the comparison of didactic lecture approaches to movie viewing and small-group discussions in a geoscience classroom (McNeal *et al.*, 2014; Morrison *et al.*, 2020) and measuring the effect of augmented reality on student engagement in the geoscience classroom (Soltis *et al.*, 2020). However, these studies were sample-size limited, and skin sensor data were not always correlated with more traditional measures such as classroom observation tools and student reflections of engagement.

In the current study, to address the need for an unbiased tool that can collect real-time biophysical levels of student engagement, we employed skin sensors with both students and the instructor in a large undergraduate introductory biology

classroom. Our primary goal was to use and explore skin sensors as a meaningful measurement tool for student and instructor engagement in large undergraduate STEM classrooms. We compared the collected GSR data with traditional measures of student engagement (e.g., classroom observations and student self-reflections of engagement) to provide triangulation in our data collection. We wanted to determine whether the GSR approach could detect differences in engagement of students during different teaching modalities (active learning vs. traditional learning) in two different sections of an undergraduate biology classroom. We also collected data on changes in student academic performance in each treatment through a pre-post concept inventory assessment. Finally, we conducted an exploratory study on the engagement of the instructor when teaching using the two different approaches employed in this study. We set out to address these research questions: How does teaching approach (active learning vs. traditional lecture) impact student engagement in the undergraduate biology classroom as measured using skin biosensors? How does instructor engagement change as a result of the chosen teaching approach?

METHODS

This research study used a mixed-methods case study approach in which multiple quantitative and qualitative measures were employed, including a pre-post assessment on introductory biology content knowledge (students only), a classroom observation tool (outside observers recording student and instructor activities), the GSR skin sensor measure of skin conductivity (students and instructor), and data via the postclass self-reflection engagement form (students only) were collected. Human subjects' research approval was obtained by the appropriate institutional review boards, and participant consent was obtained for this research project.

Study Setting and Population

Principles of Biology is a core course for science majors at the studied university, with more than 2000 students registered every year. This course is an introduction to the fundamental biological principles common to all organisms. Emphasis is placed on basic chemistry, cell structure and function, energy processing and metabolism, cell division, genetics, evolution, ecology and diversity, and other related topics.

Class Setting

The study was composed of two treatments of an introductory biology course taught by the same instructor (M.Z.) who used two different teaching approaches, one in each section of the course: active learning and traditional lecture. Both sections of the course were taught in the Fall semester of the same year. The active-learning section met two times a week, 75 minutes each time; and the traditional lecture section met three times a week, 50 minutes each time. Both classes were similar in enrollment, with 200–220 students. More than 95% of students were freshman and all were science majors, including pre-health related majors (57% traditional, 46% active-learning), animal science and pre-veterinary majors (11% traditional, 17% active-learning), other STEM majors (16% traditional, 17% active-learning), and other majors (16% traditional, 20% active-learning). Gender, ethnicity, and student classification of year in college were very similar for the two sections (Table 1).

Of the students enrolled in the course, 138 and 168 students opted into the research study in the active and traditional lecture treatments, respectively. All active-learning and traditional lecture observations occurred in the same week to ensure the same content was taught in both classes. There were three mid-term exams and a final exam during the semester. None of the classroom observations or skin sensor recordings were made during exam days. The first exam was implemented 3 weeks before the first class observation and skin sensor recording, so there was adequate time for the instructor and students to establish a classroom routine. The second observation/recording occurred the week following and was given one class before the second exam. The third observation/recording occurred the week before the third exam. The fourth observation/recording occurred between the third exam and finals. The fifth observation/recording occurred the last week of classes before the final exam period began. The final exam was given 1 week after the last recording (see Supplemental Material for course syllabus).

There are many different and effective active-learning strategies (Tanner, 2013) and various ways to implement active learning (Borrego *et al.*, 2013; McConnell *et al.*, 2017). The pedagogy of just-in-time teaching (Novak *et al.*, 1999) was used in the active-learning class with preclass assignments and multiple in-class activities with prompt feedback using in-class learning assistants (LAs). In each learning unit, the preclass materials included lecture videos, reading materials, PowerPoint slides, and practical assignments. In class, students usually took an entrance quiz followed by a brief lecture (10–15 minutes). Then multiple activities were provided based on the learning goals for the particular class, including team-based discussion, immediate feedback assessment technique quizzes, and essay- or video-based learning. Each unit usually ended with an exit quiz. In the traditional lecture class, students learned individually by listening to lectures and taking notes. Learning materials and resources from the active-learning course were also available to students enrolled in the traditional lecture. Supplemental instruction leaders sat in both classes and assisted students outside class during tutoring hours. During active-learning classes, there were 12 LAs in class to assist student teams. The main duties of LAs were to answer students' questions during class activities and assist teamwork to achieve high efficiency. There were no LAs in the traditional lecture classroom due to the lack of class activities and the standard lecture format.

The instructor (M.Z.) of the course has been teaching the course under study for 7 years at the undergraduate level. She used a traditional lecture approach for the first 3 years and transformed the class to active learning for the past 4 years. As such, both teaching modalities had been previously implemented in the same course with the same instructor (M.Z.) before this research study.

Recruitment and Incentives

Students enrolled in the active-learning and traditional lecture classrooms were recruited to the research study through course announcements given by the researchers. Although a researcher in the project, the lead instructor (M.Z.) was not allowed to recruit students, know which students had participated, or view student data during the semester as per IRB policies relating to coercion. Students who did not consent to participate in the

TABLE 1. Student demographic data for the two study treatments

	Active learning	Traditional lecture
Gender		
Female	132 (71.7%)	137 (72.87%)
Male	51 (27.7%)	50 (26.60%)
Ethnicity		
American Indian/Native American	1 (0.54%)	2 (1.06%)
Asian/Asian American	7 (3.80%)	4 (2.13%)
Black/African/African American	15 (8.15%)	10 (5.32%)
Latino(a)/Hispanic	7 (3.80%)	3 (1.60%)
White/Caucasian	150 (81.52%)	165 (87.7%)
Other	4 (2.17%)	4 (2.13%)
Classification		
Freshman	148 (80.43%)	154 (81.91%)
Sophomore	28 (15.22%)	28 (14.89%)
Junior	5 (2.72%)	4 (2.13%)
Senior	1 (0.54%)	1 (0.53%)
Super-senior	2 (1.09%)	1 (0.53%)

research study still completed pre–post content assessments as part of their regular classroom activities (these data were later removed from the data set), but did not complete any other portion of the research study. Only students who consented to participate completed a background information survey (See supplemental materials for background survey), wore skin sensors, and completed self-reflection engagement forms. Extra credit was awarded to the entire class when the course as a whole completed an 85% return rate of the background information survey. Students who wore skin sensors and completed engagement forms did not receive any additional incentives.

Skin Sensors

Due to a limited number of sensors, skin conductance readings were collected on a subset of randomly selected students from the pool of those who consented to participate in the research during each observed class. In total, 78 students were measured during the active-learning class section and 85 students during the traditional lecture section. Approximately 15–17 students in each class wore skin sensors during the five class periods that were observed for Classroom Observation Protocol for Undergraduate STEM, or COPUS measurements (described in more detail in *COPUS Observations*). Each of the students selected to wear skin sensors was notified via email before class and instructed to meet the researchers at the front of the classroom a few minutes before class started. Students were provided an Empatica E4 wrist sensor and engagement feedback form at the beginning of class, and researchers assisted them in correctly placing the sensors on their wrists, ensuring the sensors were turned on and collecting data. Once the sensor made skin contact, it automatically began collecting data. Students were asked to place it on their non-writing arm and not to press on the unit or make unnecessary movements while wearing the device. Students were shown how to press the button on the skin sensor that, when pressed, “marks” or timestamps specific events in the GSR data, allowing researchers to make whole-group alignments during data analyses. At the start of class, the instructor asked students to push the button on the skin sensor.

At the end of class, students returned the devices to the researchers along with their completed self-reflection engagement forms. The instructor was also provided a skin sensor to wear on her wrist during one of the five classes in each treatment.

The data from the sensors were then downloaded in Excel format from the Empatica software program provided by the sensor manufacturer. Data were collected every 0.125 seconds over the period of the classroom activities. The first 5 minutes of subjects wearing the sensor were considered a calibration period and the second 5 minutes of subjects wearing the sensor were considered a benchmarking period. Graphs of the data in Empatica allowed for visual inspection to ensure that data were complete with no obvious holes or gaps in skin contact. Data cleaning for calibration periods and percent change calculations were completed by the researchers, as the Empatica program was not able to perform these functions.

The skin sensors collected room temperature and accelerometer data in addition to skin conductance. Limited room temperature (0.18°C) changes were measured during each class, and students were all sitting in their chairs during each activity, with accelerometer data showing minimal movements. As such, temperature and movement were not shown to be factors on the collected skin conductance data.

Student Self-Reflection Engagement Forms and Post Instructor Discussions

At the end of the class, students who wore skin sensors completed a self-reflection form. Students were asked to rank their engagement in the class period and to compare their engagement in the class period with their engagement in other courses using Likert-scale response options (scored 1–5 with options ranging from not engaged to somewhat engaged to highly engaged). Students were also give three open-ended questions that asked them to list the activities they found most and least engaging and why and to report on any aspects of the class that may have captured their attention (see Supplemental Material for engagement form).

The instructor (M.Z.) also engaged in a brief 5- to 10-minute informal discussion with researchers after each observed class period. During this time, the instructor shared with the researchers what she felt went well, what did not, and how engaged she felt during the class. The discussions were informal, as the instructor (M.Z.) was also a researcher on the project.

COPUS Observations

COPUS is a course observation protocol designed to collect information on the range and frequency of teaching practices used in an undergraduate STEM lecture course (Smith *et al.*, 2013). The protocol consists of two major components that are organized into several subcategories: student activities (e.g. listening and taking notes, asking questions, group work) and instructor activities (e.g. lecture, posing questions, demonstration). Every 2 minutes, the observer codes the behaviors exhibited by the teacher and students. A third scale on the protocol is the Student Engagement rating. We did not use this scale due to its subjectivity.

In our study, three trained observers who had previous experience with COPUS each attended eight of the 10 observed classes during which skin sensors were used, and two of the three attended the remaining two observations.

TABLE 2. Skin conductance and self-reported engagement results for the two study treatments^a

Treatment	Skin conductance		Average self-reported engagement*	
	N	Average percent change (SD)	N	Average (SD)
Active learning	71	10.56% (29.92)	72	3.51 (0.86)
Traditional lecture	65	−1.23% (58.08)	89	3.17 (0.92)

^aPaired-sample *t* test results between treatments are shown for both the average percent change in skin conductance and the average student self-reported engagement results. Asterisk (*) indicates statistical significance at $p < 0.05$. Self-reported engagement was scored using a Likert scale from 1 (not engaged) to 5 (very engaged). Self-reported scores are for students who wore the skin sensors.

Five classes for the active-learning section and five classes for the traditional lecture section were observed, and each observation was on the same weekly topic and made when students wore GSR sensors. We coded observed behaviors and calculated interrater reliability using the adjacent agreement function in Excel. The overall percent agreement of the raters was 92.4%. In general, there was strong consistency during observations of both the traditional lecture and active-learning courses, which continued to improve throughout the semester. The values for each COPUS element from the five active-learning and five traditional lecture classes were averaged (scores are reported in Figures 3 and 4, which are discussed later in the article).

Pre–Post Conceptual Understanding Assessment

We selected a total of 37 items from three existing biology concept inventories that best matched the content taught in the introductory biology course offered in this study. The concept inventories included a focus on molecular and cell biology (Klymkowsky, 2010; Shi *et al.*, 2010) and genetics (Wick *et al.*, 2013). Because we selected questions from each assessment and then combined them into a new assessment that better fit the content taught in the course, we reanalyzed the reliability of the new assessment and obtained a Cronbach's alpha score of 0.6, an acceptable reliability score (See Supplemental Material for the pre-post concept inventory used in this study).

Data and Statistical Analyses

All data were processed in Microsoft Excel and then moved to Statistical Package for Social Sciences (v. 25.0) for more in-depth statistical analyses on aggregated data. Percent change was calculated for GSR data by subtracting the average skin conductance during the 5 minutes for benchmarking from the average skin conductance during class and then dividing by the average skin conductance during the 5-minute benchmark. Percent change values were cleaned to ensure that they were not biased toward changes between individuals with lower GSR readings by removing values that were more than 3 SDs greater than the mean (around $\pm 150\%$, 27 cases), which likely indicated potential calibration or contact issues. Then, between-treatment *t* tests were conducted on the average percent change GSR data, average pre–post concept assessment scores, and average scores for the student self-reflection forms. Correlation analysis was performed on average COPUS scores (in the major categories of listening and group work) and average GSR readings for these same COPUS time periods. All statistical assumptions (e.g. normality, homogeneity of variance, linearity, and independence) of the parametric tests were met. Gain scores were calculated as (post – pre)/(100 – pre).

RESULTS

For the GSR results, the active-learning classes on average showed the trend of an increase (10.56%) in percent change of skin conductance, while the traditional lecture showed a small decreasing trend in percent change of skin conductance (−1.23%). This difference between treatments is not statistically significant; however, statistical significance testing is likely limited due to the high standard deviations in the data set. Though skin conductance percent change was not significant between treatments, self-reported engagement between treatments was statistically significant and higher in the active-learning class with a medium effect size, $t(159) = -2.438$, $p = 0.016$, $d = 0.52$ (Table 2). Self-reported engagement compared with other courses was also higher in the active-learning section than the traditional section ($M = 2.91$ vs. $M = 2.61$), but not significantly so.

Individual student GSR differences between the two sections were noted after plotting EDA over time for randomly selected individuals. Figure 1 shows a typical student example in each treatment group plotted with the general activities in the class. The example student in the active-learning section had variable EDA during the class, with peaks in activity during periods of group work; classroom EDA measures were higher than baseline measures for this individual. In contrast, EDA measures of the example student in the traditional lecture classroom stayed relatively constant and low throughout most of the class; classroom EDA measures in this case were lower than baseline measurements for this individual.

Similar EDA plots (Figure 2) for the collected measures of instructor GSR were made on the same dates as those shown in the students' plots (Figure 1). During the traditional lecture class period, the instructor was constantly engaged throughout the duration of the class as she presented the PowerPoint lecture to her students. However, during the active-learning classroom, decreases in instructor engagement were measured during student group work. EDA readings stayed relatively constant throughout the first portion of the class, with decreases beginning about midway through the class through the latter half of the class (Figure 2).

Post classroom implementation informal discussions with the instructor (M.Z.) indicated that, during the traditional lecture course she felt the need to talk louder and be funnier (e.g., make jokes) to keep the students' attention during the class session. She also indicated that she felt more stressed, tired, and exhausted after the traditional lecture than after the active-learning sessions from having to be “on” the entire time. These additional efforts could explain the differences in EDA for the instructor between the two treatments.

There was a significant negative correlation between the time students spent listening and their percent change in skin conductance, and a significant positive correlation between the

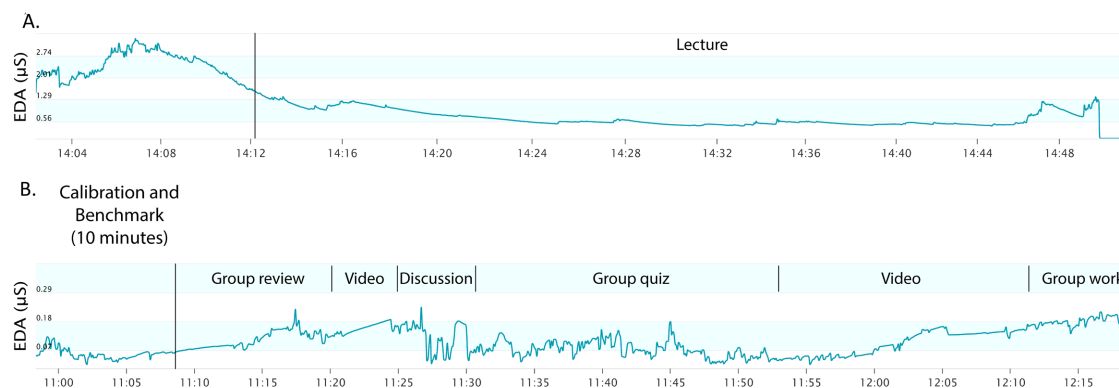


FIGURE 1. Example student skin conductance for each of the study treatments: (A) traditional lecture (50 minutes) and (B) active learning (80 minutes). For both cases, the first 10 minutes of class were considered benchmarking phases when lecture only occurred in both cases.

amount of time students spent working in groups and their percent change in skin conductance (Table 3). This suggests that students become less engaged the more time they spend listening to the instructor and more engaged when working in groups.

Additionally, there was a significant difference between the amount of time the instructor spent lecturing and the amount of time students spent listening between the traditional and active-learning classes (Figures 3 and 4). In the active-learning section, students spent 25% of the class time listening to lecture, whereas in the traditional section of the course, students spent 62% of the class time listening to lecture. The active-learning group also spent at least 16% of their time doing group work and 30% of their time doing other activities, whereas the

traditional group spent no time doing these activities. Additionally, the instructor spent less time lecturing in the active-learning section (26%) than in the traditional section (47%), and spent more time doing activities related to managing the classroom (24%), or facilitation, in the active-learning section. In the traditional lecture, none of her time was spent on classroom management activities.

For the pre-post concept inventory results, there was a significant difference in pretest biology conceptual understanding scores between the two sections, $t(304) = 3.851$, $p < 0.01$, with the traditional lecture class scoring higher (Table 4). However, the difference between the scores was no longer significant at posttest, $t(304) = 1.700$, $p = 0.09$. The increase in total number of questions correct for the active-learning class from pre to post

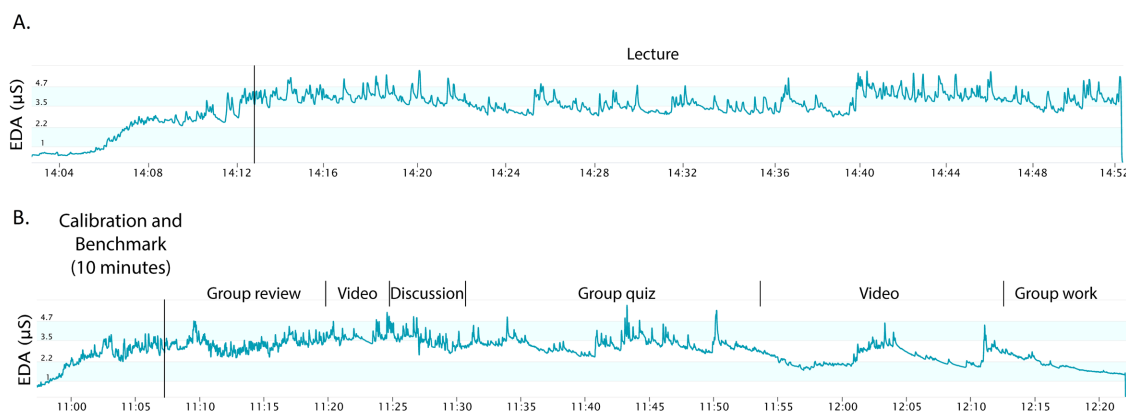


FIGURE 2. Instructor skin conductance for each of the study treatments: (A) traditional lecture (50 minutes) and (B) active learning (80 minutes). For both cases, the first 10 minutes of class were considered benchmarking phases when lecture only occurred in both cases.

TABLE 3. One-tailed Pearson correlation results of average COPUS observations codes for working in groups (WG, OG) and listening (L) with skin conductance.

		Average COPUS score: amount of time listening	Average COPUS score: amount of time working in groups
Skin conductance ($N = 136$)	Pearson correlation	-0.161	0.148
	p value	0.031*	0.043*

*Statistical significance at $p < 0.05$.

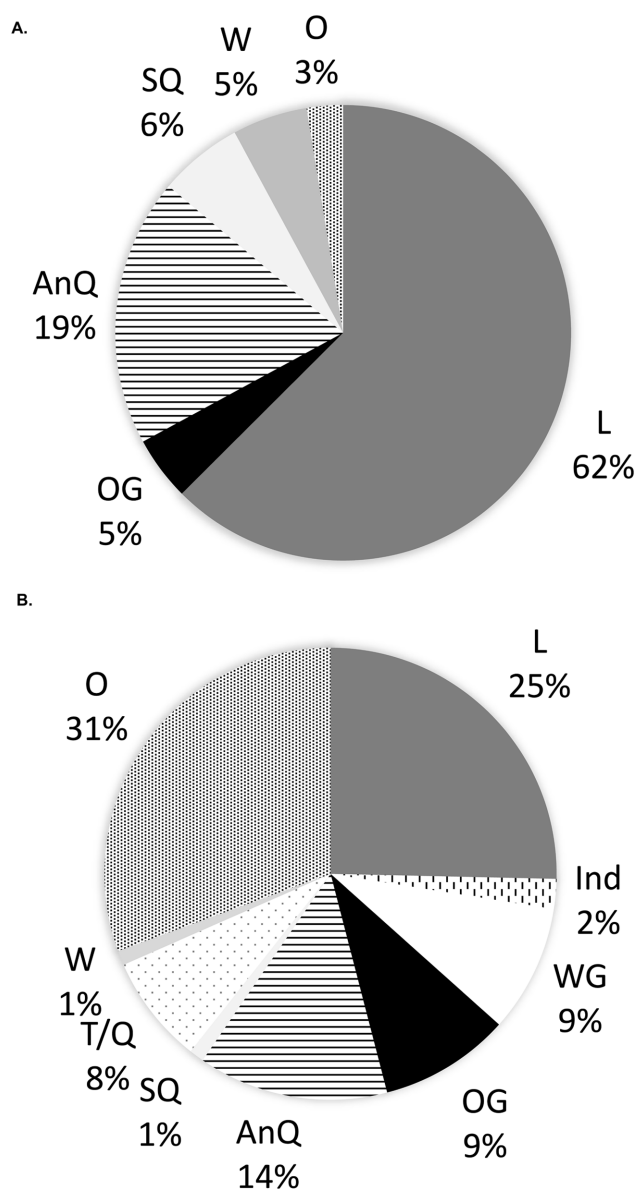


FIGURE 3. Average COPUS results for student activities in both treatments: (A) traditional lecture and (B) active learning. AnQ, students answering questions; Ind, individual thinking/problem solving; L, listening; O, other; OG, other assigned groups; SQ, students ask questions; T/Q, test/quiz; W, waiting; WG, working in groups.

was greater than for the traditional class (5.23 vs. 4.15) and this difference between treatments is significant, $t(305) = -1.974$, $p = 0.049$, $d = 0.227$). Both traditional lecture and active-learning sections had gains in pre-post scores, with the active-learning section showing slightly greater gains, but these gains were not statistically significant ($p > 0.05$) between treatments (Table 4).

DISCUSSION

Although reformed-based teaching and learning practices have long been noted for their impacts on improved student learning and performance outcomes, mostly due to increased student

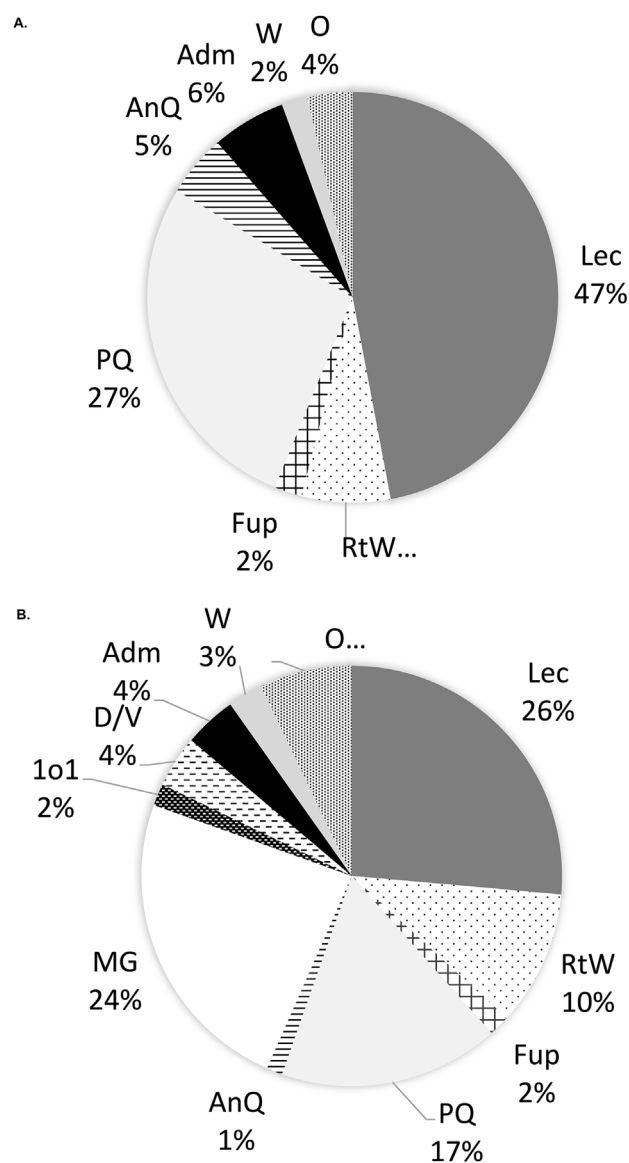


FIGURE 4. Average COPUS results for instructor activities in both treatments: (A) traditional lecture and (B) active learning. 1o1, talking individually to a student; Adm, administration tasks; AnQ, answering questions; D/V, diagram or video; Fup, follow-up; Lec, lecturing; MG, management; O, other; PQ, posing questions; RtW, writing on board; W, waiting.

engagement in the classroom (Singer *et al.*, 2012), it is often difficult to directly measure student engagement. Additionally, although research often monitors what instructors do in their teaching, it does not usually measure instructor engagement or how the chosen teaching modality impacts the instructor. When engagement is measured, it is frequently made through collecting observation information on overt classroom behavior (e.g., Smith *et al.*, 2013) or through post self-reflection surveys (e.g., Wiggins *et al.*, 2017); both of these approaches can have limitations.

This research study took a new approach in which student and instructor engagement was measured using skin conductance, or GSR, in an introductory biology course. The percent

TABLE 4. Pre–post biology content knowledge mean percent scores and between-subjects *t* test results for each of the two study treatments^a

Treatment	Active learning (<i>N</i> = 168)	Traditional (<i>N</i> = 138)	<i>p</i> value
Pre mean score (SD)	32.84 (10.13)	37.38 (10.45)	0.010*
Post mean score (SD)	47.00 (11.95)	49.35 (12.29)	0.090
Gain score ^b (SD)	0.21 (.17)	0.18 (.20)	0.060

^aMaximum content inventory scores were out of 100 total points.^bGain scores calculated using (post – pre)/(100 – pre), range 0–1.*Statistical significance at *p* < 0.05.

change in skin conductance was negatively correlated with the percent time students spent listening to the instructor and positively correlated with the time spent working in groups. These results are similar to those reported by McNeal *et al.* (2014), in which skin conductance measurements showed that students had greater engagement during group work than during lecture periods in a geoscience classroom, and also align with results reported by Morrison *et al.* (2020), in which engagement increased when students dialogued about climate change videos they viewed.

The results also showed that students in the active-learning section had a positive average percent change over baseline measurements in engagement, whereas the traditional lecture instead had a negative average percent change in engagement compared with baseline measurements. Students in the active-learning class also self-reported being more engaged than students in the traditional class section, which aligns with research by Wiggins *et al.* (2017), in which it was found that students in active-learning settings were more engaged when they valued an activity more and when they had a greater personal effort in the activity. These results align with student performance metrics: students in the active-learning classroom had slightly higher content gains than those in the traditional classroom, although these differences were not statistically significant between treatments. Despite initial significant pretest performance differences, with the active-learning group scoring lower, there were no differences in posttest performance at the end of the semester. As such, the active-learning class section made strides to close the content knowledge gap and “level the playing field,” an outcome also noted with such teaching approaches in similar studies (Derting and Ebert-May, 2010; Deslauriers *et al.*, 2011; Wick *et al.*, 2013; Freeman *et al.*, 2014; Theobald *et al.*, 2020).

Most of the attention to active-learning approaches has been placed on the students and their associated learning benefits. However, by taking the active-learning approach, instructors often have to make many changes themselves, including going from “sage on the stage” to “guide on the side” (King, 1993). The resulting classroom instructor behavior is often observed using classroom observation protocols (Sawada *et al.*, 2002; Smith *et al.*, 2013). However, direct measures of instructor engagement during active learning have not been made. In this study, we wanted to measure how the change from traditional lecture approaches to active-learning approaches affected instructor engagement. The exploratory results indicated that instructor engagement levels were consistently high during traditional lecture periods; however, during portions of the active-learning section, most often when students worked in groups, the instructor engagement levels decreased. This is

likely due to the fact that the instructor did not feel she had to be “on” for the entire duration of the active-learning class section and instead had time to facilitate, potentially relieving some of the instructor’s emotional stress and cognitive load for other tasks that may have been less demanding but potentially more beneficial to the students.

Use of Skin Sensors in the Undergraduate Classroom Setting

This study explored the use of GSR as a means to measure student engagement in the active-learning classroom, with results triangulated with COPUS observations, student performance measures, and student self-reported engagement. Higher GSR measurements during active-learning classrooms corresponded with trends of increased student performance and increased self-reports of engagement and were positively correlated with classroom activities that promote student engagement (e.g., group work). This study has shown that skin biosensors are a promising tool to measure student and instructor engagement in the undergraduate classroom. However, more work needs to be completed to validate the results of this study, given the limitations of the current study and the difficulties in analyzing the skin sensor data (large standard deviations, individual GSR differences, etc.).

Limitations

There are several limitations in this study. First, there were high standard deviations among the GSR readings due to the fact that different students wore the skin sensors in each class and data were collected on different days. The sample size was limited, which impacted the types of statistical tests that could be employed. Additionally, variables such as student mood; stress due to in-classroom experiences, such as being assigned to a student group; and out-of-class conditions can potentially be factors causing individual and day-to-day variances in skin conductance. However, these variables cannot be controlled in an experiment like the one used in this study and thus should be treated as a limitation of this research. Another limitation of the study was that the concept inventories that were used for conceptual understanding in the research were difficult for the students (post scores less than 50 out of 100), and for this reason we did not observe very large gains in scores for either section of the course. Furthermore, the sample in our study was a convenience sample of undergraduates in a course that one of the researchers taught where individuals opted to be part of the research. As such, the results of the research may not have included students from all levels of academic ability. Furthermore, this study was conducted in a single instructor’s classroom, in one STEM content area, with only certain active-learning strategies. The study was also limited to a single instructor whose engagement was

only measured during one class period for each of the two sections. Therefore, the results of this study cannot be generalized beyond the conditions presented, and future work should confirm the results in different contexts, learning environments, and student/instructor populations.

Future Recommendations

This study lays the groundwork for future studies exploring real-time student and instructor engagement during varied classroom teaching and learning practices or aim to robustly validate skin biosensors. Because this work has a variety of limitations, we recommend future work be conducted to verify the results and expand on both the student and instructor findings. There are many ways real-time engagement information could be used to continue to refine and improve undergraduate STEM classroom teaching and learning. For instance, additional biosensor studies could examine the role of high-stakes testing on student stress (Ballen *et al.*, 2017). Perhaps student biosensor data could be used as a feedback tool during faculty professional development programs to support and train instructors in using active-learning approaches more frequently and more appropriately in their classrooms. Studies could even be designed to further explore the initial data on instructor engagement during teaching presented here. There is still much more work to do to scale the use of skin biosensors in the classroom (e.g., data automation, analysis algorithms, and lower biosensor equipment costs are needed); however, strides are being made to make these tools more practical. For instance, fitness monitors have been used in a variety of research settings (Cadmus-Bertram *et al.*, 2017) and could be a more affordable option for collecting engagement-related information within the classroom. Additionally, efforts are underway to develop algorithms to analyze EDA (Greco *et al.*, 2016). We hope that this work is a starting point for future research aimed at continuing to explore the affordances of skin biosensors in undergraduate STEM teaching and learning.

ACKNOWLEDGMENTS

We would like to thank the students who participated in this research study as well as the external reviewers for their critical feedback.

REFERENCES

- Andreassi, J. L. (2007). *Psychophysiology: Human behavior and physiological response* (5th ed.). Mahwah, NJ: Erlbaum.
- Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. *Conference on Artificial Intelligence in Education*, 200, 17–24.
- Ballen, C. J., Salehi, S., & Cotner, S. (2017). Exams disadvantage women in introductory biology. *PLoS ONE*, 12(10), e0186419.
- Beeli, G., Casutt, G., Baumgartner, T., & Jäncke, L. (2008). Modulating presence and impulsiveness by external stimulation of the brain. *Behavioral and Brain Functions*, 4(33), 1–7.
- Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4), 647–658.
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: David McKay.
- Borrego, M., Cutler, S., Prince, M., Henderson, C., & Froyd, J. E. (2013). Fidelity of implementation of research-based instructional strategies (RBIS) in engineering science course. *Journal of Engineering Education*, 102, 394–425.
- Cadmus-Bertram, L., Gangnon, R., Wirkus, E. J., Thraen-Borowski, K. M., & Gorzelitz-Liebhauser, J. (2017). The accuracy of heart rate monitoring by some wrist-worn activity trackers. *Annals of Internal Medicine*, 166(8), 610.
- Chaouachi, M., Chalfoun, P., Jraidt, I., & Frasson, C. (2010). *Affect and mental engagement: Towards adaptability for intelligent systems*. In *Proceedings of the 23rd international FLAIRS conference*. Montreal, Canada: AAAI Press.
- Chi, M. T. W., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychology*, 49, 219–243.
- Coates, H. (2005). The value of student engagement for higher education quality assurance. *Quality in Higher Education*, 11(1), 25–36.
- Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In Cacioppo, J. T., Tassinari, L. G., & Bernston, G. G. (Eds.), *Handbook of psychophysiology* (pp. 159–181). Cambridge: Cambridge University Press.
- Derting, T. L., & Ebert-May, D. (2010). Learner-centered inquiry in undergraduate biology: Positive relationships with long-term student achievement. *CBE—Life Sciences Education*, 9, 462–472.
- Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, 332, 862–864.
- Dragon, T., Arroyo, I., Woolf, B. P., Burleson, W., El Kaliouby, R., & Eydgahi, H. (2008). Viewing student affect and learning through classroom observation and physical sensors. In Woolf, B. P., Aimeur, E., Nkambou, R., & Lajoie, S. (Ed.), *Intelligent tutoring systems* (pp. 29–39). Berlin: Springer-Verlag.
- Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE—Life Sciences Education*, 13(3), 453–468.
- Eddy, S. L., Converse, M., & Wenderoth, M. P. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE—Life Sciences Education*, 14(2), ar23.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.
- 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 USA*, 111(23), 8410–8415.
- Goetz, T., Cronjaeger, H., Frenzel, A. C., Ludtke, O., & Hall, N. C. (2010). Academic self-concept and emotion relations: Domain specificity and age effects. *Contemporary Educational Psychology*, 35(1), 44–58.
- Goodwin, M. S. (2016). Laboratory and home-based assessment of electrodermal activity in individuals with autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(10), S301–S302.
- Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2016). cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, 63(4), 797–804.
- Gregersen, A., Langkjær, B., Heiselberg, L., & Wieland, J. L. (2017). Following the viewers: Investigating television drama engagement through skin conductance measurements. *Poetics*, 64, 1–13.
- Haag, A., Goronzy, S., Schaich, P., & Williams, J. (2004, June). Emotion recognition using bio-sensors: First steps towards an automatic system. In *Tutorial and research workshop on affective dialogue systems* (pp. 36–48). Berlin: Springer.
- Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, 332(6034), 1213–1216.
- Hardy, M., Wiebe, E. N., Grafsgaard, J. F., Boyer, K. E., & Lester, J. C. (2013). Physiological responses to events during training use of skin conductance to inform future adaptive learning systems. In *Proceedings of the Human Factors and Ergonomics Society annual meeting* (pp. 2101–2105). Washington, DC.
- Hedman, E., Miller, L., Schoen, S., Nielsen, D., Goodwin, M., & Picard, R. (2012). Measuring autonomic arousal during therapy. In *Proceedings of the 8th International Design and Emotion Conference* (pp. 11–14). London, England.

- Hora, M., & Ferrare, J. (2010). *The teaching dimensions observation protocol (TDOP)*. Madison: Wisconsin Center for Education Research, University of Wisconsin–Madison.
- Ivory, J. D., & Kalyanaraman, S. (2007). The effects of technological advancement and violent content in video games on players' feelings of presence, involvement, physiological arousal, and aggression. *Journal of Communication*, 57, 532–555.
- Kim, P. W. (2018). Real-time bio-signal-processing of students based on an intelligent algorithm for Internet of Things to assess engagement levels in a classroom. *Future Generation Computer Systems*, 86, 716–722.
- King, A. (1993). From sage on stage to guide on the side. *College Teaching*, 41, 30–35.
- Klymkowsky, M. W., Underwood, S. M., & Garvin-Doxas, K. (2010). Biological Concepts Instrument (BCI): A diagnostic tool for revealing student thinking. Retrieved from arXiv:1012.4501 [q-bio.OT]
- 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.
- Latulipe, C., Carroll, E. A., & Lottridge, D. (2011). Love, hate, arousal and engagement: Exploring audience responses to performing arts. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1845–1854). New York, NY: Association for Computing Machinery.
- Linnenbrink, E. A., & Pintrich, P. R. (2004). Motivation, emotion and cognition. In Dai, D. Y., & Sternberg, R. J. (Eds.), *Role of affect in cognitive processing in academic contexts* (eds., pp. 57–87). Mahwah, NJ: Erlbaum.
- Malmivuo, J., & Plonsey, R. (1995). *Bioelectromagnetism: Principles and applications of bioelectric and biomagnetic fields*. New York: Oxford University Press.
- McConnell, D. A., Chapman, L., Czajka, D., Jones, J. P., Ryker, K. D., & Wigen, J. (2017). Instructional utility and learning efficacy of common active learning strategies. *Journal of Geoscience Education*, 65, 604–625.
- McNeal, K. S., Spry, J., Mitra, R., & Tipton, J. (2014). Measuring student engagement, knowledge, and perceptions of climate change in an introductory geology course. *Journal of Geoscience Education*, 62, 655–667.
- Morrison, A. L., Rosaz, S., Gold, A. U., & Kay, J. E. (2020). Quantifying student engagement in learning about climate change using galvanic hand sensors in a controlled educational setting. *Climatic Change*, 159, 17–36.
- Novak, G., Patterson, E. T., Gavrin, A. D., & Christian, W. (1999). *Just-in-time teaching: Blending active learning with Web technology*. Upper Saddle River, NJ: Prentice Hall.
- O'Haire, M. E., McKenzie, S. J., Beck, A. M., & Slaughter, V. (2015). Animals may act as social buffers: Skin conductance arousal in children with autism spectrum disorder in a social context. *Developmental Psychobiology*, 57(5), 584–595.
- Owens, M. T., Seidel, S. B., Wong, M., Bejines, T. E., Lietz, S., Perez, J. R., ... & Tanner, K. D. (2017). Classroom sound can be used to classify teaching practices in college science courses. *Proceedings of the National Academy of Sciences USA*, 114(12), 3085–3090.
- Pecchinenda, A. (1996). The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition & Emotion*, 10(5), 481–504.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91–105.
- Poh, M. Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 57(5), 1243–1252.
- Potter, R. F., & Bolls, P. D. (2011). *Physiological measurement and meaning: Cognitive and emotional processing of media*. New York: Routledge.
- Pritchard, G. M. (2008). Rules of engagement: How students engage with their studies. *Newport CELT Journal*, 1, 45–51.
- Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S., & Keltikangas-Jarvinen, L. (2008). The psychophysiology of James Bond: Phasic emotional responses to violent video game events. *Emotion*, 8(1), 114–120.
- Russell, S. H., Hancock, M. P., & McCullough, J. (2007). Benefits of undergraduate research experiences. *Science*, 316, 548–549.
- Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics*, 102(6), 245–253.
- Shen, L., Wang, M., & Shen, R. (2009). Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment. *Educational Technology and Society*, 12(2), 176–189.
- Shi, J., Wood, W. B., Martin, J. M., Guild, N. A., Vicens, Q., & Knight, J. K. (2010). A diagnostic assessment for introductory molecular and cell biology. *CBE—Life Sciences Education*, 9(4), 453–461.
- Smith, M. K., Jones, F. H., 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(4), 618–627.
- Soltis, N., McNeal, K. S., Atkins, R., & Maudlin, L. (2020). A novel approach to measuring student engagement while using an augmented reality sandbox. *Journal of Geography in Higher Education*. doi: 10.1080/03098265.2020.1771547
- Tanner, K. D. (2013). Structure matters: Twenty-one teaching strategies to promote student engagement and cultivate classroom equity. *CBE—Life Sciences Education*, 12(3), 322–331.
- Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., ... & Freeman, S. (2020). Active Learning narrow achievement gaps for underrepresented students in undergraduate science, technology, engineering and math. *Proceedings of the National Academy of Sciences USA*, 117(12), 6476–6483.
- Van Dooren, M., De Vries, J. J., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurements locations. *Physiology & Behavior*, 106(2), 298–304.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Luria, A. R., Lopez-Morillas, M., & Cole, M., Eds.; Wertsch, J. V., Trans.) Cambridge, Mass: Harvard University Press.
- Wick, S., Decker, M., Matthes, D., & Wright, R. (2013). Students propose genetic solutions to societal problems. *Science*, 341, 1467–1468.
- Wiggins, B. L., Eddy, S. L., Wener-Fligner, L., Freisem, K., Grunspan, D. Z., Theobald, E. J., ... & Momsen, J. (2017). ASPECT: A survey to assess student perspective of engagement in an active-learning classroom. *CBE—Life Sciences Education*, 16, ar32.
- Wolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognising and responding to student affect. *International Journal of Learning Technology*, 4(3), 129–164.