Identifying Student Struggle by Analyzing Facial Movement During Asynchronous Video Lecture Viewing: Towards an Automated Tool to Support Instructors

Conference or Workshop Item

The final authenticated publication will be available online in LNCS (Springer),
Proc. of the 23rd Int'l Conf. on Artificial Intelligence in Education (AIED 2022),
eds. Rodrigo, Matsuda, Cristea & Dimitrova

Identifying student struggle by analyzing facial movement during asynchronous video lecture viewing: Towards an automated tool to support instructors

Adam Linson[1[0000-0001-9602-6536]], Yucheng Xu[2[0000-0001-9023-0974]],
Andrea R. English[3[0000-0002-1351-0507]], and Robert B. Fisher[2[0000-0001-6860-9371]]

[1] School of Computing & Communications, Open University, Edinburgh EH3 7QJ, UK
[2] School of Informatics, University of Edinburgh, Edinburgh EH8 9LE, UK
[3] Moray House School of Education & Sport, University of Edinburgh,
Edinburgh EH8 8AQ, UK
adam.linson@open.ac.uk
{yucheng.xu, andrea.english, r.b.fisher}@ed.ac.uk

Abstract. The widespread shift in higher education (HE) from in-person instruction to pre-recorded video lectures means that many instructors have lost access to real-time student feedback for the duration of any given lecture (a 'sea of faces' that express struggle, comprehension, etc.). We hypothesized that this feedback could be partially restored by analyzing student facial movement data gathered during recorded lecture viewing and visualizing it on a common lecture timeline. Our approach builds on computer vision research on engagement and affect in facial expression, and education research on student struggle. Here, we focus on individual student struggle (the effortful attempt to grasp new concepts and ideas) and its group-level visualization as student feedback to support human instructors. Research suggests that instructor supported student struggle can help students develop conceptual understanding, while unsupported struggle can lead to disengagement. Studies of online learning in higher education found that when students struggle with recorded video lecture content, questions and confusion often remain unreported and thus unsupported by instructors. In a pilot study, we sought to identify group-level student struggle by analyzing individual student facial movement during asynchronous video lecture viewing and mapping cohort data to annotated lecture segments (e.g. when a new concept is introduced). We gathered real-time webcam data of 10 student participants and their self-paced intermittent click feedback on personal struggle state, along with retrospective self-reports. We analyzed participant video with computer vision techniques to identify facial movement and correlated the data with independent human observer inferences about struggle-related states. We plotted all participants' data (computer vision analysis, self-report, observer annotation) along the lecture timeline. The visualization exposed group-level struggle patterns in relation to lecture content, which could help instructors identify content areas where students need additional support, e.g. through student-centered interventions or lecture revisions.

Keywords: video analysis, data visualization, facial expression, student struggle, reflective teaching, human-centered computing.

1 Introduction

1.1 Project description

When instructors in higher education (HE) deliver in-person lectures, they can access a form of real-time student feedback simply by ‘reading the room’. That is, by scanning their students’ nonverbal cues, instructors can make inferences on the assumption that the ‘sea of faces’ may express various cognitive and affective states relevant to group instruction, such as struggle and comprehension. However, the widespread shift in HE from in-person instruction to pre-recorded video lectures means that many instructors have lost access to this form of feedback. We sought to test the hypothesis that this feedback could be partially restored by analyzing student facial movement data gathered during recorded lecture viewing and visualizing it on a common lecture timeline. To this end, we developed a software prototype, PUZZLED, inspired by current research in education, computer vision, and data visualization. The system was designed to identify when and in what respect students are struggling, and to analyze and visualize results to provide insights to human instructors.

In this report, we describe the prototype design and piloting in a small-scale exploratory and feasibility study (N=10), funded by the University of Edinburgh Regional Skills program. As a key contribution of this paper, the study showed that visual evidence can be extracted from video of student facial movement (e.g. eye gaze aversion) that aligns temporally with aspects of a corresponding viewed lecture video. That is, moments in the lecture video containing conceptually challenging content, omitted background information, or other difficulties posed to student viewers (e.g. blurry text) led to measurable student facial movements (e.g. expression changes). These contrasted with student facial movement data captured during introductory or otherwise straightforward segments of the lecture video.

A second key contribution of this paper relates to the visualization of the data. Current data visualization interfaces for time-based models are primarily anchored in either absolute time (e.g. audience feedback during an in-person lecture, which uses global timestamps), or abstract task time (e.g. time solving a problem, which is averaged across individuals). Here, we integrate both by using global timestamps indexed against a common reference timeline (feedback from asynchronous viewings of a video lecture). This timeline is used to visualize group-level feedback from a student cohort at a granular level. In this application, the visualization can inform instructors about how student struggle relates to specific segments of the lecture.

1.2 Research context in education

Research on student learning, including in HE, suggests that struggle – the effortful attempt to grasp new concepts and ideas – is important to the learning process [1–5]. More precisely, there is differentiation between ‘unproductive’ struggle, such as unresolved confusion that leads to task disengagement, and ‘productive’ struggle, as when a challenging task is cognitively engaging. Productive struggle is important for the development of students’ critical thinking and deep understanding [6–9, see also 10]. It can also indicate an appropriate level of challenge that maintains learner engagement, a key factor in HE student retention [11, 12]. In addition, research has found that equity-oriented teaching both successfully challenges all students to engage in struggle, and also supports all students through struggle [3, 5].

In online HE instruction, however, research indicates that when students struggle with recorded video lecture content (when a question or confusion arises), it often remains unreported, and thus unsupported by instructors, leading to lower student engagement and greater attrition, relative to in-person courses [13, 14]. A growing number of students are negatively impacted by this problem, given that recorded lectures are now a "mainstream" part of online HE provision [15]. 94% of UK universities make recorded lectures available to students year-round, in part "as a catalyst for inclusivity" [16]. In addition, European HE reform has promoted the expansion of e-Learning provision, including online recorded lectures (75% of universities), as a means of "widening access" to HE, while acknowledging that quality "teacher support" is needed to maintain learning with "critical thinking" and "deep understanding" in online contexts [17, 18].

In principle, instructors can only effectively support student struggle in online courses of students who self-report. This poses a serious problem, as the move to online HE places more demands on students to make their struggle known to instructors. Yet students who do not have knowledge of the subject area, and/or are not skilled in self-regulated learning, are less likely to self-assess and self-report their difficulties [19]. Additional research suggests that self-reporting may privilege the subset of students who are vocal about their struggles [20]. Thus, many students' learning needs remain unsupported in relation to recorded lecture content online [13]. Overall, this situation highlights the tension between the benefits and drawbacks of online learning and its constitutive technologies [21].

1.3 Research context in informatics

A range of literature has used estimations of student engagement or affect, primarily using well-developed standards for assessing facial action units, along with self- and/or observer-coded higher-level states correlated with quantitative face and postural data [10, 22–25]. We take inspiration from these methods and bellwethers of feasibility. To our knowledge, other studies using computer vision related analysis of student faces involve participants interacting with virtual tutors or games, but not pre-recorded lecture videos with human instructors, as we have done.[1]

The choice of educational material 'delivery mode' in our study (i.e. lecture videos with a human instructor) can be understood in relation to our primary aim of supporting human instructors, as compared to the aims of similar studies that seek to investigate learning itself or to enhance virtual tutors. For our purposes, rather than seeking to determine whether a student is 'really' struggling, we instead focus on providing inferences about the student cohort that could motivate an instructor to re-evaluate aspects of their lecture (e.g. content, delivery, pre-requisites, etc.) or provide corresponding non-lecture-based student support (e.g. forums, further readings, etc.).

---

[1] See our source video lectures (1a) and (1b), which played sequentially without interruption:
(1a) `https://media.ed.ac.uk/media/1_yd7krol3`
(1b) `https://media.ed.ac.uk/media/1_ktjie97s`

In educational technology research that does not involve face data, there are related applications of data visualization that aim to provide insights to instructors. We draw from an example that uses a horizontal time axis and vertical axis of individual students to surface salient patterns within a vast and complex student cohort (e.g. in a MOOC) [13]. We also build directly on a concept for annotating synchronously viewed content. The latter implementation uses crowd-sourced structured tagging (live audience feedback) that is visualized on a horizontal time axis, aligned with a video replay interface [21, 26].

Outside of education, computer vision research has sought to measure affective indicators in participant video under 'real world' capture conditions. In an approach related to ours, independent algorithms for classifying events in multiple low-level data streams (face, posture, etc.) are followed by fusion and high-level affect classification [27]. Our approach also relates to research on stress using multimodal biosignal information extracted from videos [28], in that we aspired to improve the accuracy of student struggle estimation by correlating low-level objective measures (e.g. gaze direction) with real-time and retrospective self-report. (Similar classification techniques have been used in education research with sophisticated instruments in controlled environments [10], in contrast to our use of webcams in everyday locations.)

1.4 Research context on facial expressions and eye gaze

Overwhelmingly, research on eye gaze centers on visual content fixation, to understand how people look at text, mathematical formulas, videos, interfaces, etc. In contrast, a smaller body of research considers eye gaze as an indicator of affective or cognitive state, primary in terms of whether or not gaze is averted. In computer vision affect detection, gaze aversion metrics have been used to improve automated classification of emotional state [29]. In developmental psychology, research has confirmed that young children judge faces to be engaged in thinking when viewing photos with subjects averting their gaze [30]. (The study uses a similar approach to ground truth as the present one, in that cognitive-affective states are inferred from images by independent raters.) Educational psychology research has also identified gaze aversion in young children as indicative of thinking; this is suggested as a cue to instructors who must gauge how soon to expect a verbal response following a question [31]. Research on inferring mental states from gaze aversion supports the idea that disengaging from perceptual demands (such as looking at an instructor) facilitates thinking [32].

2 Methodology

2.1 Participants and study design

Following ethics approval, we recruited university students (N=10) from a UK postgraduate degree program in informatics. We offered a £10 voucher in exchange for roughly 45 minutes of participation. 7 female and 3 male volunteers responded with informed consent (demographic data was self-reported in free text). 90% of ages fell in the range of 22-29, plus one 37-year-old participant, and ethnicity was entered as South Korean (1), Chinese (2), Indian (2), or white/Caucasian (5). 70% of participants had subject matter experience ranging from 1-4 years; a further two had 0.25 years, and one had 8 years. 80% selected a multiple-choice answer to report having a "little" experience learning through video lectures (20% chose a "lot", and none chose "it's new to me", the sole remaining option).

We invited the participants via an emailed link (with an anonymous unique identifier) to start and complete their session in a single sitting. The web interface provided instructions for positioning the webcam using a live feed from their own device ("ensure that your face is in the middle of the image"). No webcam video was displayed during recorded lecture viewing. A short 3-minute practice session was offered with a separate lecture not used for the study, to provide experience viewing and using the click feedback interface, which offered the following button options and instructions:

- "Feels easy" - *click when you think the content is easy for you to understand*
- "Feels challenging" - *click when you are able to follow the content, but it is not too easy for you*
- "I'm lost" - *click when you cannot follow at all what the lecturer is talking about*

The click feedback interface could be repositioned from left-to-right, while remaining in a fixed row beneath the lecture video (Fig. 1). A further instruction stated "*When viewing the lecture, as your feeling changes, continue to click these buttons, to show how you feel*". Following the practice lecture video, the participants were shown two different lecture videos (ca. 3 min + ca. 6 min = total ca. 10 min) from an existing MSc computer vision course taught by the last author. The lecture videos played back-to-back automatically, while participant webcam and click feedback was captured in real-time during viewing. Participants were not informed that the lectures were not originally intended for consecutive viewing, as the second lecture summarized a previous lecture not shown to participants. This provided an experimental control indicator of 'challenged' reactions not related to challenging subject matter. Upon completion, an 'exit survey' was provided for retrospective self-report on the lecture material and feedback on the study interface design.

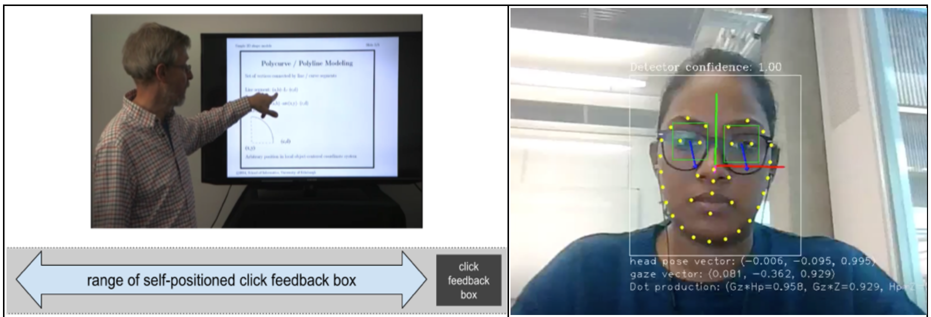
There were some problems with both the interface code (cross-browser compatibility) and comprehensibility of the instructions to participants. The data collection succeeded as intended for participants 1-5 (50%). Participant 6 successfully provided webcam data without click feedback, and we did not request a repeat session. For participants 7-10, we requested a repeat session; 7-9 succeeded in their second session and participant 10 succeeded in a third session. Thus, data on 40% of participants did not relate to their first viewing of the lecture materials. However, for an exploratory pilot and feasibility study, this was not a serious obstacle to completing its objectives.

Our hypothesis was that granular student cohort feedback on video lecture materials could be provided to instructors by students who viewed videos asynchronously. We sought to analyze student facial movement data gathered during recorded lecture viewing, and to visualize the cohort feedback on a common lecture timeline. Our expectation was that computer vision techniques could identify positional changes in head orientation and eye gaze direction, along with facial expression changes indicating struggle, and that these could be correlated with independent observations inferring struggle-related states, to provide a pathway to increasingly automated recognition. With all participants' data (computer vision analysis, self-report, observer annotation) plotted along the lecture timeline, evidence for our hypothesis would be found if group-level patterns emerged in relation to segments of lecture content. A toolchain of data processing and visualization steps would then provide an overview of the student feedback in relation to the lecture material. This could help instructors identify content that corresponded to a potential requirement for further student support.

2.2 Data analysis techniques

Facial feature analysis was performed with OpenVINO, using models adapted from Open Model Zoo.[2] Pre-existing trained models for facial landmark localization, face detection, gaze estimation, head pose estimation, and eyes open-or-closed state were integrated with PUZZLED, to analyze participant video frames for eye gaze direction and head orientation. While further analysis is needed to uncover potentially relevant patterns in head orientation, we found a notable correlation between eye gaze direction (aversion) and possible student struggle (see context in §1.4 and results in §3).

We normalized and smoothed out eye gaze direction to indicate a baseline bandwidth, above which was classified as upward gaze and below as downward. We then filtered out gaze direction data within the baseline band, treating it as direct video viewing (including left-right patterns of reading on-screen text). We also filtered out downward gaze data, since our self-report click feedback interface was below the video (Fig. 1), and participants appeared to be saccading to the interface when contemplating or performing click feedback. The remaining upward gaze data was plotted against real-time and retrospective self-report and observer annotations. Notably, we did not treat it as a universally valid measure of struggle (see below).



Lecture video player & feedback interface | Webcam frame & computer annotation

Fig. 1. Frame layout of lecture video and self-report click feedback interface (l); webcam frame of face with features detected and gaze direction (r). A planned production version would not remotely transmit facial images, only non-visual data from webcam analyses performed locally.

Individual classifications

Our approach is inclusive, in that it does not depend on generalizations across ethnicity, culture, gender, etc., and can even remain robust with respect to individual difference. For example, for participant '052', upward gaze likely indicates struggle,

[2] `https://docs.openvino.ai/latest/omz_demos.html`

whereas for ‘409’, it does not. This classification for ‘052’ is in part established through correlations with the other within-subject data points (e.g. retrospective self-report that the first lecture was challenging and the second was not, and real-time self-report of challenge coinciding with independent observer annotations of challenge).

Cohort classifications

Similar to related work [33], which uses a percentage of frames with a target classification (‘stress’) in a time window to generalize the classification to the window (stress segment), we use a relative threshold of events with a target classification (‘struggle’) within and across students to identify a relevant lecture segment in which students struggled. In our case, we allow for sparse target classification events in different data streams (Fig. 2): individual students (grey horizontal bars) self-reports of being challenged (blue stars), computer-detected upward gaze aversion events (cyan dots), and independent annotations of inferred challenge (red circles and dots), aggregated across students for each bounded lecture segment (within green vertical lines).

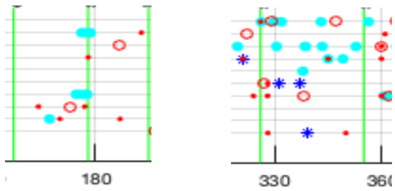


Fig. 2. Example of two plotted segments, a (left), b (right).

In the above illustration, relatively sparse struggle events do not meet the threshold within the ca. 40 second window depicted (Fig. 2a, two lecture segments, 155s-195s), while for a similar time window (Fig. 2b, one lecture segment, 320s-360s), relatively dense struggle events generalize to a segment classification of cohort struggle.

Independent observations were done by three members of the research team (AE, RF, AL), by viewing all the participant videos (without the lecture video), and using the same click feedback interface as the participants. A further annotation option of “bored” was added to the interface to allow for distinguishing between comprehension and disengagement, and thereby to increase robustness in correlating data points for the ‘feels easy’ option. A manual segmentation and annotation of the lecture was performed based solely on its content, independent of any participant or observer data (e.g. ‘3:00 - 3:25, introduction of new term’; ‘3:26 – 4:10, description of applying a technique’). Finally, data from annotations, computer vision, real-time and retrospective self-reports, and independent observations was plotted on the lecture timeline.

3 Results

Given the structure of our study, there are two relevant sets of results. The first set of results relates to the correspondence between computer vision data (eye gaze direction) and manual annotation data (including student-self report and observer annotations). These results indicate that the computer vision algorithm shows promise, and could be developed in future work to increase automated analysis and tagging.

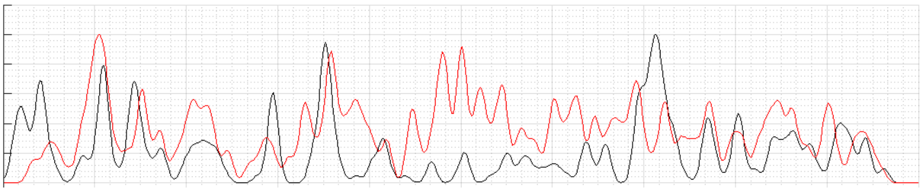


Fig. 3. The horizontal axis represents the lecture timeline from start time to end time. The black line is a normalized and smoothed plot of all students' vertical gaze direction, where the lowest points are nearest to the median gaze bandwidth (i.e. looking towards any point left-to-right on the horizon), and the highest points indicate peak vertical upward gaze. The red line is a normalized and smoothed density plot of all struggle annotations (student self-report and three independent observers). Agreement among annotators and self-reports are seen in the red peaks, and typically track upward gaze (black peaks) as seen in the confluences that occur throughout, apart from a notable red/black divergence in a middle segment of the timeline (x-axis). This period of divergence hypothetically corresponds to the increased level of click feedback during that time window, which appears to cause a downward gaze towards the click interface below the video (see Fig. 1). In future work, we will eliminate real-time student self-report and replace it with retrospective self-report during a second video viewing, to mitigate the divided attention (and gaze patterns) between initial lecture video viewing and concurrent feedback reporting.

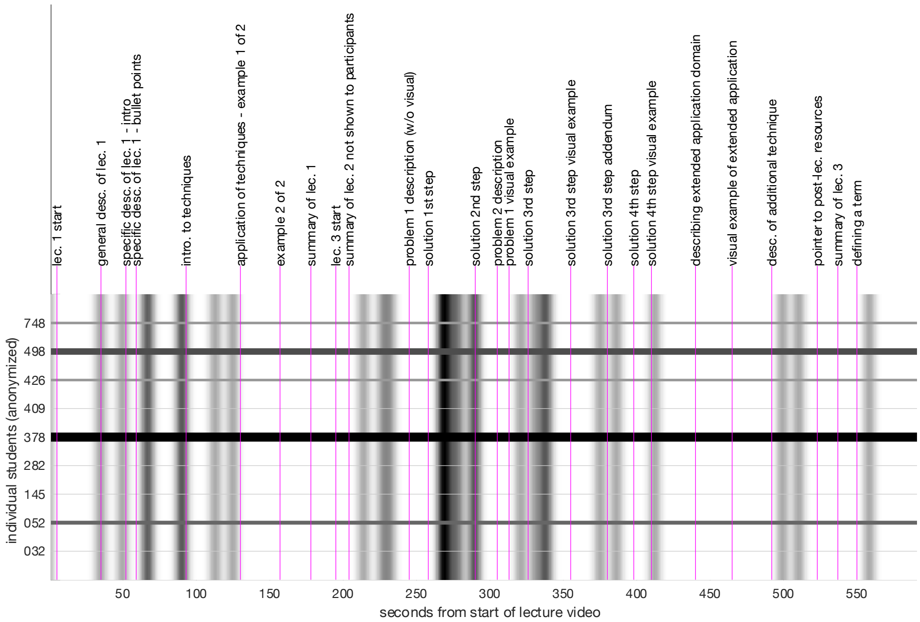


Fig. 4. Prototype of instructor data visualization interface (darker/wider bars = more self-reported student struggle). Vertical bars are density plots of the cohort. For example, at around 260s, in the lecture segment labelled 'solution 1st step', a greater proportion of students reported struggle than elsewhere. This would suggest to the instructor that the lecture video segment could be re-examined, to understand if the struggle was part of the pedagogical design (e.g. providing 'food for thought'), or alternatively, if students need further support, and how to support them. Support could include student-centered interventions (e.g. adding a forum or

group study session, linking to additional resources), or revising the lecture content as needed, by clarifying a term, adding a visual example, etc. Horizontal bars indicate frequency of report. For example, student '378' reported 'feels challenging' more frequently than all other students. The range of horizontal bars gives the instructor an overview of the student cohort, which in this figure, suggests a large proportion of the students are well-matched to the lecture content. Empirical survey data of the students' retrospective self-reports bear out this reading.

With respect to eye gaze direction, our results suggest initial evidence for a hypothesis that intermittent periods of upward gaze aversion could be related to a 'struggle state', perhaps related to increased mental effort (see §1.4), based on the correspondence depicted in Fig. 3. Data from each individual is used to establish their own baseline (median) vertical gaze direction, such that upwards gaze is a relative measure. Our aim is to use multiple indicators of struggle that may vary across individuals, but that occur consistently for a single individual. For example, an individual who does not exhibit upwards gaze when they struggle might exhibit a different indicator. Our approach should therefore be receptive to individual differences, whether cultural or idiosyncratic. If heterogenous indicators have a greater density for a given lecture segment, this becomes noteworthy for the instructor. (At present, head position data was too noisy to identify any reliable correlations.)

Our second set of results indicates how a data visualization of student struggle could benefit instructors (Fig. 4). At present, for clarity, we illustrate a minimalist version of an instructor interface prototype using only student self-report data. The figure caption describes the visualization in detail.

A full-featured interface will include embedded lecture video, for the instructor to 'seek' to relevant video positions for lecture review. It could also include anonymized background information for individual students who elect to disclose it (e.g. disabilities, non-native speakers relative to lecture language, experience level in the subject matter). These and other features would help instructors understand the overview.

4 Discussion

The results from our initial experiments show that visual evidence can be extracted from video of student facial movement that aligns temporally with aspects of a corresponding viewed lecture video. In our exploration of study design, computer vision apparatus, collected and analyzed data, and instructor interface design, we discovered strengths and limitations of our approach. A core strength of our findings is that visual evidence of struggle in facial movement analysis was often present during relevant segments of the viewed lecture video. Relevant segments challenged student comprehension due to either conceptual content or presentation issues such as a missing visual example. A core limitation of our study design was gathering self-report feedback concurrently with video lecture viewing, which interfered with facial movement data. Our study design will change in the future to allow more 'naturalistic' student viewing of the lecture. As there was a strong agreement between students' retrospective self-report and their real-time click feedback, we envision building on this to allow students to retrospectively review and annotate the video lecture.

The computer vision apparatus was largely effective, and we anticipate extending it with further "struggle" detectors and increasing automation. A follow-on study with the above-mentioned improvements in study design would provide a basis for training a machine learning model that integrated video and annotation data. We could then test how well it generalized to new student video.

We were often surprised by data we collected and analyzed, in terms of what it contained and the patterns it revealed. For example, we did not expect students to exhibit such pronounced facial movements when viewing a video alone (e.g. nodding their head). It was also interesting to see apparent visual manifestations of struggle correspond so closely to aspects of the video lecture, ranging from the use of unexpected, unusual, or new terms, to unclear lecture video imagery (e.g. blurry text), to future-oriented references such as abstract descriptions, concretized with visual examples in a following slide (a transition also reflected in the student data analysis).

Finally, taking the patterns we found in the data and visualizing them in an instructor interface was not trivial. At times, when a relevant pattern in the data was strong conceptually, it was opaque when the data was visualized. Other times, patterns in the data were easy to overlook until they were visualized. While both of these issues are typical of data visualization in the sciences, we were not able to fully anticipate how they would arise in an interface for instructors to gain insights about a student cohort.

As the last author was also the instructor who wrote and delivered the video lectures, it is of interest to report his takeaway from the study, irrespective of bias. His report is suggestive of the potential benefits to instructors that we plan to explore systematically in future work. He notes that in reflecting on his past in-person teaching, he indeed made inferences about ‘face-to-face’ student cohorts by observing behaviors, e.g. different forms of nodding in seeming comprehension, or less positive indicators such as paper rustling, mobile phone usage, or staring at the desk.

For both in-person and recorded video instruction, he received positive student feedback, collected following his lectures. Having done this study, he now sees how student feedback evolved from one recorded lecture segment to another, rather than being a gestalt post-lecture impression. He also sees at a glance how many of the times that students indicated challenge corresponded to a lecture segment with presentation issues as opposed to those with genuine challenging content (see Fig. 4).

5 Conclusion

This implementation of the PUZZLED prototype realized the aims of its design. It provided an instructor with insights about how segments of their lectures related to the students’ experience of them. Conceptually challenging lecture segments corresponded to pronounced student struggle patterns initially, which then transitioned back to a baseline. In context, for the lecturer who structured the content, this indicated that students were at least coping with and potentially learning advanced techniques in the subject matter. A few students who struggled more often than others may have needed further support to the get the most from the lesson. In still other instances, the content or order of slides could be modified to provide (e.g.) visual examples at key moments.

We imagine that PUZZLED could also indicate where instructors might increase the challenge level of under-challenging content, to encourage productive struggle. In addition, we believe it can help contribute to more inclusive online HE instruction, by facilitating instructors' ability to receive non-verbal feedback from all students using recorded video lectures. A second prototype will be tested with more instructors.

References

1. English, A.R.: *Discontinuity in learning*. Cambridge University Press (2013).
2. Boaler, J.: *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teaching*. John Wiley & Sons (2015).
3. Hiebert, J., Grouws, D.A.: The effects of classroom mathematics teaching on students' learning. *Second handbook of research on mathematics teaching and learning*. 1, 371–404 (2007).
4. Shulman, L.S.: Those who understand: Knowledge growth in teaching. *Journal of Education*. 193, 1–11 (2013).
5. Warshauer, H.K.: Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education*. 18, 375–400 (2015).
6. Alexander, R.: *Towards Dialogic Teaching: Rethinking Classroom Talk*. Dorchester Publishing Company, Incorporated (2008).
7. Murdoch, D., English, A.R., Hintz, A., Tyson, K.: Feeling heard: Inclusive education, transformative learning, and productive struggle. *Educational Theory*. 70, 653–679 (2020).
8. Lipman, M.: *Thinking in education*. Cambridge university press (2003).
9. Oser, F., Spychinger, M.: *Lernen ist schmerzhaft. Zur Theorie der Fehlerkultur und zur Praxis des Negativen Wissens*. Beltz (2005).
10. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. *Learning and Instruction*. 22, 145–157 (2012).
11. Nelson Laird, T.F., Chen, D., Kuh, G.D.: Classroom Practices at Institutions with Higher-than-Expected Persistence Rates: What Student Engagement Data Tell Us. *New Directions for Teaching and Learning*. 115, 85–99 (2008).
12. Trowler, P., Trowler, V.: *Student engagement evidence summary*. The Higher Education Academy (2010).
13. Y. Chen, Q. Chen, Mingqian Zhao, S. Boyer, K. Veeramachaneni, H. Qu: DropoutSeer: Visualizing learning patterns in Massive Open Online Courses for dropout reasoning and prediction. In: *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*. pp. 111–120 (2016).
14. Lee, Y., Choi, J.: A review of online course dropout research: Implications for practice and future research. *Educational Tech. Res. and Dev*. 59, 593–618 (2011).
15. Rios-Amaya, J., Secker, J., Morrison, C.: *Lecture recording in higher education: Risky business or evolving open practice*. LSE/University of Kent (2016).
16. Newland, B.: *Lecture Capture in UK HE 2017*. HeLF UK (2017).
17. Gaebel, M., Kupriyanova, V., Morais, R., Colucci, E.: E-*Learning in European Higher Education Institutions: Results of a Mapping Survey Conducted in October-December 2013*. European University Association. (2014).
18. High Level Group on the Modernisation of Higher Education: *Report to the European Commission on improving the quality of teaching and learning in Europe's higher education institutions*. Publications Office of the European Union (2013).

19. Meltzer, D.E., Manivannan, K.: Transforming the lecture-hall environment: The fully interactive physics lecture. *Am. Journal of Physics*. 70, 639–654 (2002).
20. Phuong, A.E., Nguyen, J., Marie, D.: Evaluating an Adaptive Equity-Oriented Pedagogy. *Journal of Effective Teaching*. 17, 5–44 (2017).
21. Bourgatte, M., Fournout, O., Puig, V.: Les technologies du numérique au service de l'enseignement: Vers un apprentissage instrumenté et visuel. In: Châteauvert, J. and Delavaud, G. (eds.) *D'un écran à l'autre*. pp. 437–456. l'Harmattan (2016).
22. J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, J. C. Lester: Automatically Recognizing Facial Indicators of Frustration: A Learning-centric Analysis. In: 2013 *Humaine Association Conference on Affective Computing and Intelligent Interaction*. pp. 159–165 (2013).
23. J. Whitehill, Z. Serpell, Y. Lin, A. Foster, J. R. Movellan: The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions. *IEEE Transactions on Affective Computing*. 5, 86–98 (2014).
24. Mohamad Nezami, O., Dras, M., Hamey, L., Richards, D., Wan, S., Paris, C.: Automatic recognition of student engagement using deep learning and facial expression. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. pp. 273–289. Springer (2019).
25. Bosch, N., D'Mello, S.K., Ocumpaugh, J., Baker, R.S., Shute, V.: Using Video to Automatically Detect Learner Affect in Computer-Enabled Classrooms. *ACM Trans. Interact. Intell. Syst.* 6, (2016).
26. Huron, S., Isenberg, P., Fekete, J.D.: PolemicTweet: Video annotation and analysis through tagged tweets. In: *IFIP Conference on Human-Computer Interaction*. pp. 135–152. Springer (2013).
27. Schuller, B., Müller, R., Eyben, F., Gast, J., Hörnler, B., Wöllmer, M., Rigoll, G., Höthker, A., Konosu, H.: Being bored? Recognising natural interest by extensive audiovisual integration for real-life application. *Image and Vision Computing*. 27, 1760–1774 (2009). https://doi.org/10.1016/j.imavis.2009.02.013.
28. Nagasawa, T., Takahashi, R., Koopipat, C., Tsumura, N.: Stress Estimation Using Multimodal Biosignal Information from RGB Facial Video. In: *IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*. pp. 292–293 (2020).
29. Y. Zhao, X. Wang, E. M. Petriu: Facial expression analysis using eye gaze information. In: *2011 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings*. pp. 1–4 (2011).
30. Baron-Cohen, S., Cross, P.: Reading the Eyes: Evidence for the Role of Perception in the Development of a Theory of Mind. *Mind & Lang*. 7, 172–186 (1992).
31. Doherty-Sneddon, G., Phelps, F.G.: Teachers' Responses to Children's Eye Gaze. *Educational Psychology*. 27, 93–109 (2007).
32. Glenberg, A.M., Schroeder, J.L., Robertson, D.A.: Averting the gaze disengages the environment and facilitates remembering. *Mem. & Cog*. 26, 651–658 (1998).
33. Gao, H., Yüce, A., Thiran, J.-P.: Detecting emotional stress from facial expressions for driving safety. In: *2014 IEEE International Conference on Image Processing (ICIP)*. pp. 5961–5965. (2014).