Sensing Behaviors of Students in Online vs. Face-to-Face Lecturing Contexts

Rebecca Medina, Rochester Institute of Technology, rcm3517@rit.edu
Daniel Carpenter, Siena College, d22carp@siena.edu
Joe Geigel, Rochester Institute of Technology, jmg@cs.rit.edu
Reynold Bailey, Rochester Institute of Technology, rjb@cs.rit.edu
Linwei Wang, Rochester Institute of Technology, lxwast@rit.edu
Cecilia O. Alm, Rochester Institute of Technology, coagla@rit.edu

***Abstract*—University students are often presented with the choice between a traditional classroom and an online learning environment. Given the growing interest in web-based learning, it is essential to understand if students' needs are met in these learning environments. Sensing mechanisms enable real-time monitoring of students' reactions as they view and engage with course content. We use galvanic skin response and facial expression analysis to identify differences in behaviors associated with learning via a face-to-face versus an online lecture. We also explore the effects of incentives on learning. Findings indicate that physiological data recorded during a lecture is a good indicator of content difficulty, potentially providing a way for instructors to adjust their materials and delivery to benefit students' understanding. The data further suggests that subjects react more negatively to online lecturing and that learning incentives may have the adverse effect of increasing stress on students as opposed to improving performance.**

***Index Terms*—Multimodal sensing; online lectures; learning performance; content difficulty; facial expressions; galvanic skin response**

I. Introduction

Online coursework is a now a staple in educational settings. Students can enjoy the flexibility of time and location, and educational establishments can benefit from the cost-effectiveness of producing a course with prerecorded lectures that can be reused [12], [30]. The Online Learning Consortium reported that, in the United States, more than one in four students (28%) were taking an online course in 2015 [1]. They also reported close to a 4% increase in the number of distance education students from 2014 to 2015 and pointed out that over 2.8 million American students take courses exclusively online. Studies have suggested equivalent learning outcomes between the two modes, but there is still a difference in the perceptions of online courses among students and teachers [6]. Less than a third of academic leaders report that their faculty accept the "value and legitimacy of online education" [1]. Moreover, studies have shown that face-to-face classes better motivate students [16]. This lack of motivation is one reason for the higher attrition rate among students in online courses [11].

This paper presents the results of an experiment to explore the differences between online and traditional learning environments. Our study aims to observe and quantify the nonverbal feedback that is lost in an online versus a face-to-face lecturing environment. Instructors giving a lecture in a traditional classroom use this in situ feedback to adapt to their students' needs, but online instructors do not have access to similar information. To help provide this information, we collect galvanic skin response (GSR) and facial expression data on students in both online and face-to-face lecture environments. By combining these sensing modalities, our goal is to analyze the attention and behavior of students in each environment. We explore whether student physiological data recorded during a lecture, and subsequently linked to assessment results, provides insight about content difficulty. In addition, we explore if offering incentives has an impact on learning performance in both settings.

Fig. 1: In the study, half of the subjects viewed an online lecture in the video structure shown (top). Audio and video captions were available to students. The other half were given a face-to-face lecture in a classroom with an instructor and two subjects at a time (bottom).

The broader problem in which we are interested is the issue of student-teacher interaction in online courses. Research has shown that student-teacher interaction positively correlates with better student satisfaction and retention [9]. As such, providing feedback to teachers about their students may help improve online learning. Our study attempts to answer the following research questions, in order to identify and create mechanisms that could help improve the interaction and feedback loop between online teachers and students:

RQ1: Does learning performance change between online and face-to-face learning contexts or in the presence of monetary incentives?

RQ2: Do GSR or facial expressions differ in online vs. face-to-face contexts?

RQ3: Can GSR or facial expressions indicate when a student is struggling to learn?

RQ4: Do GSR or facial expressions change in the presence of monetary incentives?

Our contributions include the following:

1) We provide insights into how noninvasive sensing mechanisms can serve as instructor feedback tools in educational environments.
2) We discover relationships between GSR or facial expressions and student behaviors or learning performance.
3) We verify and expand a methodological framework that links lecture content to quiz questions.

II. Related Work

Students and universities like online courses because they are flexible and cost-effective [12], [30]. Students can watch lectures when they choose without being in a physical classroom, and they can rewatch lectures as much as they want. Colleges and universities save money on online courses because, over time, they may require less lecturing time commitment from teachers and can be reused with minor changes. In addition, online courses can achieve similar learning outcomes as face-to-face courses [6], though they tend to differ on measures of student satisfaction [30]. While students perceive online courses as more flexible, they report face-to-face courses as more interactive, thus teaching them more [24].

In education, interaction is important for learning [17]. Studies have noted significant, positive relationships between interaction and perceived engagement, learning, confidence, and student satisfaction [14], [15], [21]. The lack of direct interaction in online courses can result in an increased likelihood that a student will drop out of a class [14], [18], [28] and contributes to the perception among students and teachers that online learning is less effective than traditional learning [1].

Interaction in educational settings is typically characterized as: student-student, student-content, or student-instructor [9]. Meaningful learning can occur when at least one interaction form is present at a high level [2]. Of the forms of interaction, research suggests that student-instructor interaction is the best predictor of course satisfaction [4], [5], [29].

Face-to-face classrooms provide instructors an opportunity to observe student behaviors and take into consideration nonverbal feedback cues, which help them assess when students are confused and struggling with materials or when they are not being attentive. Our study is a step toward addressing the lack of nonverbal feedback loops in online learning environments. We investigate the use of sensing for providing enhanced real-time feedback mechanisms to instructors.

Students experience a wide variety of affective states that may impact motivation and performance [23]. Boredom and confusion have been linked to learning [7]. However, many studies involve self-reported rather than measurement-based affect analysis. In contrast, we measure galvanic skin response (GSR) and facial expressions to explore affective reactions while learning and linking these to performance.

Prior research has used GSR, among other sensing modalities, to measure the cognitive load experienced while performing challenging tasks [26]. One study showed a significant correlation between cognitive load and both accumulated GSR and average GSR for reading and arithmetic tasks [20]. Another study indicated that GSR values, as well as pupil dilation, change during cognitive tasks, but did not establish a clear link between the changing values and cognitive load [8]. GSR and pupil dilation measures have also been used to determine if subjects are in relaxed or stressed states, finding that pupil dilation features resulted in improved classification performance [25]. However, pupil dilation remains impractical in the educational context as it is sensitive to changes in lighting, among other factors, and thus requires a strictly controlled environment [13], [22].

The use of incentives to motivate learning has long been an area of interest to researchers. Performance incentives can lead to greater motivation and effort, but whether this affects learning is uncertain [3], [19]. One study argues that the effect of performance incentives depends on the existence of feedback and experience [27]. Much focus has been on the performance and learning in a work setting rather than a classroom. Therefore, we explore whether a monetary performance incentive promotes student learning or whether it may stress individuals, as measured by GSR and facial expressions.

III. Experimental Design

A. Subjects

Twenty-eight subjects (13 M, 15F) participated in this study. Their ages ranged from 18 to 50, with most being between 18 and 27. Five subjects were not native English speakers. Many were college students and had a wide range of majors.

As shown in Figure 2, we randomly split the subjects evenly into two conditions: viewing an online lecture or

TABLE I: Grading scale provided to most subjects who had performance incentives on the post-lecture quiz with 18 quiz questions.

Questions Correct	Payment
0-3 out of 18	$10
4-7 out of 18	$11
8-11 out of 18	$12
12-14 out of 18	$13
15-17 out of 18	$14
18 out of 18	$15

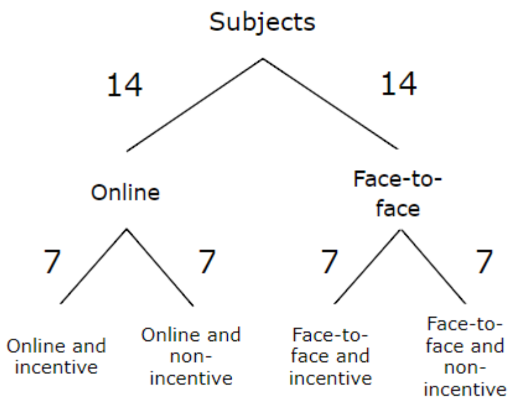

Fig. 2: Twenty-eight subjects were equally split between online and face-to-face lectures. In each condition, half received a performance incentive on a post-lecture quiz.

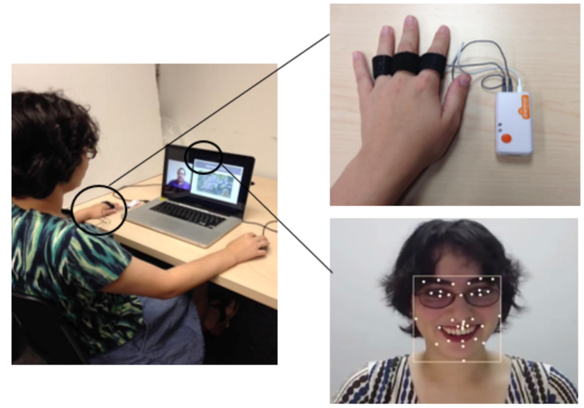

Fig. 3: Experimental setup for online subjects (left). All subjects had a GSR sensor on the hand not occupied by the computer mouse (upper right). Facial video was captured by the built-in webcam (bottom right).

participating in a face-to-face lecture. Subjects were not asked which type of lecture they would prefer so that our findings would apply to a general population, regardless of learning environment preferences. Within both groups, half were given learning incentives and half were not. Participants were briefed individually. This resulted in four test groups consisting of seven subjects each. Non-incentive subjects were paid a $15 participation fee regardless of performance while incentive subjects were told that they would be paid $10 and could earn up to $15 based on the number of quiz questions they answered correctly. To ensure that subjects understood the incentive, a payment scale was provided during the consent process before the experiment (see Table I). Regardless of performance all subjects were actually paid $15 in the end.

B. Setup

In both online and face-to-face lectures, subjects had a MacBook laptop in front of them which they used to enter demographic information in a pre-survey and take a quiz of 18 questions on lecture material. Online students watched the lecture on the laptop. For face-to-face subjects, the laptop was set to display a black screen during the lecture so that subjects would not be distracted from the instructor; however, the laptop camera remained on to capture their facial expressions. The average self-reported prior experience in the lecture material was 2.3 on a scale of 1 to 7. Non-incentive subjects reported more prior experience than incentive subjects (2.6 and 1.8 respectively). Online subjects reported more prior experience than face-to-face subjects (2.4 and 1.9 respectively).

Data collection lasted from the beginning of the pre-survey to completion of the lecture. During this time, we used the MacBook's built-in webcam to record a video of the subject's face and a screen capture video was recorded using Open Broadcaster Software. To record GSR data, we used a Shimmer3 sensor with electrodes placed on the index and middle fingers of the subject's non-dominant hand. Subjects were asked to keep that hand still for the duration of the experiment, as movement creates artifacts in the GSR data. The time of the pre-survey was used to calibrate the GSR sensor.

Online and face-to-face lectures were taught by the same instructor using the same content slides. An effort was made to keep lecture length consistent across online and face-to-face conditions, with an average time of 18 minutes. The lecture covered introductory topics in the 2D Computer Graphics pipeline, which the instructor has expertise in. For the face-to-face lecture, slides were projected onto a screen at the front of a classroom. The video recorded for online lectures used a side-by-side style, with the instructor's face on the left and slides on the right (see Figure 1).

We held face-to-face lectures in a college classroom, with two subjects participating in the lecture at the same time. Subjects sat at the front of the classroom and were told they could ask questions (see Figure 1). Online lectures were held in a conference room one subject at a time. Subjects were oriented so that they faced away from the door, thus minimizing external distractions (see Figure 3). Closed captions were available for any subject who needed them. In addition, we told subjects that they could pause the video and re-watch content, or change the playback speed of the video if they wished. They were instructed that once they proceeded to the quiz to not return to the lecture. In both cases, the researchers provided instructions then exited the room prior to the start of the experiment. This was motivated by observations from a similar study that found that subjects were more at ease and more expressive when they were not actively being watched [10].

IV. Data

Facial videos were post-processed with Affectiva SDK[1] to determine subjects' facial expressions. Affectiva outputs a confidence rating for each of the 7 basic emotions encoding

[1] https://www.affectiva.com/product/emotion-sdk/

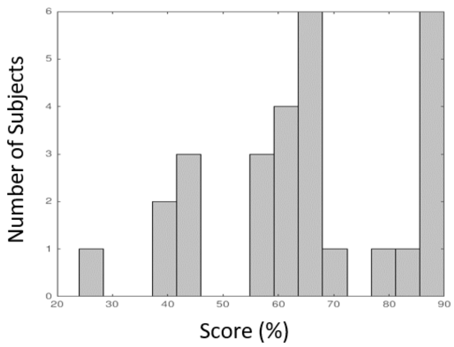

Fig. 4: Quiz scores by number of subjects achieving a certain percentage score (N = 28 subjects). A wide range of scores was obtained from the quiz results, with the lowest being roughly around 30% and the highest being around 90%.

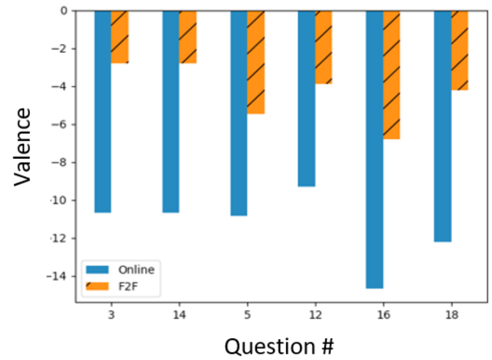

Fig. 5: Facial expression valence was averaged for online and face-to-face subjects (N = 21 subjects). A greater negative valence was detected for online rather than face-to-face subjects. Overall, negative valence was the trend across subjects.

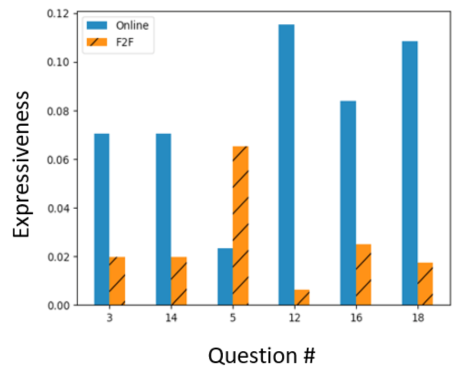

Fig. 6: Facial expressiveness of subjects was averaged for online and face-to-face subjects (N = 21 subjects). Faces were more expressive in the online than the face-to-face condition.

the likelihood that the subject being analyzed is displaying a particular emotion. The facial analysis has a tendency to identify resting faces falsely as expressing contempt or disgust. To address this or other misclassification, we applied a empirical threshold of at least 50 percent confidence in an emotion being identified. For overall facial expressiveness, we calculated the percentage of time that a subject expressed emotion.

GSR data was processed through a median filter for smoothing and noise reduction. Then we normalized each subject's data by dividing by their average GSR value during the entire lecture [20]. This allowed us to compare GSR across subjects. Finally, peak detection was conducted.[2]

After processing was complete, we synchronized data collected from different devices and created a mapping from lecture content to quiz questions. First, we carefully mapped GSR data with analyzed facial expression data based on timestamps. Next, we adapted a method developed by Edwards et al. [10] in which sensing data are mapped to spans of the lecture corresponding to material covered in the quiz questions. To create such question mappings, we identified the time during the lecture in which the answer to each question was first presented. This was done manually to account for face-to-face lecturing variation and that online subjects were allowed to rewind or speed up the lecture. Once these mappings were created, data was segmented by question time in order to view changes in behaviors for each question.

After excluding subjects due to data loss, poor data quality, or uncertain mappings, 21 subjects' sensing data (11 online/10 face-to-face, 11 incentive/10 non-incentive) were analyzed.

V. Results and Discussion

ANOVA tests were used to compare between multiple factors. Results are as follows:

[2]Using relative maxima detection on a continuous wavelet transform of our data, as implemented in SciPy (https://www.scipy.org).

A. Learning Performance

Neither the online vs. face-to-face nor the incentive vs. non-incentive factors showed a significant difference in performance on the quiz. Figure 4 shows that the average score was 65% (σ = 18%).

The observation that performance does not differ significantly between online and face-to-face lecture conditions concurs with prior findings for similar performance across learning settings [6]. The result that incentives do not improve quiz performance adds insight compared to prior work's focus on non-academic settings and use of incentives.

B. Online vs. Face-to-Face

Several differences were found between sensing data in online and face-to-face groups. Emotional valence as measured by facial expressions was more negative for online subjects ($p < 0.001$) (see Figure 5). Online subjects were also more expressive than face-to-face subjects throughout the lecture ($p < 0.05$) (see Figure 6). Online subjects expressing more facial emotion and with greater negative valence may lend support

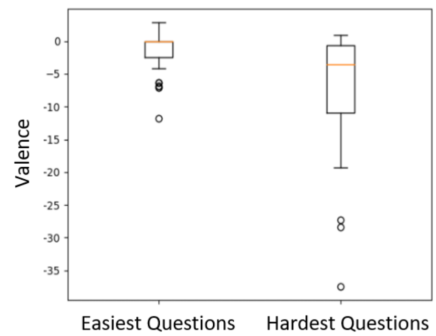

Fig. 7: The facial valence of non-incentive subjects was averaged for the three easiest and three hardest questions (N = 11 subjects), and greater negative valence was observed for difficult questions than for easy questions.

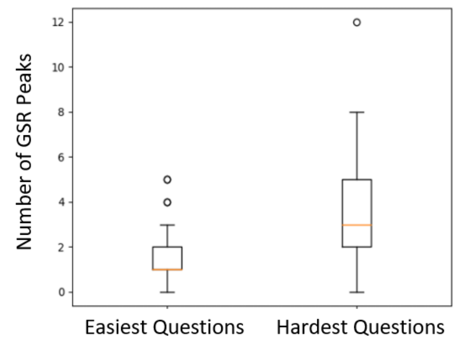

Fig. 8: The number of GSR peaks was averaged for the three easiest and three hardest questions (N = 21 subjects). The number of GSR peaks for easy and difficult questions differed significantly.

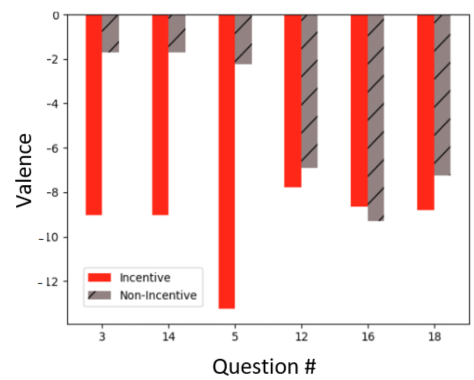

Fig. 9: Facial expression valence of subjects was averaged for incentive and non-incentive subjects (N = 21 subjects). Generally, incentive subjects displayed a greater negative valence than non-incentive subjects.

to prior observations that students are less satisfied with their experience in the online setting [16]. Alternatively, concerns about violating social norms in face-to-face interactions may inhibit overt negative expressions. Another potential reason is that, while online subjects were almost level with the webcam, face-to-face subjects were at an angle as they were looking up at the instructor. Overall, facial expressiveness appeared low, with subjects expressing emotion no more than 12% of the time in either condition; this may reflect our thresholding.

C. Content Difficulty

While not strong enough to predict an individual's question correctness, facial valence may indicate content difficulty. For incorrect responses, subjects showed greater negative facial expression valence than for correct questions ($p < 0.005$). Similarly, non-incentive subjects showed greater negative valence for harder questions ($p < 0.005$) (see Figure 7).

Across all subjects, there was a significant relationship between the number of GSR peaks during the three easiest versus the three most difficult questions in the quiz. A question which simply asked about the topic of the lecture was ignored. Question difficulty was determined by the percent of subjects who answered the question correctly. To consider the complete available evidence regarding the difficulty level of a question, the responses of all 28 subjects were used to determine question difficulty. Subjects had more GSR peaks during difficult questions than for easy ones ($p < 0.001$) (see Figure 8). This result also held when exclusively examining online subjects ($p < 0.001$), or subjects with incorrect responses ($p < 0.001$).

These results affirm prior findings that sensing data can suggest when subjects struggle to learn with challenging material [10]. This indicates that sensing may provide timely feedback to instructors about content difficulty. Given the sample size, and outliers suggested in Figures 7 and 8, more work is needed to additionally confirm the finding.

D. Incentive vs. Non-incentive

Facial expression valence was more negative for incentive subjects ($p < 0.001$) (see Figure 9). Incentive subjects also had a higher average GSR ($p < 0.01$) (see Figure 10) suggesting a stress response among this group.

E. Limitations

While our findings link correctness and difficulty to sensing data, and learning incentives to more negative reactions, the study involved a modest population on a college campus which may be representative of face-to-face but not of online students. Retention, rather than immediate recall, is also the goal of most courses. Investigating effects on content retention is left for future work. Also, not all online courses emphasize video lectures. As we only focused on this aspect of online teaching, we are not able to generalize to other teaching mechanisms. Furthermore, our controlled environment is not representative of the variety of environments that students may take an online class in. Future work should verify findings with a larger sample more representative of the online population and measure retention weeks after experiencing the lecture.

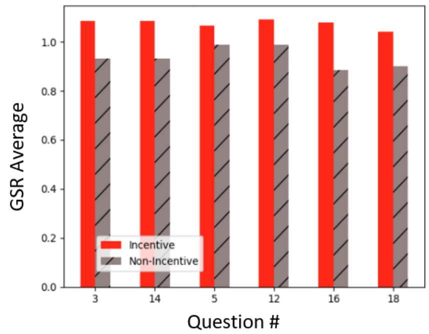

Fig. 10: GSR average for incentive and non-incentive subjects (N = 21 subjects). Incentive subjects overall had a higher average GSR than non-incentive subjects.

VI. Conclusion

This study suggests a link between learning and physiological sensors such as GSR and facial expressions. The results indicate relationships between these sensing modalities and content material difficulty or correctness. There is potential to explore if such real-time information could help instructors improve their delivery or course materials to better support student learning. Moreover, being able to determine proactively when a student is having trouble with a concept during the lecture could help personalize instruction.

We also found distinctions in facial valence and expressiveness for online and face-to-face contexts. More negative valence displayed by online subjects may indicate a way to monitor less-pleasant experience, which potentially could help avoid attrition and lower performance. Additionally, incentives appear to not improve student performance but rather increase stress and negative affective experiences.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Award No. IIS-1559889. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] I. E. Allen, R. Seaman, Jeff with Poulin, and T. T. Straut. Online report card - Tracking online education in the United States. 2015.

[2] T. Anderson. Modes of interaction in distance education: Recent developments and research questions. *Handbook of Distance Education*, pp. 129–144, 2003.

[3] R. D. Banker, S.-Y. Lee, and G. Potter. A field study of the impact of a performance-based incentive plan. *Journal of Accounting and Economics*, 21(2):195–226, 1996.

[4] J. Battalio. Interaction online: A reevaluation. *Quarterly Review of Distance Education*, 8(4):339, 2007.

[5] D. U. Bolliger and T. Martindale. Key factors for determining student satisfaction in online courses. *Intl. J. on E-learning*, 3(1):61–68, 2004.

[6] J. Cavanaugh and S. J. Jacquemin. A large sample comparison of grade based student learning outcomes in online vs. face-to-face courses. *Online Learning*, 19(2), 2015.

[7] S. Craig, A. Graesser, J. Sullins, and B. Gholson. Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3):241–250, 2004.

[8] M. E. Crosby, B. Auernheimer, C. Aschwanden, and C. Ikehara. Physiological data feedback for application in distance education. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, pp. 1–5. ACM, 2001.

[9] R. A. Croxton. The role of interactivity in student satisfaction and persistence in online learning. *Journal of Online Learning and Teaching*, 10(2):314, 2014.

[10] A. A. Edwards, A. Massicci, S. Sridharan, J. Geigel, L. Wang, R. Bailey, and C. O. Alm. Sensor-based methodological observations for studying online learning. In *Proceedings of the 2017 ACM Workshop on Intelligent Interfaces for Ubiquitous and Smart Learning*, pp. 25–30. ACM, 2017.

[11] A. V. Fonolahi, A. Jokhan, et al. Are students studying in the online mode faring as well as students studying in the face-to-face mode? Has equivalence in learning been achieved? *Journal of Online Learning and Teaching*, 10(4):598, 2014.

[12] O. John, S. Main, and M. Cooper. Student perceptions of online interactive versus traditional lectures; Or how I managed not to fall asleep with my eyes open. *Journal of Online Learning and Teaching*, 10(3):405, 2014.

[13] A. F. Kramer. *Physiological Metrics of Mental Workload: A Review of Recent Progress*. Navy Personnel Research and Development Ctr., 1991.

[14] X. Liu, R. J. Magjuka, C. J. Bonk, and S.-h. Lee. Does sense of community matter? An examination of participants' perceptions of building learning communities in online courses. *Quarterly Review of Distance Education*, 8(1):9, 2007.

[15] M. Mahle. Effects of interactivity on student achievement and motivation in distance education. *Quarterly Review of Distance Education*, 12(3):207, 2011.

[16] G. A. Mentzer, J. R. Cryan, and B. Teclehaimanot. Two peas in a pod? A comparison of face-to-face and web-based classrooms. *Journal of Technology and Teacher Education*, 15(2):233, 2007.

[17] M. G. Moore and G. Kearsley. *Distance Education: A Systems View of Online Learning*. Cengage Learning, 2011.

[18] L. V. Morris, C. Finnegan, and S.-S. Wu. Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, 8(3):221–231, 2005.

[19] G. R. Morrison, S. M. Ross, M. Gopalakrishnan, and J. Casey. The effects of feedback and incentives on achievement in computer-based instruction. *Contemporary Educational Psychology*, 20(1):32–50, 1995.

[20] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pp. 420–423. ACM, 2012.

[21] B. Offir, Y. Lev, and R. Bezalel. Surface and deep learning processes in distance education: Synchronous versus asynchronous systems. *Computers & Education*, 51(3):1172–1183, 2008.

[22] T. Partala and V. Surakka. Pupil size variation as an indication of affective processing. *Intl. J. of Hum.-Comp. Stud.*, 59(1):185–198, 2003.

[23] R. Pekrun, T. Goetz, W. Titz, and R. P. Perry. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Edu. Psychol.*, 37(2):91–105, 2002.

[24] C. A. Platt, N. Amber, and N. Yu. Virtually the same?: Student perceptions of the equivalence of online classes to face-to-face classes. *Journal of Online Learning and Teaching*, 10(3):489, 2014.

[25] P. Ren, A. Barreto, Y. Gao, and M. Adjouadi. Comparison of the use of pupil diameter and galvanic skin response signals for affective assessment of computer users. *Biomedical Sciences Instrumentation*, 48:345–350, 2012.

[26] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen. Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, pp. 2651–2656. ACM, 2007.

[27] G. B. Sprinkle. The effect of incentive contracts on learning and performance. *The Accounting Review*, 75(3):299–326, 2000.

[28] S. F. Tello. An analysis of student persistence in online education. In *Information Communication Technologies: Concepts, Methodologies, Tools, and Applications*, pp. 1163–1178. IGI Global, 2008.

[29] V. A. Thurmond. *Examination of Interaction Variables as Predictors of Students' Satisfaction and Willingness to Enroll in Future Web-based Courses while Controlling for Student Characteristics*. Universal-Publishers, 2003.

[30] S. Young and H. E. Duncan. Online and face-to-face teaching: How do student ratings differ? *Journal of Online Learning and Teaching*, 10(1):70, 2014.