Sensor-based Methodological Observations for Studying Online Learning

Ashley Edwards
SUNY Geneseo
ae5@geneseo.edu

Anthony Massicci
Onondaga Community College
anthony.massicci@outlook.com

Srinivas Sridharan
Stevens Institute of Technology
ssridha3@stevens.edu

Joe Geigel
Rochester Institute of Technology
jmg@cs.rit.edu

Linwei Wang
Rochester Institute of Technology
lxwast@rit.edu

Reynold Bailey
Rochester Institute of Technology
rjb@cs.rit.edu

Cecilia Ovesdotter Alm
Rochester Institute of Technology
coagla@rit.edu

ABSTRACT

Online learning has gained increased popularity in recent years. However, with online learning, teacher observation and intervention is lost, creating a need for technologically observable characteristics that can compensate for this limitation. The present study used a wide array of sensing mechanisms including eye tracking, galvanic skin response (GSR) recording, facial expression analysis, and summary note-taking to monitor participants while they watched and recalled an online video lecture. We explored the link between these human-elicited responses and learning outcomes as measured by quiz questions. Results revealed GSR to be the best indicator of the challenge level of the lecture material. Yet, eye tracking and GSR remain difficult to capture when monitoring online learning as the requirement to remain still impacts natural behavior and leads to more stoic and unexpressive faces. Continued work on methods ensuring naturalistic capture are critical for broadening the use of sensor technology in online learning, as are ways to fuse these data with other input, such as structured and unstructured data from peer-to-peer or student-teacher interactions.

ACM Classification Keywords

K.3.1 Computers and Education: Distance Learning.; H.5.2 User Interfaces: Evaluation/methodology.

SmartLearn'17, March 13 2017, Limassol, Cyprus
© 2017 ACM. ISBN 978-1-4503-4904-8/17/03…$15.00
DOI: http://dx.doi.org/10.1145/3038535.3038536

Author Keywords

Online learning; multimodal sensing; sensing methodology; learning outcomes.

INTRODUCTION

Online learning enrollments in the United States continue to increase annually and currently outpace the annual growth of enrollment in traditional higher education courses [1]. Online courses allow for a wider reach, including students who may not otherwise have access to these classes such as students from different geographic areas or individuals who cannot travel to the university, and it also offers convenience for non-traditional students with busy schedules. One limitation of online learning is that teacher observation and intervention is lost, creating a less individualized learning experience. This creates a need for the exploration of technologically observable characteristics that can compensate for this limitation.

Recent studies have shown that psychophysiological responses of the user can be used to create flexible e-learning contexts, and *adaptive e-learning* is the term used to describe the ability of such an environment to recognize and adjust to these reactions [15]. This allows for a more customized learning experience in which the system can tailor the instruction to meet the needs of each individual. The notion of customized curriculum has become more idealized in the education field in which educators are moving away from the "one size fits all" approach and are more cognizant of differing learning styles, rates, and strengths/weaknesses. Adaptive e-learning allows for real-time recognition and customization of curriculum, creating a learning context that can both anticipate and change to meet the needs of the individual learner.

In this paper, we report on a study that attempts to identify technologically observable characteristics that relate to learn-

ing outcomes during online lectures. Educational technology can build upon such observable cues to identify problem areas during the learning process towards differentiating learned and unlearned material in real-time. There are multiple uses for such cues gathered during the learning process. For instance, they can help instructors improve their courses by indicting pedagogically problematic material, or help reduce self-assessment time spent by students to identify unlearned material that has yet to be mastered.

With science, technology, engineering, and mathematics (STEM) education as a national priority, it is important to explore how online learning of STEM content can be improved. Specifically, we used computer graphics as a use case to investigate online learning of technical content. We chose this area as several co-authors are university educators with expertise in this domain.

BACKGROUND

Attention and engagement are vaguely defined constructs understood subjectively and widely variable operationally. It is debated in the literature what method to use for monitoring attention and engagement. Ramesh et al. (2014) proposed that engagement can be explored in terms of the interaction between behavior-based, language-based, and other characteristics [13]. One study used electrocardiogram (ECG), heart flux (HF), and electroencephalogram (EEG) data to categorize attention and non-attention [3]. Another approach measured engagement of TV viewers using video cameras to track head and facial features [7]. Further, researchers have also used webcams to monitor engagement levels by analyzing facial expressions [17]. In addition, there is evidence for a link between gaze behavior and learning [16]. We conducted an experiment to investigate feasible monitoring methods that link to learning in a setting involving online video lecturing.

The present study uses eye tracking, galvanic skin response (GSR), facial expression analysis, and summary note-taking characteristics in an attempt to simulate a naturalistic environment more representative of a common online learning setting using only non-invasive sensors. Although these sensors are not currently widely accessible in online learning settings, with the progression of technological accessibility, it is likely that commercial laptops will include eye-tracking capabilities in addition to a webcam and wearable fitness technology may soon include GSR data as well.

Although GSR has been shown to provide a measure of cognitive load, the application of this is often a challenge. Haapalainen et al. (2010) observed such challenges in attempting to use psychophysiological metrics to gauge cognitive load [6]. They used a GSR armband which they noted may not have been sufficiently sensitive to provide an accurate indication of cognitive load.

Our data collection experiment attempted to simulate a naturalistic environment more representative of a common online learning scenario. While subjects watched a video lecture, non-invasive sensors monitored the study subjects via eye tracking, GSR, and facial expression capture. After viewing the lecture, the subjects engaged in summary note-taking followed by a quiz. The study is operating under the assumption that answer correctness on the quiz suggests learning. As opposed to linking these measures to attention and engagement and then attention and engagement to learning, the present study attempts to link these measures directly to learning, to investigate the link between these technologically observable psychophysiological responses and learning outcomes.

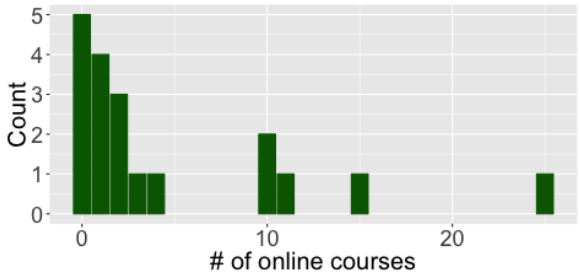

Figure 1. Histogram showing participants' level of experience with online courses.

The current study is operating under the assumption that answer correctness suggests learning. We hypothesized that answer correctness can be linked to observable human reactions occurring during the sections of the video lecture corresponding to the material covered by the quiz question. We further hypothesized that human biofeedback response can point to the challenge level of the educational content. The present study attempts to investigate the link between these technologically observable psychophysiological responses and learning outcomes.

METHOD

Participants

Participants consisted of 21 college students. They were paid $20 for their involvement. All participants reported having normal or corrected-to-normal vision, as well as having no known learning or attention disorders. Their age ranged from 18 to 65 with a mean age of 24.1. Gender was evenly distributed: 11 participants were men and 10 were women.

Participants self-reported the number of online courses they had previously taken. As shown in Figure 1, many participants had taken none or few classes online, however, a few participants did have significant experience with online learning. Data from two participants were excluded from further analysis as they did not correctly follow instructions.

Stimuli and Apparatus

The experimental setup is illustrated in Figure 2. Stimuli were presented using SMI Experiment Center 3.5.169 on a 22 inch LCD monitor with a resolution of 1680 X 1050 pixels with a refresh rate of 60 Hz. Participants were seated 60-80cm from the screen. A Shimmer3 GSR sensor collecting data at a rate of approximately 50 Hz was placed on the participant's non-dominant hand. Eye movements were monitored using a SensoMotoric Instruments iView X Remote Eye Tracking Device desktop-mounted eye tracker which operated at 60Hz, with a manufacturer-reported gaze position accuracy $< 0.5^{\circ}$.

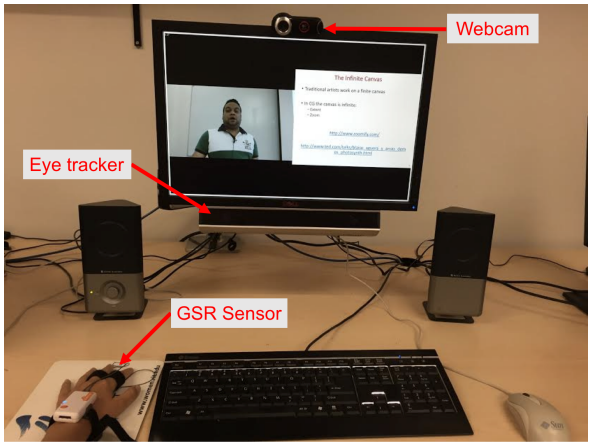

Figure 2. Experiment set up with webcam mounted above and eye tracker mounted below the monitor. The GSR sensor is attached to the hand of the participant.

A Logitech Webcam Pro 9000 was mounted on top of the monitor to collect video of participants in order to analyze facial expressions. The OpenFace software toolkit was used to capture and quantify 68 different facial landmark locations [2]. Additional facial analysis was performed using Affectiva's Affdex SDK[1] which reports on the emotions conveyed and also includes measures of attention and engagement.

Procedure

The GSR sensor was attached to the participant's non-dominant hand. Participants completed a demographic survey including questions about academic major, previous experience with computer graphics, previous experience with online learning etc., using an online survey on a desktop computer. Calibration of the GSR sensor was performed during the demographic survey.

After the survey, participants were seated in front of the eye tracker with their non-dominant hand (with the GSR sensor) flat on the table. They were instructed to refrain from moving that hand during the lecture video. A nine point calibration was used to calibrate the eye tracker, followed by a four point validation using manufacturer-provided techniques. Next, participants viewed a lecture video, lasting 16 minutes and 34 seconds. This video consisted of two panels presented side by side. The video of the instructor was presented on the left side of the screen and corresponding lecture slides containing relevant content were presented on the right. When the lecture video was complete, participants were given five minutes to type a summary of important points presented in the video. A countdown timer on the screen helped participants keep track of the remaining time. Then, participants took a 15 question quiz without a completion time limit. Participants also rated their confidence in their answer for each question.

Lastly, participants completed a post-experiment survey about their perceived experiences during the experiment including questions such as how well they feel they paid attention, how engaged in the material they felt, how interesting they found the material, how well the lecture prepared them for the quiz, and how difficult they found the material.

[1] www.affectiva.com

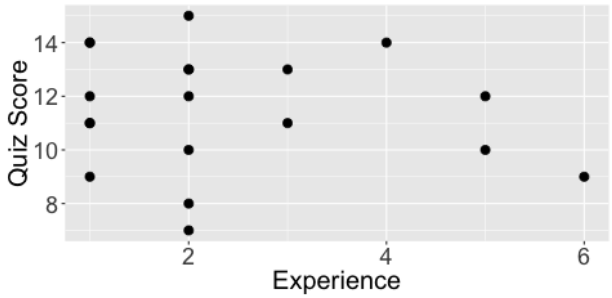

Figure 3. Participants self-reported previous experience with computer graphics on a 7-point Likert scale (1 = none, 7 = expert) and quiz scores were not strongly correlated [r(17) = -0.143, p = 0.559], which indicates that quiz score was not influenced by previous knowledge of lecture material.

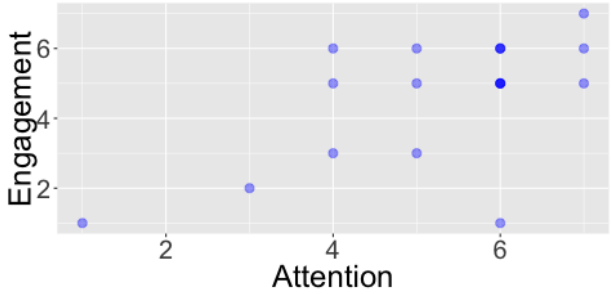

Figure 4. Self-reported attention during the lecture was correlated with self-reported engagement in the lecture material [r(17) = 0.615, p = 0.005]. Both used a 7-point scale.

RESULTS

Data from two participants were removed due to not following instructions. Data from the remaining 19 participants showed no strong correlation [r(17) = -0.143, p = 0.559] between self-reported previous experience with computer graphics and quiz score; see Figure 3. Therefore, all participants were analyzed together regardless of previous experience. Results also showed that the amount of time spent looking at the instructor (or the slides) did not have a significant impact on the number of correct and incorrect responses on the quiz.

As shown in Figure 4, participants' self-reported level of attention and self-reported level of engagement throughout the lecture were correlated [r(17) = 0.615, p = 0.005].

Furthermore, self-reported interest in lecture material was correlated with self-reported level of attention [r(17) = 0.739, p < 0.001] as well as with self-reported level of engagement [r(17) = 0.876, p < 0.001]; see Figures 5 and 6.

Participant responses to the survey question asking how well they felt the lecture prepared them for the quiz were correlated with quiz score [r(17) = 0.550, p = 0.015]. This suggests that participants had some intuition of their quiz performance and how well they remembered lecture material.

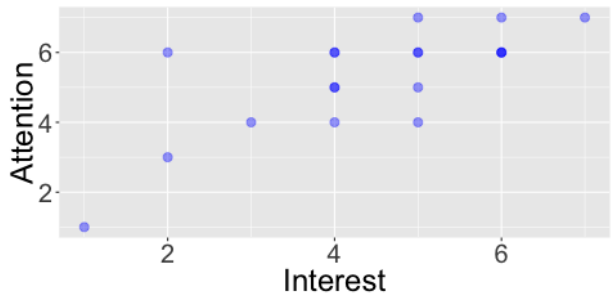

Figure 5. Self-reported interest in lecture material was correlated with self-reported attention during the lecture [r(17) = 0.739, p < 0.001].

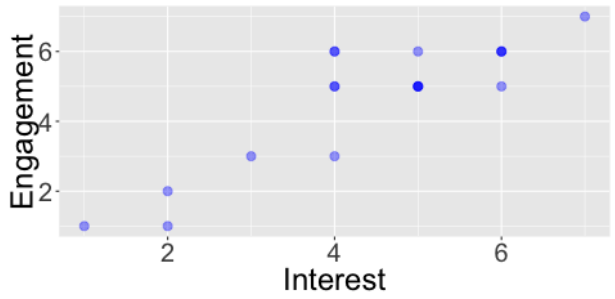

Figure 6. Self-reported interest in lecture material was correlated with self-reported engagement during the lecture [r(17) = 0.876, p < 0.001].

In order to remove noise from the GSR data, a median filter was used. This data was then normalized for comparison between participants and split into sections corresponding to quiz questions. These sections consist of the time interval in which the question material was covered in the lecture video. Time intervals ranged from 5 to 69 seconds.

Given the clustering of quiz scores on the high end, as shown in Figure 8, it was also useful to investigate challenge level for individual questions at the extremes in terms of performance level. We examined two questions that represented the extremes in terms of quiz score accuracy. They were:

- *Q3: Complete the following sentence: The part of the canvas that is drawn on the display is called the.....* Answer: Clip window. Correctly answered by all participants.
- *Q12: In what coordinate system are the vertices for graphics primitives specified?* Answer: World coordinates. Correctly answered by only 5 participants.

GSR measurements are reported in kOhms - a measure of resistance. Lower resistance indicates higher skin conductance (sweat level). Range was chosen as a GSR measurement because it indicates changes across a specific period of time and it can eliminate influences of individual differences in resting levels or drift that may occur in sensor data. In order to investigate whether the GSR range differed by challenge level, a paired samples t-test was conducted. As shown in Figure 7, the range of the GSR data during the section of the video that corresponded to the question that most people got correct

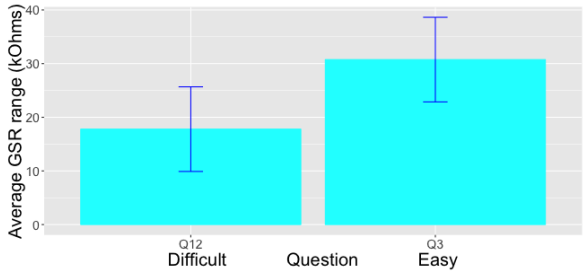

Figure 7. GSR relationship with question difficulty. Question 3 was answered correctly by all participants. Question 12 was answered correctly by only 5 participants.

(easy material) was significantly higher [t(18) = 2.442, p = 0.025] than for the material that corresponded to the question that most people got wrong (difficult material). This suggests that cognitive load was higher during more difficult content material.

DISCUSSION

General Discussion

Results revealed a significant difference in GSR range between easy and difficult content material. This suggests that cognitive load was higher during more difficult material. This is consistent with a prior study that found GSR to be an accurate measure of cognitive load level for different levels of task difficulty [12]. Furthermore, another study used GSR and the number of blinks to classify cognitive load [11]. Our experiment found no significant relationship between number of blinks and performance on the quiz; however, further research should investigate whether number of blinks can provide additional information when combined with GSR data.

In addition to the implications for online education, these findings can be applied to assistive technology for students with learning disabilities. For instance, measures of attention can be useful in aiding students with attention disorders. Due to the prevalence of attention deficit hyperactivity disorder (ADHD) as the second most common learning disability, the advantage is important. Assistive technology has been shown to not only improve learning outcomes, but also boost self-confidence and reduce stress associated with learning difficulties [5, 10, 14, 8, 9].

Lessons Learned from Evolving Experimental Design

Facial analysis revealed very little variation of facial expressions over the course of the lecture, which could have been related to asking participants to keep still in order to obtain accurate eye tracking and minimize GSR sensor noise. We hypothesized that actively trying to remain still contributed to the stoic and unexpressive faces that were observed. In order to explore this, six additional participants were included with two changes to the methodology. The experimental design changes can be characterized as follows:

1. Twenty-one participants completed the experiment as described in Procedure section above.

2. Four new participants were not asked to keep still. Also, researchers left the room during the lecture in order to further imitate a natural online learning setting and eliminate the influence of being watched. These measures may have contributed to a more comfortable setting, encouraging more natural facial expressions.
3. Two final participants further did not wear the GSR sensor in addition to the instruction change.

First, the latter two groups of participants seemed to not only move more but also seemed to be generally more expressive in the face.

Second, the last group seemed to move the most; both their entire body as well as their face.

Limitations

The GSR sensor is sensitive to hand movement, thus taking notes during the lecture would interfere with the GSR data. Therefore, the experimental design was modified to include summary note-taking after the lecture rather than during. This deviates from the natural habit of online learning. Upon initial inspection, no significant relationships were uncovered between simple metrics (e.g., number of tokens) for summaries and quiz score.

The extended sample size was of modest size. Future research should continue to explore the impact of methodological distinctions in order to make definitive judgments.

Because performance on the quiz had no impact on participants, motivation to learn was not present as in typical online learning in which students are graded on performance. Also, the goal in education is for long-term retention of information rather than a quiz directly after material is learned. Future research should involve a longitudinal dimension to investigate retention of information and learning patterns over time.

In addition to the quiz only being 15 questions which is atypical for most courses, the distribution of scores was skewed. Most people did well on the quiz, as seen in Figure 8.

Another limitation of studying online learning in a laboratory is the likelihood of participants acting differently because they know they are being watched. This Hawthorne effect makes it difficult to generalize laboratory findings to online learning as a whole [4].

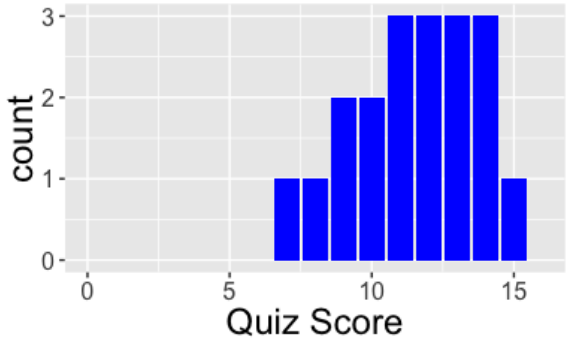

Figure 8. Quiz scores were skewed such that most participants answered more than 10 of the 15 questions correctly.

Due to the fact that eye tracking data was applied only to understand which side of the screen the viewer was attending to, it was not beneficial to consider links between eye tracking and GSR data.

CONCLUSION

We reported on a study that made the following contributions:

1. It presented a new methodology for studying online learning.
2. It identified issues in using sensor technology to study online learning.
3. It proposed suggestions for experimental design modifications to improve the use of sensor technology in studying online learning.

Eye tracking and GSR remain a challenge to capture when monitoring online learning as the requirement to remain still impacts natural behavior and can lead to unexpressiveness.

Continued work on methods ensuring naturalistic capture are important and will help broaden the use of sensor technology in online learning.

We also see a need for studying ways to fuse learner sensor data with other input from online educational settings, such as structured and unstructured data from peer-to-peer or student-teacher interactions.

Acknowledgements

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. Elaine Allen, Jeff Seaman, Russell Poulin, and Terri Taylor Straut. 2016. Online Report Card: Tracking Online Education in the United States. (Feb 2016).
2. Tadas Baltru, Peter Robinson, Louis-Philippe Morency, and others. 2016. OpenFace: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1–10.
3. Ashwin Belle, Rosalyn Hobson, and Kayvan Najarian. 2011. A physiological signal processing system for optimal engagement and attention detection. In *Bioinformatics and Biomedicine Workshops (BIBMW), 2011 IEEE International Conference on*. IEEE, 555–561.
4. Jack P Campbell, Vada A Maxey, and William A Watson. 1995. Hawthorne effect: implications for prehospital research. *Annals of emergency medicine* 26, 5 (1995), 590–594.
5. Kim K Floyd and Sharon L Judge. 2012. The Efficacy of Assistive Technology on Reading Comprehension for

Postsecondary Students with Learning Disabilities. *Assistive Technology Outcomes and Benefits* 8, 1 (2012), 48–64.

6. Eija Haapalainen, SeungJun Kim, Jodi F Forlizzi, and Anind K Dey. 2010. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 301–310.
7. Jaime Hernandez, Zicheng Liu, Geoff Hulten, Dave DeBarr, Kyle Krum, and Zhengyou Zhang. 2013. Measuring the engagement level of TV viewers. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 1–7.
8. Helena Lindstedt and Õie Umb-Carlsson. 2013. Cognitive assistive technology and professional support in everyday life for adults with ADHD. *Disability and Rehabilitation: Assistive Technology* 8, 5 (2013), 402–408.
9. Anna Milani, Maria Luisa Lorusso, and Massimo Molteni. 2010. The effects of audiobooks on the psychosocial adjustment of pre-adolescents and adolescents with dyslexia. *Dyslexia* 16, 1 (2010), 87–97.
10. Charlotte A Mull and Patricia L Sitlington. 2003. The Role of Technology in the Transition to Postsecondary Education of Students with Learning Disabilities A Review of the Literature. *The Journal of Special Education* 37, 1 (2003), 26–32.
11. Nargess Nourbakhsh, Yang Wang, and Fang Chen. 2013. GSR and blink features for cognitive load classification. In *IFIP Conference on Human-Computer Interaction*. Springer, 159–166.
12. Nargess Nourbakhsh, Yang Wang, Fang Chen, and Rafael A Calvo. 2012. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*. ACM, 420–423.
13. Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daumé III, and Lise Getoor. 2014. Uncovering hidden engagement patterns for predicting learner performance in MOOCs. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 157–158.
14. Marshall H Raskind and Eleanor L Higgins. 1998. Assistive technology for postsecondary students with learning disabilities an overview. *Journal of Learning Disabilities* 31, 1 (1998), 27–40.
15. Jonathan L Rosch and Jennifer J Vogel-Walcutt. 2013. A review of eye-tracking applications as tools for training. *Cognition, technology & work* 15, 3 (2013), 313–327.
16. Hsiao-Ching She and Yi-Zen Chen. 2009. The impact of multimedia effect on science learning: Evidence from eye movements. *Computers & Education* 53, 4 (2009), 1297–1307.
17. Jacob Whitehill, Zewelanji Serpell, Yi-Ching Lin, Aysha Foster, and Javier R Movellan. 2014. The faces of engagement: Automatic recognition of student engagement from facial expressions. *Affective Computing, IEEE Transactions on* 5, 1 (2014), 86–98.