Leveraging Eye Tracking in Digital Classrooms: A Step Towards Multimodal Model for Learning Assistance

Sean Anthony Byrne
sean.byrne@imtlucca.it
MoMiLab, IMT School for Advanced Studies Lucca
Lucca, LU, Italy

Nora Castner
nora.castner@uni-tuebingen.de
Human-Computer Interaction, University of Tübingen
Tübingen, Germany

Ard Kastrati
akastrati@ethz.ch
ETH Zurich
Zurich, Switzerland

Martyna Plomecka
martyna.plomecka@uzh.ch
Department of Psychology, University of Zurich
Zurich, Switzerland

William Schaefer
william.schaefer@utsa.edu
University of Texas at San Antonio
USA

Enkelejda Kasneci
enkelejda.kasneci@tum.de
Human-Centered Technologies for Learning, Technical University of Munich
Munich, Germany

Zoya Bylinskii
bylinski@adobe.com
Adobe Research
USA

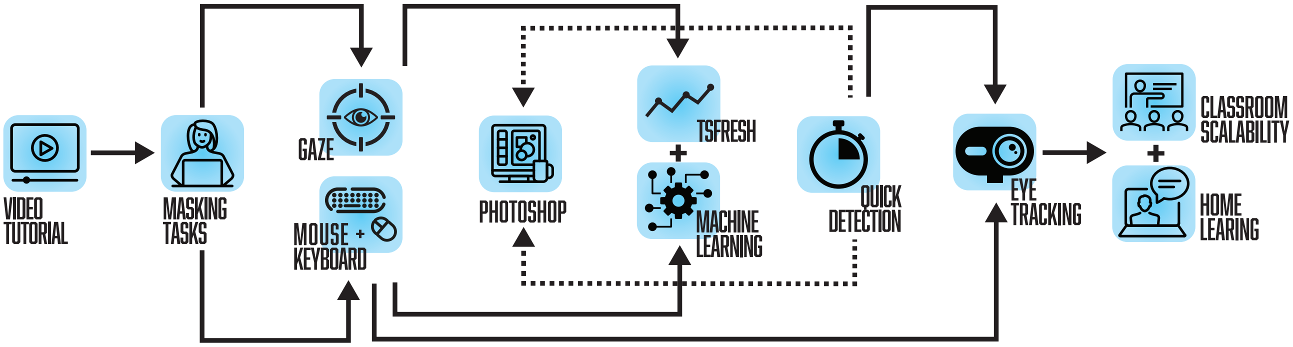

Figure 1: Our preliminary investigation uses a multimodal design to examine how students complete assigned workflows using Adobe Photoshop. After watching a video tutorial, participants (new to Photoshop) completed two masking tasks, with the second task being more challenging than the first. Eye tracking data, mouse clicks, and key strokes were recorded while participants completed the tasks and could toggle between a task and the video tutorial at any point. An initial ML model combines the inputs to predict overall task performance in the first five seconds of user interaction data. We outline the current challenges and discuss our visions for building on this pipeline to identify areas of difficulty for students and provide timely interventions, such as tailored tutorial prompts or in-classroom assistance, and expand this pipeline to incorporate students who learn at home via webcam-based eye tracking.

ABSTRACT

Instructors who teach digital literacy skills are increasingly faced with the challenges that come with larger student populations and online courses. We asked an educator how we could support student learning and better assist instructors both online and in the classroom. To address these challenges, we discuss how behavioral signals collected from eye tracking and mouse tracking can be combined to offer predictions of student performance. In our preliminary study, participants completed two image masking tasks

ETRA '23, May 30–June 02, 2023, Tubingen, Germany

ACM ISBN 979-8-4007-0150-4/23/05…$15.00
https://doi.org/10.1145/3588015.3589197

in Adobe Photoshop based on real college-level course content. We then trained a machine learning model to predict student performance in each task based on data from other students, as a step towards offering automated student assistance and feedback to instructors. We reflect on the challenges and scalability issues to deploying such a system in-the-wild, and present some guidelines for future work.

CCS CONCEPTS

• **Applied computing** → **Psychology**; **Interactive learning environments**; • **Human-centered computing** → *Usability testing*.

KEYWORDS

Eye Tracking, Usability, Scanpath Analysis, Adobe Photoshop, Multimedia Learning, Intelligent Tutoring Systems, Digital Tools

ACM Reference Format:
Sean Anthony Byrne, Nora Castner, Ard Kastrati, Martyna Plomecka, William Schaefer, Enkelejda Kasneci, and Zoya Bylinskii. 2023. Leveraging Eye Tracking in Digital Classrooms: A Step Towards Multimodal Model for Learning Assistance. In *2023 Symposium on Eye Tracking Research and Applications (ETRA '23), May 30–June 02, 2023, Tubingen, Germany*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3588015.3589197

1 INTRODUCTION

The rising need for digital literacy and growing digital classrooms. In our present economy, software competency is a much desired skill across many industries. Colleges are stepping up to the challenge by growing their computer science/engineering departments and labs and offering online certification programs to increase student throughput [Crick et al. 2020; Guzdial and Morrison 2016]. These trends are putting an increasing burden on instructors, who are faced with larger classrooms and more required teaching, often leading to teachers being under resourced. Meaning, over divided attention could lead to missing when a student is in need of assistance. Furthermore, some classes and certifications are moving entirely online, where often the teacher cannot see the students [Garris and Fleck 2022]. Overall, it becomes increasingly difficult for instructors to observe student struggles or offer individualized guidance. In this paper, we ask how we can support student learning while providing tools to assist instructors both online and in the classroom. From this question, we set out to tackle a topic that has often tried to be solved, but comes with glaring issues related to real world environments and scalability. Here, we present our work towards assistive models for supporting educators and students, but more important, we present potential solutions for known problems that large scale systems can face in an educational environments.

Motivations and guiding questions. The foundations for this paper emerged out of discussions with an Adobe Education Specialist who teaches an average of 3-4 courses a semester, and holds both large training sessions and one-on-ones, with a throughput of 500-700 trainees per semester. *"When classes are in person in the lab, we can typically accommodate 10-15 students. When we moved some of these courses online, 50-person classes filled up within a day, and it is not uncommon to see 200-person classes for the same content."* This education specialist leads training courses and produces educational material (e.g., video tutorials) to teach students how to use creative tools like Adobe Photoshop and Illustrator. Given the complexity of these tools, a lot of in-person, individualized guidance is required to help students when they get stuck. However, it is not possible when classes grow and move online. These challenges motivated the following guiding questions:

(1) What behavioral signals can be automatically captured while students complete digital workflows as additional feedback to the instructor about where students commonly get stuck?
(2) Can behavioral signals be used to automatically alert an instructor when a given student is about to or has gotten stuck?
(3) Can such automatic predictions of student confusion be used to trigger tutorial content, and lessen the load on the instructor?

The rest of this paper is our initial attempt to answer these questions and provide a set of guidelines for future systems that could be deployed in computer labs/classrooms. The system endeavors to analyze student behaviors, offer automated help, and provide feedback to instructors (see pipeline in figure 1). It will align with the guidelines based both on a survey of prior work and our initial investigation. We include the results of a preliminary investigation where we captured student eye movements, mouse movements, and keyboard presses as they completed digital workflows using Adobe Photoshop, and a machine learning model that used these behavioral signals to predict student performance. This work can be considered the beta version of a real-time teaching assistant we are developing. Moreover, we present some challenges, outline considerations for scaling such approaches to future digital classrooms, and foster discussion surrounding the usefulness of multimodal inputs into these systems.

2 SURVEY OF PRIOR WORK

Intelligent tutoring systems. Intelligent tutoring systems (ITS) have been shown to be effective for a vast array of use cases, including mechanisms that can predict users' states [Azcona et al. 2019; Hutt et al. 2021; Wang et al. 2021], automatic generation of instructional content [O'Rourke et al. 2015; Ramesh et al. 2011; Wambsganss et al. 2021] or videos [Chi et al. 2012; Pongnumkul et al. 2011], and even attention guiding [Castner et al. 2020]. These systems have been widely adopted in online learning, but their applicability to in-person settings should not be overlooked. In digital classrooms, where every student has a computer or other digital device to work on, ITSs have the potential to provide educators with direct feedback on student behavior. In this context, Ramesh et al. developed an adaptive tutorial interface that was capable of transferring knowledge between software applications (e.g., Photoshop to GIMP) to help users accomplish specific tasks. Other researchers have relied on gamification to improve task performance or learn new tools in software like AutoCAD [Li et al. 2012, 2014].

Multimodal signals for user understanding. Recent studies have also demonstrated the potential of combining eye tracking with other modalities to improve adaptive interfaces. Xu et al. used an interface coupled with keyboard and mouse input to predict visual attention, while Gong et al. combined signals from eye tracking, tool sensors, and environmental sensors to accurately recognize fabrication expertise and activity being performed. Fuhl et al. found that eye tracking combined with mouse events was the best indicator for distinguishing users, and Saboundji and Rill also used gaze and mouse movements for error detection. These findings suggest that the combination of gaze and software interaction can aid in the design of intelligent assistive systems that can identify specific users and potential pain points. Moreover, multimodal input has the potential to strengthen system interpretation, where one input could be lacking or ambiguous at any moment.

Eye tracking in classrooms. A growing body of literature has leveraged eye tracking technology in real classroom settings [Keller et al. 2022]. Previous research has focused on using commercial, off-the-shelf eye tracking systems, field cameras, or 3D depth sensors to analyze the activities and engagement of students and teachers [Sumer et al. 2018]: By capturing students' gaze [Bidwell and Fuchs 2011], head pose and motion statistics [Ventura et al. 2016]. In a classroom setting, a lot is happening at once, and information is constantly changing [Jarodzka et al. 2021]. This can be challenging for both the teacher, who must manage and educate students in a personalized way, and for the students themselves, who need to extract relevant information and interact with their peers [Goldberg et al. 2021]. Eye tracking can be used to record eye movements in relation to an external stimulus to understand what a person saw. This method is traditionally used in laboratory experiments, but can be adapted to study visual perception in real-life classroom scenarios [Keller et al. 2022]. Here, we discuss the previous studies' findings that assess visual perception in the complex, dynamic classroom setting with eye tracking.

Rosengrant et al. utilized an attentional model to track students' focus and attention during lectures by monitoring gaze patterns, and varied the number of times students wore eye tracking glasses to observe if awareness affected focus over time. Results support the idea that well-structured classes with interactions between students and instructors can effectively maintain student attention throughout the class. Jarodzka et al. investigated using eye tracking the processes behind the split-attention effect in realistic settings and found that students largely neglected additional information in split designs, and ignored information they deemed optional to solve the task. Hutt et al. developed attention-aware learning technology (AALT) that detects and responds to mind wandering using gaze-based eye tracking, along with two classroom studies with 287 high-school students that demonstrated that AALT could successfully reorient attention, reduce mind wandering, and improve retention for students with low prior knowledge. Yang et al. analyzed the data from university students, aiming to capture visual attention during a presentation in a real classroom setting. Eye movements were recorded using an eye tracking system while a teacher gave a 12-15 minute presentation. The results showed that students paid more attention to text zones and the teacher's narration, but had a longer average fixation duration when viewing picture zones. Additionally, when viewing slides containing scientific hypotheses, the difference in attention between text and picture zones decreased. Finally, the earth-science majors were found to have better information decoding and integration abilities than the rest of the students. Sumer et al. approached the problem from the classroom teacher's perspective, showing that teachers' attentional processes provide essential information about their ability to focus on relevant information in the complexity of classroom interactions and distribute their attention among students to identify their learning needs. They combined mobile eye tracking with computer vision, using a state-of-the-art face detector and a novel method to cluster faces into a number of identities.

3 PRELIMINARY INVESTIGATION

Our preliminary investigation was based on our education specialist's classroom experience: *"Masking in Photoshop is a day one skill. The very first course lecture covers masking and how you would apply the skill to working with the tool on a daily basis. Just about every project in the course will build upon this skill, as it is a foundation of photo editing."* We thus selected tasks that, after a short video tutorial, could be completed by untrained participants. Prior to the current investigation, we had run a pilot study where we tested the feasibility of the software, which integrates gaze, user input, and screen recording during the same Photoshop task and tutorial interaction [Castner et al. 2022].

Participants and study design. A total of 53 students (ages 18-27, 34 females) with no prior experience in Photoshop were recruited through mass emails and word of mouth at two large European universities. The participants were asked to complete two masking tasks, in which they were required to extract a subject from one image and overlay it on another background image. Task 1 used a man as the subject of the masking task, while Task 2 used a woman as the subject, with the latter being more challenging due to the need to fix the hair (figure 2). Before starting the masking tasks, participants watched a two-minute video tutorial created by our Adobe Education Specialist. Participants then spent an average of 20 minutes completing both masking tasks, using Adobe Photoshop version 23.2.2 on a *Thinkpad X1 Carbon Gen 10, 16 GB LPDDR5 6400MHz, 1 TB M.2 2280 SSD.*

Multimodal data collection & setup. To collect data on participants' interactions with Adobe Photoshop, we measured a combination of inputs from eye tracking, mouse clicks, and key strokes for a more complete picture of how users interact with the software. To collect the eye tracking data, we used the Tobii Pro Fusion eye tracker running at 250 Hz on a monitor with full HD resolution. We used the software Titta [Niehorster et al. 2020] in Python running the eye tracker and OpenCV [Bradski 2000] to record the screen. However, due to technical errors, we were only able to use data from 39 participants for the final analysis. The raw eye tracking data was cleaned and event detection was performed using I-VT with a minimum fixation duration of 60 ms and a velocity threshold of 30 °/s using the Perception engineer's toolkit [Kübler 2020]. Mouse and keyboard events were also recorded and linked to timestamps. For more information regarding the experiment design, we refer the readers to [Castner et al. 2022].

Data processing. To process the data, we used the TSFresh Python package [Christ et al. 2018]. extracts features from the time series data independently and then concatenates the extracted features from each sensor to form a feature set [Christ et al. 2018]. In our case, the feature set is comprised of gaze data, keystrokes, and mouse clicks that can then be used as input to a machine learning model for analysis and prediction.

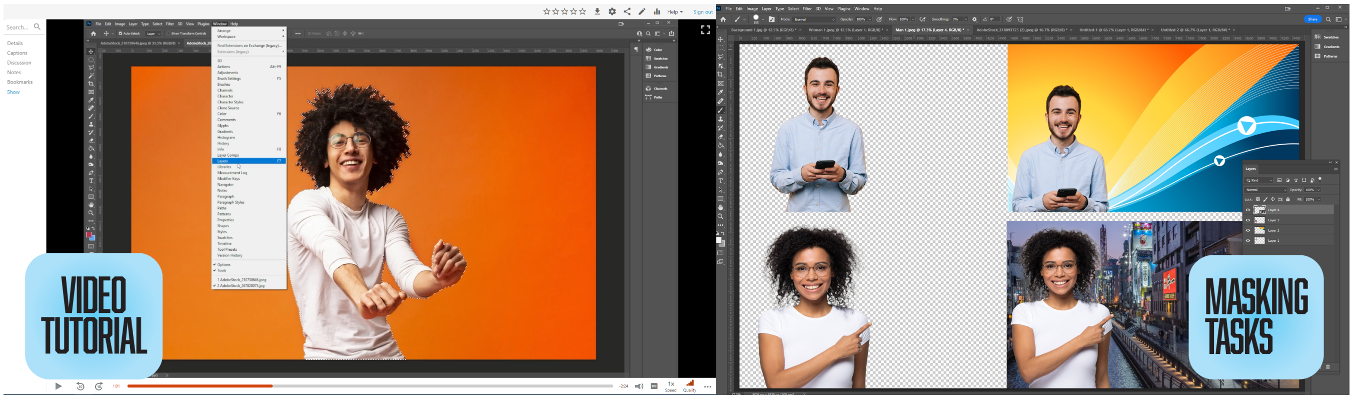

Figure 2: The initial and final stages of a masking task performed by participants using Adobe Photoshop, with the video tutorial accessible by a key stroke at any point during the task (left panel). We also illustrate the starting point of Tasks 1 & 2 and the expected outcome of each task (right panel). Adobe stock photos featured ©stock.adobe.com.

We used a supervised learning approach with evaluations provided by our education specialist as labels for our model. The educator individually assessed the quality of the final mask for each task, giving each mask a score between 1-5. The educator focused on factors such as the visibility of edges from the original image, the correction of hair in Task 2, and the presence of background elements in the mask. A score of 5 corresponds to a mask with clean edges, corrected hair, and fully masked subjects, while a score of 1 would be given to masks that ignored the video tutorial, had uneven mask lines, or inadvertently revealed the background rather than the masked individual. *"In a real classroom, I use a 5 point or 10 point rubric, depending on the project. The key difference between how I scored this project from a course project is that, in the course, I'm typically looking at a variety of other skills as well. Each project will have 10 categories of 10 points or 10 categories of 5 points."*

Our labeling method resulted in a dataset with a significant imbalance, which can present challenges in training a model, but is more reflective of real-world scenarios. To simplify the prediction task during the initial stages of data collection, we converted the labels of each task into pass/fail categories. A passing grade was assigned to students who scored 3 or higher, while scores below 3 were considered as failing. In Task 1, 54% of students "passed" while in Task 2, 28% of students received a passing grade.

Machine learning pipeline. Features obtained from the TSFresh package were used as input to machine learning models: A K-Nearest Neighbors Algorithm (KNN) and Random Forest Classifier. Our classification tasks involved predicting whether an unseen participant would pass or fail a masking task based on the first five seconds of recorded sensor data. We empirically chose this threshold to approximate an application scenario where a teaching intervention could be delivered in a timely manner if student struggle is detected. We split our 39 participants into a training and testing set using an 80-20 split. Our tests included: Predict how an unseen participant would perform on Task 1, predicting how an unseen participant would perform on Task 2, and predicting how a participant would perform on Task 2 after the first five seconds based on a model trained on the full recordings from Task 1.

Initial insights. Due to limited data and imbalanced labeling, many of our models suffered from overfitting to the majority class. However, we were able to predict pass or fail outcomes for Task 2 for a holdout sample of participants using only the first 5 seconds of sensor data, using a KNN model trained on the same time period. Accuracy was 75% at predicting pass/fail in Task 2. In initial tests, to investigate how much data is needed to make a prediction, we tested at 2, 5, 10 & 20 seconds and found that 5 seconds performed optimally, with diminishing returns after that point. As eye tracking is a costly and cumbersome technology to add to digital classrooms, we evaluated how crucial the gaze features were to our models. Towards this goal, we used Random Forest feature importance to investigate the role of each input sensor in our model. We found that the features extracted from the eye tracker accounted for 48 of the top 50 most important features out of 5482 total features for all of our classification tasks. We hypothesize that this is due to students spending a significant amount of time visually inspecting each task before attempting a workflow through mouse or keyboard use.

4 CHALLENGES AND GUIDELINES

Though eye tracking technology has gained momentum in classroom settings, there are major challenges to be addressed for a truly proficient assistive system to be actualized. Here, we want to stress the realities that come with realistic environments.

Unconstrained Tasks. In a study where there is a lot of freedom in how participants accomplish a task, it can be difficult to analyze the data because there is no clear way to identify patterns or trends. Without clear steps in the task design, it can be difficult to determine how participant behavior relates to the research question. Therefore, we designed the preliminary study with these attributes:

(1) Presenting video tutorials that describe a *sequence of steps* required to complete a task.
(2) Creating a *short* task design: The video tutorial and the masking task can be finished in a short sitting. Our initial study iterations showed that it is possible to capture and build

prediction models based on behavioral signals in a short sitting.

Signal loss. Although eye tracking technology has advanced significantly, signal loss remains one of the key challenges. This becomes even more problematic when considering large-scale eye tracking studies. High-end equipment becomes impractical, while cheaper alternatives like webcam eye trackers are even more prone to data loss [Valliappan et al. 2020; Wisiecka et al. 2022]. This was a major challenge in our preliminary study, prompting the following steps:

(1) Data preprocessing: We cleaned and preprocessed the data to remove any missing or corrupted values, which required removing 14 out of 53 participants from our collected data.
(2) Model selection: We used models robust to missing data, including decision trees or random forests, which can handle missing values without the need for imputation.
(3) Cross-validation: We used techniques such as cross-validation to ensure model robustness to data loss and variance between participants.

Deployment. There are several challenges to deploying eye tracking technology in the classroom. By collecting our data at two different universities, we already encountered deployment and generalizability issues. Some of the main lessons learned from our study iterations include:

(1) Integrating sensor data with screen recordings and syncing them to interactions in a complex UI requires a custom software engineering solution that is currently unavailable. The use of Adobe Photoshop simultaneously with the eye tracking and screen recording software put a strain on the computer's resources, leading to the CPU being overutilized and causing the system to slow down at certain timepoints in data collection, contributing to eye tracking signal loss.
(2) Ensuring that the software runs correctly and consistently on different operating systems, devices, and hardware configurations is a significant challenge. Different platforms come with their own technical limitations. This was a problem in our preliminary study, since the sampling rate of our screen recording software was dependent on the system specifications. This led to synchronization difficulties for performing the data analysis.
(3) A given classroom may have its own configuration of operating systems, Photoshop version, and eye tracking hardware, all of which evolve over time and pose generalization challenges to models trained in different settings. This would require continuous model training and re-deployment.

5 CONCLUSION

In this preliminary investigation, we provide initial evidence that user interaction and gaze behavior can be viable predictors of task performance in digital workflows. We present the early evaluation of a machine learning model that takes a features-based approach to time series via the TSFresh package [Christ et al. 2018]. Our model offers a promising first step towards automated prediction of student performance in real time. Our aim is that future iterations of this model could provide instructors with direct feedback on student behavior and alert them when a student is in need of assistance. Further extensions of such a model could be used in detecting student struggles and automatically triggering tutorial content, which would help to reduce the load on instructors and improve the effectiveness of online classes.

While we presented an initial model for student performance prediction, more work would be needed to deploy such a model in a classroom setting. To integrate such a system into the classroom further requires that it aligns with and supports instructors' existing workflows; this would allow the system to augment the instructor, rather than further burden them. To this end, we structured our tasks based on real college-level course material. Our research questions were motivated by the real needs and workflows of an educator. Below we summarize our findings with respect to our initial motivating questions from the introduction.

Behavioral signals in digital workflows. We combine gaze with digital tool interactions (mouse and keyboard) to handle both cognitive strategies (conveyed by eye movements) and specific task behavior (mouse movements). We found that input from both modalities can be at low frequencies, while still offering relevant information.

Automatic detection and instructor notification. Using both gaze and mouse events offers faster recognition of inefficient or less than ideal task behavior, which is critical for appropriate assistance [Fuhl et al. 2021; Saboundji and Rill 2020]. At the moment the pain point happens, it needs to be addressed. We found that even the first five seconds of interaction data was informative to future task performance.

Trigger content and lessen instructor load. A multimodal approach can be used to quickly trigger an assistive system that would advocate an appropriate level of support to either a teacher or an at-home learner [Fuhl et al. 2021; Gong et al. 2019; Saboundji and Rill 2020; Xu et al. 2016]. Future iterations of an assistive system can strive to be lean and unconstrained, in order to foster creativity so users are not confined to a one-style-fits-all workflow. This approach could smooth over knowledge gaps of learners. More important, human teacher resources can be more effectively allocated over multiple students, using an automated teaching assistant as a second "pair of eyes" in an educational environment. It can lower educational costs while increasing convenience as we learn to support at home or self-regulated learning.

Scalability challenges. As eye tracking technology continues to improve and costs decrease, it's likely that we will see its more widespread adoption, including the growth of webcam-based eye tracking [Valliappan et al. 2020; Wisiecka et al. 2022]. While webcam tracking has the potential to provide valuable insights into student behavior, there are a number of technical and practical challenges that must be overcome before deployment in a remote learning environment. In an online setting, students may be working in different, unconstrained environments with varying lighting and camera angles, which can make it difficult to accurately track and analyze behavior. Students may also be less willing to have their behavior tracked via webcam due to issues related to privacy and consent.

Reflections from an Education Specialist. Higher education instruction grows in complexity on a daily basis. *"Faculty are constantly asked to learn new software, bring new skills to their students, and produce artifacts that the university can track and analyze. Any tool that can be provided to instructors to find student pain points before they become disruptions will be game-changing."* By tracking students' eye movements, instructors would be able to quickly and efficiently discover where a student may be struggling with a new tool. This would allow either the instructor or a software trainer to intervene. Additionally, by collecting this data, we can better serve new instructors by noting common pain points and teaching them about them as they prepare their courses.

Closing thoughts. The rising need for digital literacy and the growing number of digital classrooms are putting an increasing burden on instructors [Goldberg et al. 2021]. We have started to address the questions of how we can support student learning while providing tools to augment instructors' capabilities motivated by an expert perspective from an educator. We have also gained some initial validation that behavioral signals like eye movements, mouse movements, and keyboard presses, can be used to automatically assess student progress [Hutt et al. 2021; Jarodzka et al. 2017; Rosengrant et al. 2021; Sumer et al. 2018; Yang et al. 2013]. We're working towards the goal of integrating such a system into in-person classrooms with the potential to improve the effectiveness of online education and provide new insights into the relationship between student behavior and learning outcomes.

REFERENCES

David Azcona, I-Han Hsiao, and Alan F Smeaton. 2019. Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints. *User Modeling and User-Adapted Interaction* 29, 4 (2019), 759–788.

Jonathan Bidwell and Henry Fuchs. 2011. Classroom analytics: Measuring student engagement with automated gaze tracking. *Behav Res Methods* 49, 113 (2011).

G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

Nora Castner, Lea Geßler, David Geisler, Fabian Hüttig, and Enkelejda Kasneci. 2020. Towards expert gaze modeling and recognition of a user's attention in realtime. In *24th International Conference on Knowledge-based and Intelligent Information & Engineering Systems.*

Nora Castner, Bela Umlauf, Ard Kastrati, Martyna Beata Płomecka, William Schaefer, Enkelejda Kasneci, and Zoya Bylinskii. 2022. A gaze-based study design to explore how competency evolves during a photo manipulation task. In *2022 Symposium on Eye Tracking Research and Applications.* 1–3.

Pei-Yu Chi, Sally Ahn, Amanda Ren, Mira Dontcheva, Wilmot Li, and Björn Hartmann. 2012. MixT: automatic generation of step-by-step mixed media tutorials. In *ACM UIST.* 93–102.

Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. 2018. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing* 307 (2018), 72–77.

Tom Crick, Cathryn Knight, Richard Watermeyer, and Janet Goodall. 2020. The impact of COVID-19 and "Emergency Remote Teaching" on the UK computer science education community. In *United Kingdom & Ireland Computing Education Research Conference.* 31–37.

Wolfgang Fuhl, Nikolai Sanamrad, and Enkelejda Kasneci. 2021. The gaze and mouse signal as additional source for user fingerprints in browser applications. *arXiv preprint arXiv:2101.03793* (2021).

Christopher P Garris and Bethany Fleck. 2022. Student evaluations of transitioned-online courses during the COVID-19 pandemic. *Scholarship of Teaching and Learning in Psychology* 8, 2 (2022), 119.

Patricia Goldberg, Ömer Sümer, Kathleen Stürmer, Wolfgang Wagner, Richard Göllner, Peter Gerjets, Enkelejda Kasneci, and Ulrich Trautwein. 2021. Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. *Educational Psychology Review* 33, 1 (2021), 27–49.

Jun Gong, Fraser Anderson, George Fitzmaurice, and Tovi Grossman. 2019. Instrumenting and analyzing fabrication activities, users, and expertise. In *ACM CHI.* 1–14.

Mark Guzdial and Briana Morrison. 2016. Growing computer science education into a STEM education discipline. *Commun. ACM* 59, 11 (2016), 31–33.

Stephen Hutt, Kristina Krasich, James R. Brockmole, and Sidney K. D'Mello. 2021. Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. In *ACM CHI.* 1–14.

Halszka Jarodzka, Kenneth Holmqvist, and Hans Gruber. 2017. Eye tracking in Educational Science: Theoretical frameworks and research agendas. *Journal of eye movement research* 10, 1 (2017).

Halszka Jarodzka, Irene Skuballa, and Hans Gruber. 2021. Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom. *Educational Psychology Review* 33, 1 (2021), 1–10.

Lena Keller, Kai S Cortina, Katharina Müller, and Kevin F Miller. 2022. Noticing and weighing alternatives in the reflection of regular classroom teaching: Evidence of expertise using mobile eye-tracking. *Instructional Science* 50, 2 (2022), 251–272.

Thomas C Kübler. 2020. The Perception Engineer's Toolkit for Eye-Tracking data analysis. In *ACM ETRA.*

Wei Li, Tovi Grossman, and George Fitzmaurice. 2012. GamiCAD: a gamified tutorial system for first time autocad users. In *ACM UIST.* 103–112.

Wei Li, Tovi Grossman, and George Fitzmaurice. 2014. CADament: a gamified multi-player software tutorial system. In *ACM CHI.* 3369–3378.

Diederick C Niehorster, Richard Andersson, and Marcus Nyström. 2020. Titta: A toolbox for creating PsychToolbox and Psychopy experiments with Tobii eye trackers. *Behavior research methods* 52, 5 (2020), 1970–1979.

Eleanor O'Rourke, Erik Andersen, Sumit Gulwani, and Zoran Popović. 2015. A framework for automatically generating interactive instructional scaffolding. In *ACM CHI.* 1545–1554.

Suporn Pongnumkul, Mira Dontcheva, Wilmot Li, Jue Wang, Lubomir Bourdev, Shai Avidan, and Michael F Cohen. 2011. Pause-and-play: automatically linking screencast video tutorials with applications. In *ACM UIST.* 135–144.

Vidya Ramesh, Charlie Hsu, Maneesh Agrawala, and Björn Hartmann. 2011. ShowMeHow: translating user interface instructions between applications. In *ACM UIST.* 127–134.

David Rosengrant, Doug Hearrington, and Jennifer O'Brien. 2021. Investigating student sustained attention in a guided inquiry lecture course using an eye tracker. *Educational psychology review* 33, 1 (2021), 11–26.

Rachid Riad Saboundji and Róbert Adrian Rill. 2020. Predicting Human Errors from Gaze and Cursor Movements. In *International Joint Conference on Neural Networks.* IEEE, 1–8.

Omer Sumer, Patricia Goldberg, Kathleen Sturmer, Tina Seidel, Peter Gerjets, Ulrich Trautwein, and Enkelejda Kasneci. 2018. Teachers' perception in the classroom. In *CVPR Workshops.* 2315–2324.

Nachiappan Valliappan, Na Dai, Ethan H. Steinberg, Junfeng He, Kantwon Rogers, Venky Ramachandran, Pingmei Xu, Mina Shojaeizadeh, Li Guo, Kai J. Kohlhoff, and Vidhya Navalpakkam. 2020. Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nature Communications* 11 (2020).

Jonathan Ventura, Steve Cruz, and Terrance E Boult. 2016. Improving teaching and learning through video summaries of student engagement. In *Workshop on Computational Models for Learning Systems and Educational Assessment.*

Thiemo Wambsganss, Tobias Kueng, Matthias Soellner, and Jan Marco Leimeister. 2021. ArgueTutor: An adaptive dialog-based learning system for argumentation skills. In *ACM CHI.* 1–13.

Xi Wang, Zoya Bylinskii, Monica Castelhano, James Hillis, and Andrew Duchowski. 2021. EMICS'21: Eye Movements as an Interface to Cognitive State. In *CHI Extended Abstracts.* 1–6.

Katarzyna Wisiecka, Krzysztof Krejtz, Izabela Krejtz, Damian Sromek, Adam Cellary, Beata Lewandowska, and Andrew Duchowski. 2022. Comparison of Webcam and Remote Eye Tracking. In *ACM ETRA.* https://doi.org/10.1145/3517031.3529615

Pingmei Xu, Yusuke Sugano, and Andreas Bulling. 2016. Spatio-temporal modeling and prediction of visual attention in graphical user interfaces. In *ACM CHI.* 3299–3310.

Fang-Ying Yang, Chun-Yen Chang, Wan-Ru Chien, Yu-Ta Chien, and Yuen-Hsien Tseng. 2013. Tracking learners' visual attention during a multimedia presentation in a real classroom. *Computers & Education* 62 (2013), 208–220.