Using Eye-tracking to Examine Student Attention

MATTHEW RE*, Clemson University, USA

This is a final report for a CPSC8810 project that focuses on developing a proof-of-concept for the use eye-tracking technology to identify patterns in college student attention and focus. This project will make use of the Pupil Labs eye-tracker and Pupil Capture software to track student eye movement and fixation during a Java lecture. This report outlines an introduction to the potential study, a dive into the background of existing related work, the methodology used for this proof of concept, an explanation of how each point of interest can be evaluated using this mythology, and an evaluation of the methodology for answering some relevant hypotheses.

CCS Concepts: • **Human-centered computing** → Visualization; • **Applied computing** → **Education**;

Additional Key Words and Phrases: eye-tracking, education, attention

ACM Reference Format:
Matthew Re. 2023. Using Eye-tracking to Examine Student Attention. *ACM Trans. Graph.* XX, X, Article XXX (X 2023), 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Today's students are in a position where the day-to-day workload from all different sources is constantly increasing, and with technology being commonplace in the classroom, especially in tech-related fields, it is no surprise that students might be physically in class but mentally on a different topic. This project aims to study how college students participate in their course instruction. Specifically, this project will investigate student attention and engagement by examining how, when, and why students focus on specific and relevant material for a university lecture through eye-tracking. There are many different components to any particular course. Of the components that can easily be tracked using an eye-tracker, namely, note-taking, lecturer-provided material such as PowerPoints and figures, and the lecturer themselves, all of these provide vital information for a student's success in any particular course. While some lecturers emphasize one component over another for presenting the critical content to the students, it stands to reason that if a lecturer chooses to include a slide or figure, the lecturer wants students to focus on that slide or figure for a non-zero amount of time. This study's results will help identify possible ways to measure student focus and identify patterns in habits that revolve around attention.

An article written by Keller *et al.* proposes a framework for attention that takes place in two dimensions: internal and external, and on-topic and off-topic [Keller et al. 2020]. This study will focus on Keller *et al.*'s proposal for using recent advances in portable eye-tracking to measure how and why students move between these dimensions for attention. This study will occur in a research lab on Clemson's campus, where the researcher will present a set of four different slides. The format of these slides follows the given format:

- A "good" presentation with black text, white background, easy-to-read font, and subtle use of figures.
- A "bad" presentation with grey text, blue background, difficult-to-read font.
- A "normal" presentation focused on visual teaching (heavy use of figures).
- A "normal" presentation with no figures, only text.

During the study, the researcher will attempt to gauge which components of each of the four presentation leads to students moving between the different dimensions of attention outlined in Keller *et al.*, with a special focus on the on-topic and off-topic dimension. Specifically, the points of interest are:

- What components of the presentation are students focused on?
- How long are students fixated on Areas of Interest (AoI)?
- When a student's gaze leaves an AoI, where is it going?
- Are there areas outside of AoIs that attract student gaze?
- Is there a defined pattern that appears to explain why students move to and from the different dimensions of attention outlined in Keller *et al.*?

In the next section of this report, previous work related to this project is examined to explore what similar methods other researchers have employed to investigate similar points of interest. Based on this previous work, some hypotheses will be theorized to help guide the direction of the study.

2 BACKGROUND WORK

2.1 Eye-tracking in Education

Eye-tracking usage has been prevalent in education research even before more affordable versions of portable eye-tracking became available. A study conducted back in 2005 presented a set of PowerPoint slides and attempted to generate an average fixation profile for each slide [Slykhuis et al. 2005]. The results of this study allowed the researchers to determine which components of the PowerPoint (photos, text, title, etc.) participants were focused on and their frequency. Ultimately, a lot of what this study was focused on was whether narration changed which components of the PowerPoint participants focused on. While this proposed study will not look into narration, it does give an idea of potential areas of interest that could be defined during the lecture. A later paper by Busjahn *et al.* conducted a similar project where researchers presented Java source code to participants to track eye movement as the participants read through the code [Busjahn et al. 2014]. This project focused primarily on code comprehension and understanding. Still, it showed that gaze analysis is merited in education and comprehension, often correlated with attention and focus [Stern and Shalev 2013].

Author's address: Matthew Re, mjre@clemson.edu, Clemson University, Clemson, SC, USA.

0730-0301/2023/X-ARTXXX $15.00
https://doi.org/XXXXXXX.XXXXXXX

2.2 Examining Student Attention

In 2020, Rosengrant *et al.* conducted a study that emphasized investigating student attention [Rosengrant et al. 2021]. This study attempted to examine the belief that student attention decreases after the 10-15 minute mark of class. Researchers found that this belief is primarily unfounded and that students can focus for up to 70 minutes so long as the lecture contains techniques that engage the learner. The paper concludes by calling for future work to examine gaze focus and duration across different materials such as PowerPoints, notes, and other visual aids. Finally, another study published in 2020 investigated whether students who give visual attention to the instructor in an online lecture affect the student's performance and satisfaction with the course [Wang et al. 2020]. While these results are significant, it is essential to remember that the lectures in the study took place online, one with a camera-on instructor and one without the view of an instructor. In the lecture with the instructor, only the instructor's headshot was recorded, meaning that aspects such as instructor movement, hand movement, and the instructor physically pointing to parts of the PowerPoint could potentially yield different results and warrant another study.

2.3 Hypotheses

Based on this prior work, this study proposes the following hypotheses to assist in guiding the implementation of the methodology and analysis of the results:

- H1: For the "good" and "bad" PowerPoints, the number of gaze points on an AoI (PowerPoint, notes, and figures) will be similar across both mediums. However, focus will waver earlier with the second PowerPoint due to the unreadability.
- H2: For the "good" PowerPoint, student attention will follow a pattern similar to what was outlined in Slykhuis *et al.,* 2005. The "bad" PowerPoint will see more gaze points aimed at figures when compared to the text of that same PowerPoint.
- H3: Slides with more extended amounts of text and fewer figures in any PowerPoint will result in less time spent in external and on-topic focus as described in Keller *et al.,* 2020.
- H4: There will be a definitive trend with the time students spend directing their attention to figures.

Given the proof-of-concept nature of this study, this report will focus less on determining whether these hypotheses are true or false. Instead, this report will attempt to prove that the methodology provided in the next section has the potential to yield data that could prove these hypotheses true or false and investigate the points of interest outlined in the introduction. The hypotheses will be investigated in the Discussion section of this report.

3 METHODOLOGY

3.1 Hardware and Software

This project utilizes the Pupil Labs Core eye-tracker and its accompanying software. For hardware, the Core eye-tracker has three cameras, one world camera (60Hz@720p) and two individual eye cameras (200Hz@192x192x). These cameras have a 0.60-degree gaze accuracy and a 0.02 precision. The world camera has two different lenses: wide-angle and narrow-angle. While the wide angle lens

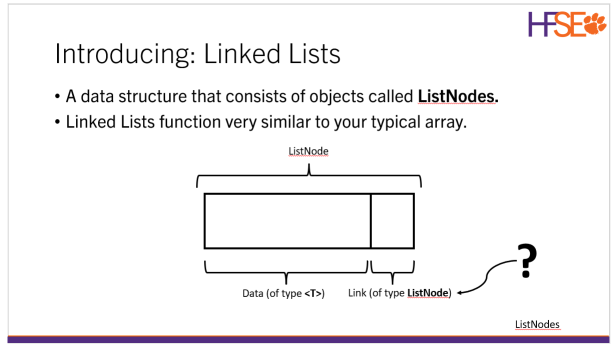

Fig. 1. A "good" presentation with subtle use of figures.

allows for a much larger Field of View (FoV) in both the vertical and horizontal axes, keeping two objects in different positions on the Z axis in focus is difficult. As such, the narrow-angle lens was used for the final study.

Pupil Labs uses two main software: Pupil Capture for recording and data collection and Pupil Player, which is Pupil Labs' media and data visualizer. Both pieces of software allow for the usage of different plug-ins to aid in collecting and analyzing the data. For recording the data (Pupil Capture), the following plug-ins were used:

- Annotation Capture - for confirming calibration.
- Blink Detector - for removing invalid data points.
- Fixation Detector - for detecting fixations with the following criteria:
  - Maximum Dispersion (degrees) - 1.50
  - Minimum Duration (milliseconds) - 120
  - Maximum Duration (milliseconds) - 300

For analyzing the data (Pupil Player), the following plug-ins were used:

- Surface Tracker - for tracking AoIs using AprilTags.
- Fixation Detector - for detecting fixations with the same criteria outlined for the Pupil Capture.
- Vis circle & polyline - for visualizing the gaze positions.

In order to utilize the Surface Tracker plug-in, Pupil Labs makes use of AprilTags, which is Pupil Labs' chosen fiducial system. The version of AprilTags used for this project was version **tag36h11**. Calibration for the eye-tracker was done using Pupil Capture's built-in calibration system. The participant gazed at a defined point on a slide to confirm the calibration while activating the annotation plug-in. This allowed for a match between an expected data point and an actual data point.

3.2 Data Collection

In order to investigate how the different components of a lecture affect student attention, four PowerPoints, outlined in the introduction section of this report, were created and presented to the participant wearing the eye-tracker. Each presentation took about two minutes and examples of the used PowerPoints are given below in **Figures 1, 2, 3,** and **4.**

In order to investigate the role of a presenter and how students give attention to one, the first two presentations had a "presenter"

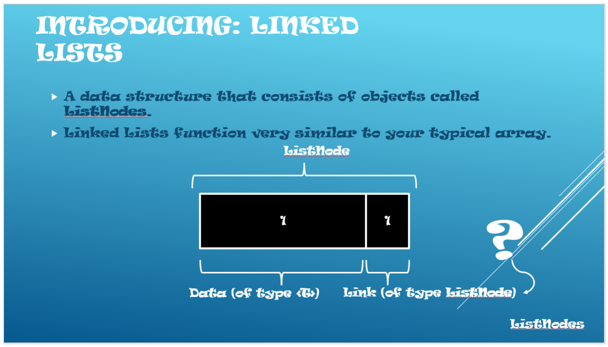

Fig. 2. A “bad“ presentation with barely readable font.

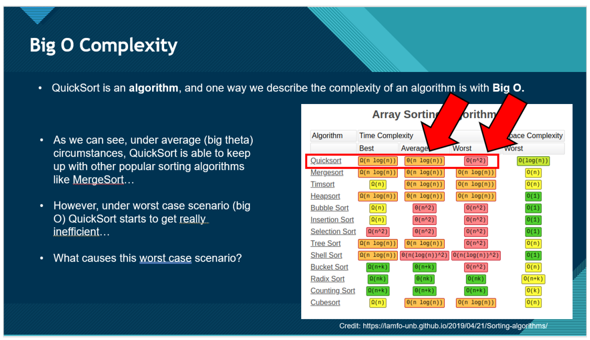

Fig. 3. A “good“ presentation with an emphasis on figures.

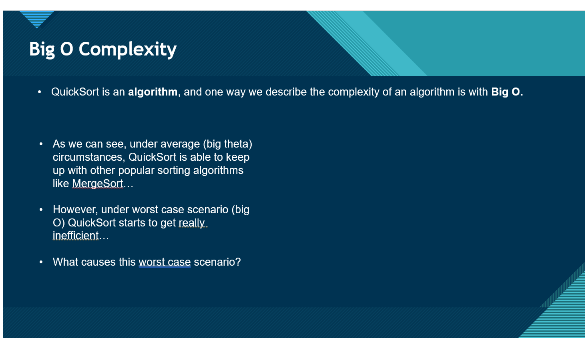

Fig. 4. A “good“ presentation with no figures, only text

who would read aloud the text, explain the figures (when present), and point to various points on the PowerPoint. The original goal was to use Densepose[1] to look at when and where the participant was looking at the presenter. However, it was soon found that Densepose was incompatible with the Pupil Labs Core headset. Before each PowerPoint was presented, a calibration was run, and it was ensured that the material presented to the participant was new.

Once the data is recorded, the Pupil Player software will export the world video with the vis circle and polyline, data pertaining to the AoIs, and other general gaze data. The data is exported into CSV files, except for the world video. Once the CSV files are collected, they are converted into Microsoft Excel Worksheets (.xlsx) and parsed to extract meaningful data for this study. The Pupil Player exports about 15 CSV files, varying depending on how many surfaces were defined using AprilTags. Two to Three surfaces were defined for each presentation video. Namely, the PowerPoint itself, any figures on that PowerPoint (where applicable), and the student's notes. It was intended that a surface would be defined for the participant's phone. However, that surface's AprilTag was not picked up by the eye-tracker. As such, the surface is ignored and grouped in with the non-defined surface category. Of all of the CSV files exported, the ones that are of significant use for this study are:

- gaze_positions_on_surface_<SurfaceName>.csv
- blinks.csv
- gaze_positions.csv
- fixations.csv
- surface_events.csv
- surface_gaze_distribution.csv
- surface_visibility.csv

The following section will investigate how the raw data was parsed through and used to explain how a study of this nature would be able to investigate the points of interest presented in the introduction section and the hypotheses presented in the background work section.

4 RESULTS

The structure of this section is that each of the points of interest will be a subsection. Within each subsection will be an explanation of which parts of the relevant CSV files help investigate that point of interest and a presentation of some of the data collected from this study.

It is important to note that all data presented in this report is “dummy data.“ The student/participant was already aware of the study's goals, and this test study was conducted in a non-natural environment. Furthermore, the goal of presenting data was to show the *potential* to yield good data in a natural, non-biased study. A Google Drive with all the “dummy data“ can be found at this link[2].

4.1 What Components of the Presentation are Students Focused On?

For each presentation, one AprilTag tracked anything pertaining to the PowerPoint. Using that one AprilTag, it is possible to track multiple surfaces. For each of the presentations, text and figures were laid out in the same format, where text typically was placed on the top or left of the PowerPoint slide, and figures were placed at the bottom and left parts of the slide. With this setup, it was possible to track when the student was looking at the text, when they were looking at figures, when they were looking at white space, and when they were looking at none of the above. For this question, only the first three PowerPoints were considered since the fourth did not contain any figures, and most white space was used for text. Using surface_gaze_distribution.csv for each presentation, the numbers can be compiled into one CSV file and combined to show the total

[1] https://docs.pupil-labs.com/alpha-lab/dense-pose/

[2] https://drive.google.com/drive/folders/1xSSPfFbQHem-Qb9LoNK8WZ5HlhPZnnvk?usp=sharing

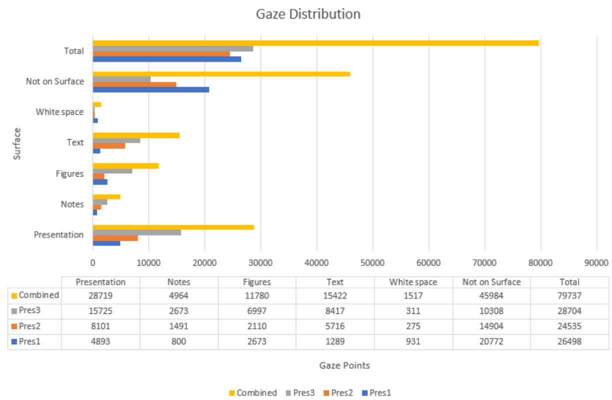

Fig. 5. Combined gaze distribution for the three presentations combined.

gaze distribution between the presentation components. Using this, the following visualization given in **Figure** 5 can be produced and examined to determine where students are dedicating their focus and attention to the presentation.

One potential factor to make apparent is that for this to work, the components of the presentation must be in a consistent layout within the presentations. Once a surface has been defined, it cannot be edited later in the recording, meaning that while it is still possible, it is not recommended to be inconsistent in text, figure, and white space placement and still expect the same possible result.

4.2 How Long are Students Fixated On Designated AoIs?

The third presentation will be used for this point of interest, given its emphasis on figures. When a surface is defined in the Pupil Player and once the data is exported, a gaze_positions CSV file with the surface name will appear. Opening the CSV file for the figure on presentation 3, there were two instances when the figures appeared on the screen. Both instances were in the exact location on the presentation. Once the data about when the figures are on screen is isolated, the relevant columns for this point of interest become world_timestamp, world_index, on_surf, and confidence. An extra column can be added called time_between that tracks the amount of time, in milliseconds, between each frame. With the specified camera spec values, the time between each frame was either 0.036 or 0.032 milliseconds. Getting a count of all the unique world indices represents the total number of frames the figure was on screen. When combined, we can identify the total number of frames and milliseconds that the student was gazing at the figure, along with a timeline of when these gazes took place across the time the figure was on screen. When designating true (for on_surf) as equal to one and false as equal to zero, the following visualization presented in **Figure** 6 with a trendline can be produced.

4.3 When a Student's Gaze Leaves an AoI, Where is it Going?

Regardless of the number of defined surfaces, a surface_events.csv file will be generated so long as there is at least one. This file, containing four columns, tracks when (by frame and world index) a

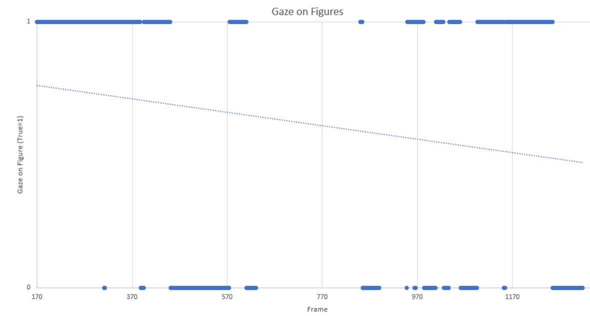

Fig. 6. Trend of student gaze on figures across time.

gaze has entered and exited a surface. When combining this CSV file with the vis circle and polyline, qualitative data on where a student's gaze is going once it leaves an AoI. For the second presentation, the surface_events.csv file reports the following:

World_index	surface_name	event_type
983	Presentation	exit
983	Text	exit
983	Figures	exit

In frame 983, the student's gaze left all three designated AoIs. Playing the video on the Pupil Player, enabling the ability to view the specific frame numbers and fixations, we can see where the next fixations occurred, shown in **Figure** 7. In this situation, the student

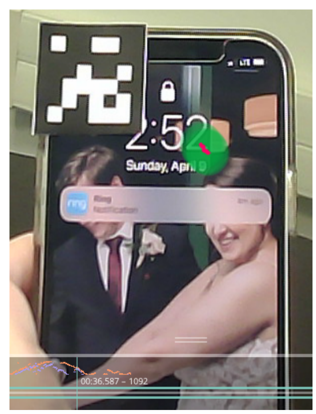

Fig. 7. Where the next fixation took place after leaving the AoIs.

took their attention off the relevant AoIs to check a notification on their phone. An attempt was made to track the phone using a separate AprilTag; however, the AprilTag did not register with the eye-tracker, resulting in the inability to track it as a dedicated surface. Using a methodology similar to the previous point of interest, it is then possible to determine the amount of time spent away from AoIs while distracted by a phone.

4.4 Are There Areas Outside of AoIs That Attract Student Attention?

Ultimately, anything we want to track would be marked with an April tag. As shown in **Figure 7,** phones are often a distraction tool for students. Using the methodology outlined in the appropriate section, there was no good way to collect data pertaining to this point of interest. The student's head was not locked in place, meaning that without a more abundant usage of AprilTags, attempting to detect patterns in fixations using the gaze x and y coordinates would not yield reliable results. Using a similar methodology similar to the previous point of interest would still be possible, but this would be a very tedious process.

4.5 Trends in How and Why Do Students Move Across The Dimensions Outlined in Keller *et al.,* 2020?

Chelazzi *et al.,* 1995, showed that gaze shifts often correlate with a shift in focus [Chelazzi et al. 1995]. Most of the points of interest that were investigated during this study show that there are many different ways eye-tracking can be leveraged in order to discover trends in what students tend to fixate on and how we can track when these fixations end, only to be replaced on something that may or may not be on-topic to the presentation. Namely, the results of the points of interest outlined in subsections 4.2 and 4.3 could be of increased use for attempting to scale the concepts defined in Keller *et al.,* 2020 into a stand-alone project for use in a real classroom. More research would have to be conducted to properly correlate the specific potential findings outlined in this paper to the definitive shifts in attention.

In the next section of this report, the findings outlined with regard to the points of interest will be used to investigate whether an actual study similar to the one outlined in this paper could prove or disprove the hypotheses outlined in Background Work, subsection 2.3.

5 DISCUSSION

The results of this project show great promise for a genuine run. While examining attention and focus is very difficult, especially when only considering the one factor of visual attention, this study showed that reproducing the methods and processes could yield potentially significant results for identifying patterns in visual attention. A future study could then be conducted to determine if the visual attention patterns parallel students' overall attention and focus. However, this sort of study is beyond the scope of just eye-tracking and requires much more emphasis on biology (EEG) and psychology.

For the hypotheses proposed in subsection 2.3, it is still possible to use the “dummy data“ to indicate whether it could verify each hypothesis's validity. For H1, a combination of the methodology from the first and second points of interest could be implemented. Pupil Player exports contain the breakdown of the distribution of gaze points (in gaze_distribution.csv). First, the percentage of gaze points on versus off the presentation is calculated to determine if the number of points between the two presentations is similar. Then, whether or not the student(s) are fixated on the presentation for each frame can be determined and plotted, with a true (one) or false (zero) value for each frame on the Y axis and the individual frames on the X. A trendline will show the general trend of attention across time, and a conclusion can be made to verify H1.

For H2, Slykhuis *et al.,* 2005 reported that students would primarily focus on text and provide, typically, only quick glances at any supplemental figures or images. Gaze path was not an immediate concern with this study; however, with the usage of the vis circle and polyline, in conjunction with surface tracking and using it to determine when a participant's gaze has moved into and out of a specific portion of a presentation, could easily be determined and visual. The settings within the vis polyline allow for a “gaze history,“ or the duration of past gaze included in the polyline. However, this is not a perfect solution. The polyline is very noisy and can be challenging to follow. A Gaussian filter should be applied to the line to reduce the noise, which can be done using Python Pandas[3].

H3 attempts to predict whether the amount of text on each slide would influence the percentage of on-topic and external focus. The third and fourth presentations shown in figures 3 and 4 attempted to demonstrate how this hypothesis could be tested. Many other sections of this report have already elaborated on how to leverage eye-tracking to determine if students are visually focusing on text versus figures versus nothing relevant, so attempting to predict on-topic versus off-topic attention habits in this sense would be trivial. However, the other dimension defined by Keller *et al.,* 2020, external versus internal, could not easily be measured via eye tracking. Simply put, external and internal focus deals more with how the brain filters out information and memory, respectively. It is possible to fixate directly on text on a slide. However, the ability to discern whether or not one's brain is adequately focused on the text and committing that text to memory instead of just looking at it with an “empty head“ is not possible with eye-tracking alone. This is where other similar technology, such as an electroencephalogram, could be required, but as it stands, thoroughly verifying this hypothesis is outside the scope of this project and methodology.

The final hypothesis, H4, attempts to theorize when and how long students give attention to figures. This hypothesis is very similar to the second point of interest detailed in subsection 4.2 and one proposed in a similar study by Slykhuis *et al.,* 2005. Figure 66 shows a mock-up of what this data could look like if collected in an actual study. Overall, claiming that there is a “definitive“ trend could utilize a similar methodology but would require many more participants and span different ways of presenting figures and images. Overall, there is a lot of potential with this methodology/study to answer the calls for future research made by Wang *et al.,* 2020, Slykhuis *et al.,* 2005, and Keller *et al.,* 2020. The next steps would be to put this proof-of-concept into practice with a larger-scale pilot study and then an actual implementation.

6 LIMITATIONS

Given the proof-of-concept nature of this study, limitations did not play a significant factor. Anything considered majorly limiting was removed from the study. However, some components were removed out of necessity but would vastly enhance the study if a means to implement them could be designed. First, one of the points of interest that was dropped from the initial proposal was how the role of a presenter affects student attention. This required the use of Densepose or similar methods of identifying body parts in real-time

[3]https://pandas.pydata.org/

and attempting to see if student gaze ever shifted or changed due to how a presented moved their hands, arms, or body. Densepose is incompatible with the Pupil Core headset, and other human mapping methods, such as Google's MediaPipe[4] has no known existing usage with this type of project while using the Pupil Core eye-tracker. A future study would have to be run to determine the viability of accurately detecting full body or at least arm and hand movement with the Pupil Core eye-tracker before something the abandoned point of interest can be added back to this study.

Furthermore, the most significant limitation of the whole study comes from the use of AprilTags. Entire studies have been conducted on the efficiency and ease of use of AprilTags alone. This project saw many failures with the placement, medium (physical or digital), and the amount of AprilTags used to track the different AoIs. Future iterations of this project would dedicate significantly more time to determining proper and efficient AprilTag usage to help clean up some of the data.

7 CONCLUSION

This project aimed to develop a proof-of-concept for using the Pupil Labs Core eye-tracker to examine student visual attention and focus in the classroom. There were many different points of interest in this study, especially ones that identify patterns or trends in where and when a student is giving attention to a presentation. This study found that while it is possible to conduct a study to investigate the phenomenon of student focus, the choice of only using eye-tracking can be very limiting. Still, eye-tracking in education is not a subfield exclusive to the last few years, and there have been multiple studies that have derived their methodology for leveraging eye-tracking for this use. The researcher welcomes all feedback from the eye-tracking and education research communities on the proposed design and methodology. Feedback can be sent to the author's address, located on the first page.

REFERENCES

Teresa Busjahn, Carsten Schulte, Bonita Sharif, Andrew Begel, Michael Hansen, Roman Bednarik, Paul Orlov, Petri Ihantola, Galina Shchekotova, and Maria Antropova. 2014. Eye tracking in computing education. In *Proceedings of the tenth annual conference on International computing education research*. 3–10.

Leonardo Chelazzi, Monica Biscaldi, Maurizio Corbetta, Andrea Peru, Giancarlo Tassinari, and Giovanni Berlucchi. 1995. Oculomotor activity and visual spatial attention. *Behavioural brain research* 71, 1-2 (1995), 81–88.

Arielle S Keller, Ido Davidesco, and Kimberly D Tanner. 2020. Attention matters: How orchestrating attention may relate to classroom learning. *CBE—Life Sciences Education* 19, 3 (2020), fe5.

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 (2021), 11–26.

David A Slykhuis, Eric N Wiebe, and Len A Annetta. 2005. Eye-tracking students' attention to PowerPoint photographs in a science education setting. *Journal of Science Education and Technology* (2005), 509–520.

Pnina Stern and Lilach Shalev. 2013. The role of sustained attention and display medium in reading comprehension among adolescents with ADHD and without it. *Research in developmental disabilities* 34, 1 (2013), 431–439.

Jiahui Wang, Pavlo Antonenko, and Kara Dawson. 2020. Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education* 146 (2020), 103779.

Received 25 April 2023

[4]https://mediapipe.dev/