

NOVA

IMS

Information
Management
School

MGI

Master Degree Program in
Information Management

Using (Webcam Based) Eye-tracker to Measure Students' Attentiveness and Engagement: Systematic Review

Filipa Pato Rebocho

Master Thesis

presented as partial requirement for obtaining the Master Degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**Using (Webcam Based) Eye-tracker to Measure Students' Attentiveness and Engagement:
Systematic Review**

by

Filipa Pato Rebocho

Master Thesis presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence.

Supervised by

Roberto Henriques, PhD, NOVA Information Management School (NOVA IMS)

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

[Lisbon, 12 July 2024]

ABSTRACT

Eye-tracking technology is recognized as a valuable tool for understanding the cognitive states of students. However, traditional eye-tracking methods are expensive and not scalable for general real-life use. In this context, webcam-based eye-tracking presents a promising alternative due to its affordability and accessibility. Here, using the PRISMA methodology for a systematic review of the relevant literature, we address the research question: "How have studies used webcam eye-tracking technology to measure attention and engagement of students in educational settings, and what are the reported methodologies outcomes?". Our findings indicate that webcam-based eye trackers, while cost-effective, face several challenges related to data accuracy which is still very unstable to what it needs to be for future real-life use, the influence of environmental factors such as lighting that still influence the quality of results, and privacy concerns expressed by the participants of several studies that represent in general the users of this kind of solutions in the future. However, despite these obstacles, the technology still shows substantial promise in improving educational research and practice. Our study highlights the need to refine webcam-based eye-tracking methods further to enhance reliability and utility in academic settings. Future research should continue to explore these avenues to overcome the current limitations and unlock the full potential of this innovative technology. Considering the above, our study provides a solid foundation for future advancements in the field, capacitating future researchers with a comprehensive database of studies with all relevant information already synthesized. This thesis will contribute to more efficiency in future studies.

KEYWORDS

Eye-Tracking; Attention; Human Behaviour; Identify patterns and trends; Webcam

Sustainable Development Goals (SDG):



TABLE OF CONTENTS

Statement of Integrity	
Dedication (optional)	
Acknowledgements (optional)	
Abstract	
List of Figures.....	
List of Tables.....	
List of Abbreviations and Acronyms.....	
1. Introduction.....	1
2. Literature review	4
2.1. Traditional eye-tracking's emergence and its limitations	4
2.2. Webcam-based eye-trackers as an alternative	5
3. Methodology	8
3.1. Protocol development.....	8
3.2. Information sources	9
3.3. Search strategy	9
3.4. Study selection and data collection process	10
3.5. Data synthesis.....	11
3.6. Risk of bias assessment	12
4. Results.....	13
4.1. Document type	13
4.2. Publication.....	13
4.3. Keywords	14
4.4. Funding	16
4.5. Publisher and source	17
4.6. Authors	19
5. discussion.....	20
5.1. Advanced algorithms/hardware enhancements.....	20
5.2. Eye-tracking in user interaction and UX studies	21
5.3. Integration with educational platforms	22
5.4. Open-source software (WebGazer)	23
5.5. Proprietary/custom solutions	26
5.6. Standard webcams without special hardware	27
5.7. Web-based platforms and remote data collection	28
6. Conclusions and future works	29

References.....31
Appendix A – Systematic review: Overview.....36

LIST OF FIGURES

Figure 1.1 - Query used in Scopus for systematic review	9
Figure 2.1 - PRISMA Flow Diagram.....	11
Figure 3.1 - Document type analysis	13
Figure 4.1 - Publication distribution over time	14
Figure 5.1 – Authors keywords word cloud	15
Figure 6.1 – Index keywords word cloud	15
Figure 7.1 – Funding analysis	16
Figure 8.1 – Number of papers per author for the top seven authors	19

LIST OF TABLES

Table 1.1 - Keywords combination for systematic review	9
Table 2.1 - Funding sponsors.....	16
Table 3.1 – Publisher distribution	17
Table 4.1 - Publication sources	17

LIST OF ABBREVIATIONS AND ACRONYMS

ACM	Association for Computing Machinery
AIPS	Artificial Intelligence Proctoring System
CAEs	Convolutional Auto-Encoders
CLM	Constrained Local Model
CNN	Convolutional Neural Network
CSV	Comma-Separated Values
ECIIN	Eye-Context Interaction Inferring Network
IEEE	Institute of Electrical and Electronics Engineers
ITS	Intelligent Tutoring System
LSTM	Long Short-Term Memory
MOOC	Massive Open Online Courses
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RNN	Recurrent Neural Network
TJS	Tracking.js
UX	User Experience
YOLOv7	You Only Look Once, version 7

1. INTRODUCTION

In the contemporary world, characterized by its fast pace and intense competition, developing a deep understanding of how people behave and perform is crucial for success across various domains.

As such, eye-tracking is an invaluable concept in this pursuit, offering an objective lens to comprehend the direction of individuals' attention and what captures their focus. Eye-tracking provides an external perspective on the distribution of attention, thereby bypassing the need for subjective explanations (Semmelmann & Weigelt, 2018).

Nevertheless, despite its promise, the pragmatic use of gaze-based techniques to observe and shape learning in real-world scenarios confronts critical challenges, mainly arising from the substantial costs and the restricted availability of suitable eye-tracking systems (Hutt et al., 2023).

That is why the arrival of webcam-based eye-tracking technology marks a significant transformation in uncovering student engagement patterns. It's not just about making things more efficient. It's about enriching the learning experience, fine-tuning course materials, and nurturing individual growth and performance (Wong et al., 2023).

Nevertheless, at the moment, there exists a multitude of methodologies and software applications for employing webcam-based eye-tracking, but a unified framework or standardized approach is still lacking. This fragmentation may slow down the systematic understanding and effective use of this technology to assess and enhance students' cognitive engagement and learning experiences. Our study steps into the scene to explore such gap, especially within academic settings (Hutt et al., 2023). We're immersing ourselves in the innovative landscape of webcam-based eye-tracking technology, exploring its potential to assess and elevate students' focus. Thus, this study will review existing literature on applying webcam-based eye trackers to measure students' attentiveness and engagement in educational environments.

Following the above, the methodology entails an extensive examination of research papers, documents, and scientific literature related to the use of webcam-based eye-trackers for understanding cognitive processes in educational contexts, giving special emphasis to the ones that investigate attention, as well as other relevant cognitive aspects, such as engagement, comprehension, and performance impact.

In this sequence, our objective will be to synthesize the studies conducted by employing webcam eye-tracking technology to understand students' cognitive capabilities during academic tasks. This will allow us to consolidate and analyse diverse methodologies, identify prevalent challenges, and uncover potential applications of webcam-based eye-tracking in educational contexts. In addition, through this effort, we intend to provide a foundational

groundwork that not only clarifies the current state of knowledge but also provides a roadmap for future research endeavours.

Overall, with this approach, we plan to provide a holistic overview of the methodologies employed, the challenges faced, and the insights gained from diverse applications of webcam-based eye-tracking in education. As such and in summary, by conducting this study and such a comprehensive review, our goal is to (i) contribute with valuable insights that can guide future researchers and studies on the application and effectiveness of webcam-based eye trackers in measuring attention and cognitive capabilities in the academic environment and (ii) foster a positive impact on the learning experience and shape an academic environment attuned to its audience.

However, we already anticipate potential challenges to future studies, including privacy concerns when using webcams to monitor students' gaze (X. Yang & Krajbich, 2021) and the reliability of webcam data, given that it may not always be the optimal choice in terms of quality. Thus, it's essential to gain a deeper understanding of these challenges through thorough examination.

As we dig into the intricate links between how students observe, focus, and do in their studies, it is crucial to emphasize how important this research is for the bigger picture in education.

The usual methods of figuring out how interested students are and how well they're performing often rely on personal judgments or limited timeframes designed for standard evaluation, leading to a narrow path (Kreitzer & Sweet-Cushman, 2022). That's where we expect the benefits of using webcam-based eye-tracking technology to come in – giving us a new way to get around these limitations, offering us a more precise and more detailed picture of how students get into their learning.

In conclusion, our final goal with this study is to make a real difference, not just in academic talks but in the everyday reality of education. The insights we hope to gain from this investigation could inform the creation of designed plans and teaching approaches that match each individual's learning style. It might just revolutionize the entire landscape of how education is conducted and experienced. By tuning into the subtle details of how people focus and perceive information, educators and instructional designers could employ powerful tools to boost the learning experience.

Lastly, and to guide the reader through this study, this thesis is structured as follows. This section, introduction, where we provide an overview, topic background, research gaps, problem statement, objectives, and significance of this study. Section 2 presents the literature review, where we review the story line of eye-tracking to measure students' attentiveness and engagement in education that got us to our study. Section 3 details the methodology, where we describe the use of PRISMA, the research protocol, information sources, search strategy, study selection and data collection process, data synthesis, and risk of bias assessment. Section 4 presents the results, where we present the findings of the study, including types of

documents, publications, keywords analysis, funding sources, publishers, sources, and authors involved in the reviewed literature. Section 5 covers the discussion, where we interpret the findings and discuss their implications, limitations, and potential for future research. Section 6 presents the conclusions and future works, where we summarize the study's contribution, highlight key findings, and propose directions for future research in the field.

2. LITERATURE REVIEW

In the current dynamic and highly competitive educational setting, deciphering the intricate workings of students' cognition emerges as a vital approach to navigating their academic trajectories.

In this context, Wang et al., (2021) emphasize the challenges associated with assessing student attention and engagement within ever-changing learning environments, stressing the urgent necessity to gain a deeper understanding of these aspects. In this sequence, attention, characterized as a cornerstone of human behaviour, takes centre stage in educational research (X. Yang & Krajbich, 2021) which leads these authors to stress the critical nature of attention allocation, often indirectly measured through assumptions from choice data and response times. Overall, attention is not only a marker of engagement but also a fundamental element influencing cognitive approaches pivotal for academic success.

Additionally, and adding another layer to the cognitive landscape, mind wandering, as described by Bonifacci et al., (2023) and Mills et al., (2018), involves a shift in attention from external tasks to internally generated thoughts. This omnipresent phenomenon, often underestimated, poses critical challenges to sustained focus and task-related thoughts (Bixler & D'Mello, 2016). Several studies have estimated that these unintentional shifts occur between 20% and 50% of the time, significantly impacting students' learning experiences (Bixler & D'Mello, 2016).

In light of the above and further investigating the understanding of cognitive capabilities, it becomes clear that a profound analysis of attention and engagement is essential. In this context, eye-tracking technologies have emerged as a promising solution to uncover these particularities and improve the learning experience in our rapidly evolving educational system.

Thus, the following sub-sections will be focused on related work regarding the evolution of eye-tracking technology and its potential to address the challenges posed by these cognitive phenomena.

2.1. TRADITIONAL EYE-TRACKING'S EMERGENCE AND ITS LIMITATIONS

Eye-tracking technology has seen a remarkable journey, evolving into a crucial tool for cracking the complexities of how our minds operate.

Over the past few decades, there have been significant efforts to create eye-tracking technology that is more user-friendly and versatile, elevating its importance in the field (Hutt et al., 2023). These tech improvements have made eye-tracking more widespread and helped us better understand how gaze patterns connect with our thinking (Hutt et al., 2023). Therefore, as pointed out by Hutt et al., (2023), gaze-based detectors hold massive promise as research tools, allowing us to dive into the cognitive factors intertwined with learning and, in a particular way, it's also necessary to recognize that eye-trackers have already become

leaders in creating adaptive solutions that improve the effectiveness of students' educational experiences.

Based on the eye-mind concept, eye-tracking use reveals an engaging story that takes advantage of people's intrinsic propensity to look at data that supports their existing ways of thinking, ultimately offering a singular view into cognitive engagement (X. Yang & Krajbich, 2021). As such, eye-tracking extends beyond its origins in choice research and is used across several distinct psychological fields, including reading and emotion recognition. In addition, it's also applied in non-psychological fields, like advertising and the study of driving behaviour (X. Yang & Krajbich, 2021).

Essentially, eye-tracking – which is defined by Wang et al., (2021) as the meticulous measurement of eye and/or body movements during interactions with learning environments – represents a paradigm shift in our understanding of the complexities of cognitive engagement in education as well as technological advancement. This technology has shown its potential to transform teaching and learning methods, as evidenced by the several studies discussed later in this thesis. Nevertheless, eye-tracking technologies come with many limitations that pose a substantial obstacle to applying these solutions to resolve daily issues and challenges.

In alignment with the preceding idea, using traditional eye-trackers for widespread and real-world application poses a significant challenge given their exorbitant price, consequently making them unaffordable for most people (Wong et al., 2023). For example, Tobii TX300 – a reference for traditional eye-trackers– requires expensive equipment and software expertise for effective deployment (D'Mello et al., 2012) (Bixler & D'Mello, 2016). While these are the most reliable technologies for the purpose under scope, they are generally associated with lab experiences and position themselves as too sophisticated and expensive for large-scale use. Unfortunately, even those that are more accessible, like Eye Tribe and Tobii EyeX, are still quite expensive to employ as real-life solutions (Bixler & D'Mello, 2016).

As a result, there is a persistent necessity for easier, more accessible options, with webcam-based eye-tracking one of the fresh solutions in the market.

2.2. WEBCAM-BASED EYE-TRACKERS AS AN ALTERNATIVE

Through the past few years, webcam-based eye-tracking has arisen as a game changer, attempting to offer a viable alternative to existing methodologies (practical or scalable options) for comprehending cognitive processes. Thus, following the growing need for new solutions, researchers have investigated the potential benefits of using webcam-based eye-tracking technology.

As such, researchers present in (Wong et al., 2023) have been at the vanguard of exploring the validity and applicability of scalable, webcam-based eye-tracking in educational environments, focusing on neurodivergent students. In their study, 43 participants engaged

in reading and answering questions while a webcam-based eye-tracker, WebGazer (Papoutsaki et al., n.d.), assessed their gaze locations. Wong et al., (2023) has concluded that, despite the lower accuracy compared to commercial-grade trackers, WebGazer has demonstrated its capability to replicate findings from various studies, revealing its potential significance in educational research.

In addition, the accuracy and precision of webcam-based eye-tracking have been evaluated across multiple scenarios. Both Semmelmann & Weigelt, (2018) and Hutt et al., (2023) used JavaScript-based eye-tracking algorithms with consumer-grade webcams, having concluded that, despite the acknowledged limitations in resolution and accuracy, these techniques proved applicability for tasks such as fixation (where participants are asked to fixate on a dot that appears on the screen to measure the ability to maintain gaze on a stationary point), pursuit (which involves following the movement of a target stimulus across the screen to assess the ability to track moving objects smoothly with one's gaze, and free vision (where participants view an image of a face and are free to look at whatever catches their attention with the objective of creating an attentional map of the image to reveal which parts draw more gaze and attention). Recognizing the several obstacles, researchers have highlighted the practicality of employing webcams to simulate interior states during learning, pointing at the transformational potential of this technology (Hutt et al., 2023).

As previously mentioned, some researchers have been using WebGazer in their studies in this department. WebGazer is an online eye-tracker designed to work with standard webcams that has been pointed out as a huge step forward in the field. It self-calibrates in real-time by analysing user interactions and mappings of eye features to on-screen gaze positions (Papoutsaki et al., n.d.). In the opinion of the owners and developers of this solution – Papoutsaki et al., (n.d.) – despite its reduced precision when compared to specialized equipment, WebGazer endeavours to deliver a natural user experience during normal web browsing, establishing it as a beneficial tool for usability studies, web research, and real-time gaze interactions.

However, it is critical to recognize that webcam-based eye-tracking is not without obstacles. Researchers have voiced concerns about lower accuracy, reliance on lighting conditions, and potential privacy issues (Wong et al., 2023). Also, some suggest that to maintain acceptable sample rates and data quality, participants should be limited to particular webcam quality levels (Semmelmann & Weigelt, 2018). Furthermore, Wong et al., (2023) stressed the importance of gaining a deeper grasp of WebGazer's limits, especially in ecological situations.

Additionally, we shall not forget that implementing webcam-based eye-tracking in research investigations poses several challenges beyond technological considerations, including programming and web development knowledge. As de Leeuw, (2015) has emphasized, scripting experiments within a browser environment demands a high level of competence, particularly for more complex designs. Therefore, researchers interested in webcam-based eye-tracking must deal with coding difficulties, which require skill to overcome the challenges

of experiment conception and design. Moreover, the difficulties extend to assuring browser compatibility, a vital aspect of every web development project (de Leeuw, 2015). Thus, addressing potential vulnerabilities across multiple browsers becomes critical, influencing the accessibility and usefulness of the technology across a varied audience.

It is safe to say that these programming-related issues and the lack of dedicated tools for browser-based experiments highlight the multidimensional nature of the challenges researchers confront when incorporating webcam-based eye-tracking into their investigations. As such, future efforts to overcome these roadblocks will contribute to the seamless integration and widespread adoption of this breakthrough technology in behavioural research. Despite these challenges, researchers are optimistic about the future potential of webcam-based eye-tracking for real-time, scalable applications, particularly in educational contexts (Wong et al., 2023), as the technology offers a fast, accessible, and potentially more representative approach to studying visual attention, aligning with the increasing popularity of online experiments (X. Yang & Krajbich, 2021).

In light of the above, webcam-based eye-trackers symbolise a paradigm change in cognitive research, providing practical and scalable options for examining cognitive processes. Despite the problems and restrictions we have observed above, current research and potential technology improvements may be the game changers to elevate these technologies for global use and, who knows, offer intriguing prospects for the future of cognitive research and education.

3. METHODOLOGY

As we previously mentioned, the main goal of this study is to conduct a comprehensive review of existing literature on the application of webcam-based eye-trackers to measure attention and engagement in educational environments.

Thus, we will employ the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) methodology (Page et al., 2021) that provides a robust and structured framework for conducting systematic reviews and ensures a transparent and rigorous process, making it an ideal choice for our research goal. In addition, its systematic approach allows for a thorough and unbiased evaluation of the existing literature, contributing to the reliability and credibility of the review.

Overall, PRISMA is particularly suitable for this study for several reasons. First, PRISMA is designed to handle the scope and complexity of systematic reviews, which is crucial for a topic like webcam-based eye-tracking in educational settings. When confronted with a large diversity of study designs, participant populations, and distinct outcomes, there is a need to apply a structured methodology that can manage and synthesize this complexity effectively. Second, PRISMA encourages the inclusion of grey literature, such as conference proceedings and unpublished studies, to mitigate publication bias. This is essential for our research to ensure we consider a broad spectrum of sources, including those that might not be published in traditional academic journals but still provide valuable and interesting insights. Third, PRISMA's detailed protocols for systematic reviews strengthen the reproducibility of research, allowing future researchers to easily add new value to previous studies, which we consider extremely relevant given the continuous growth of this area. Fourth, this framework allows us to ensure consistency in how studies are evaluated and compared which is particularly beneficial when dealing with a multidisciplinary topic like webcam-based eye-tracking where studies might vary significantly in design and execution. Lastly, PRISMA accentuates transparency and rigour in reporting, ensuring that all steps of the review process are documented and reproducible to add credibility to our potential findings.

In this sequence and given the complexity of the topic under scope and the need to analyse a diverse range of studies, PRISMA's structured guidelines will guide the process of identifying, screening, and synthesizing relevant information to achieve a clear and comprehensive overview of the existing research on webcam-based eye-tracking technology in academic settings.

3.1. PROTOCOL DEVELOPMENT

The central research question guiding this review is “How have studies used webcam eye-tracking technology to measure attention and engagement of students in educational settings, and what are the reported methodologies outcomes?” which directs the scope of the review,

emphasising the attention and engagement aspects of student behaviour assessed through webcam eye-tracking technology.

As we need to ensure a targeted and relevant selection of studies, specific criteria will be employed either to include or exclude studies.

Concerning the search strategy to be employed, a systematic search for relevant studies will be conducted using academic databases and relevant search terms related to variations of webcam-based eye-tracking technology and student behaviour and patterns.

Nevertheless, acknowledging potential biases is integral to maintaining transparency and rigour in the review process. As such, publication bias will be considered as positive results are more likely to be published. We'll also explore the possibility of including grey literature such as conference proceedings and unpublished studies to mitigate this situation.

3.2. INFORMATION SOURCES

Concerning information sources for this review, we'll use Scopus given that this database is well known for its comprehensive coverage of academic literature and relevance to the research focus. Also, Scopus includes a large amount of grey literature that will be potentially relevant to this study.

3.3. SEARCH STRATEGY

Developing an efficient search strategy requires identifying relevant terms that capture the core of the research objective. Thus, the following keywords combined with "Eye-tracking" and "Webcam" and "Student" will lead the search for studies related to webcam-based eye-tracking technology and its application in educational settings: "WebGazer", "Attention", "Engagement", "Human Behaviour", "Gaze", "Mind wandering", "Patterns", "Online", "Remote".

Table 1.1 - Keywords combination for systematic review

"Eye-tracking" AND "Webcam" AND "Student" AND	"WebGazer" OR "Attention" OR "Engagement" OR "Human Behaviour" OR "Gaze" OR "Mind wandering" OR "Patterns" OR "Online" OR "Remote"
--	---

```
( TITLE-ABS-KEY ( eye-tracking ) AND TITLE-ABS-KEY ( webcam ) AND TITLE-ABS-KEY ( student ) OR TITLE-ABS-KEY ( webgazer ) OR TITLE-ABS-KEY ( attention ) OR TITLE-ABS-KEY ( engagement ) OR TITLE-ABS-KEY ( human AND behaviour ) OR TITLE-ABS-KEY ( gaze ) OR TITLE-ABS-KEY ( mind AND wandering ) OR TITLE-ABS-KEY ( patterns ) OR TITLE-ABS-KEY ( online ) OR TITLE-ABS-KEY ( remote ) )
```

Figure 1.1 - Query used in Scopus for systematic review

These carefully selected keywords are designed to cast a wide net, capturing studies that explore the nuanced intersections of webcam eye-tracking, student attention and engagement within educational environments.

Nevertheless, as mentioned above, to maintain the precision and relevance of this review, we need to ensure that certain limitations are imposed on the selection of studies.

Thus, as the inclusion criteria, we'll have in consideration (i) studies where webcam-based eye-tracking technology have been used, (ii) studies conducted in educational settings, (iii) studies that focus on the attention and engagement of students and (iv) publications written in English to facilitate comprehensive analysis.

On the other hand, for the exclusion criteria, we'll not consider (i) studies not employing webcam-base eye-tracking, (ii) studies conducted in non-educational environments, (iii) studies that do not focus on attention and engagement and (iv) publications in languages other than English.

3.4. STUDY SELECTION AND DATA COLLECTION PROCESS

In the first part of this systematic review, a comprehensive query (shown in Figure 1.1 - "Query used in Scopus for systematic review") was run in Scopus, generating a dataset of 198 records that were then carefully reviewed using Mendeley Reference Manager to check the existence of any duplicates and ensure the uniqueness of each record.

After concluding that the dataset obtained with the query was unique, it was exported in Scopus as a CSV file, serving as the basis for the primary database's structure.

A careful screening of titles and abstracts followed to address the relevance of each record to the research question, resulting in the inclusion of 67 records and the exclusion of 131 records. In this step, it's important to mention that exclusions were driven by the criterion that if a record appeared relevant based on the title but failed to correspond with the research focus upon detailed examination of the abstract, it was a determinant for exclusion.

The subsequent step was focused on collecting the full text for reports sought for retrieval, with ten reports proving unattainable. Thus, we retrieved 57 reports for eligibility assessment, which involved a comprehensive review of the whole text to evaluate if there is any contribution to our research question and, ultimately, being included in the final output for the systematic review.

During the eligibility assessment phase, reports that did not meet the specific inclusion criteria, namely (i) English language, (ii) mention of attention or engagement, (iii) conducted in educational settings or comparable contexts, and (iv) contain the use of webcam-based eye-tracking technology, were systematically excluded and, as such, we obtained the final set of 33 studies to include in our systematic review.

For an overview of the process mentioned above, see Figure 2.1 – "PRISMA Flow Diagram".

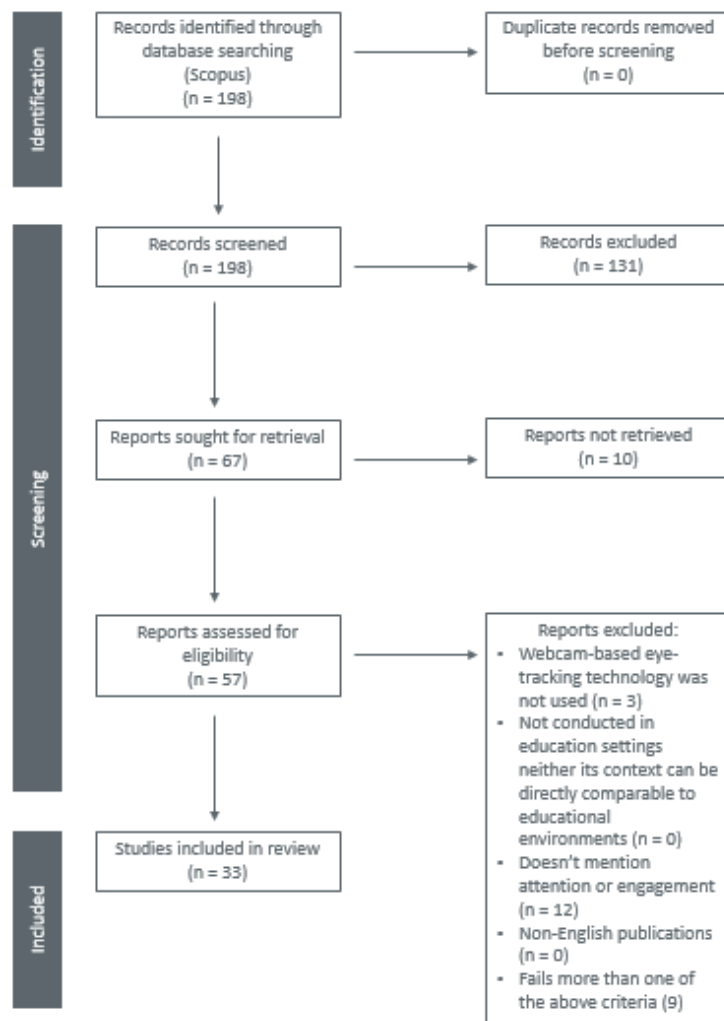


Figure 2.1 - PRISMA Flow Diagram

3.5. DATA SYNTHESIS

To assemble the synthesis of findings from the several studies included in our systematic review, we will provide a comprehensive overview table as output.

Our objective is that this output serves as a condensed repository, offering a panoramic overview of each study's relevance to addressing the research question: "How have studies utilized webcam eye-tracking technology to gauge the attention and engagement of students in educational settings, and what are the reported methodological outcomes?".

This overview table will consist of the following key components:

- Article Reference: This column will display the complete reference for each article, encompassing authorship, publication year, and title. It serves as a unique identifier of each study.

- Sampling Method: We will provide a succinct description of the sampling method employed in each study, shedding light on participant selection strategies, particularly if an experimental dimension is incorporated.
- Number of Participants: This section will include the initial participant count in each study, with specific attention to any subsequent adjustments following exclusions based on eye-tracking performance criteria.
- Stimuli: We will collect the nature of stimuli employed in each study, encompassing educational materials, academic tasks, or any pertinent content shaping the participant experience.
- Equipment: This part will detail the technological equipment leveraged in each study, delineating the specific eye-tracking technology applied.
- Methodology: We intend to present a concise synthesis of the overall research methodology employed in each study, providing a holistic understanding of the study design.
- Key Findings: The crux of each study will be displayed in this column, spotlighting the principal discoveries and outcomes achieved in each study.
- Limitations: This section will encapsulate any limitations recognized by the authors.
- Future Works: Lastly, this column will sum up insights into future avenues of exploration proposed by the authors, offering a path for subsequent research within the same field.

This structured overview table will not only serve as a comprehensive resource for researchers and academics but will also function as a guidepost for future scholars navigating the landscape of webcam eye-tracking applications in educational research.

3.6. RISK OF BIAS ASSESSMENT

Every decision made during this study employing the PRISMA methodology was diligently recorded in the CSV database, serving as the central hub for data management and guidance until the systematic review reached its conclusion. However, it's essential to acknowledge the potential for inherent biases, as the screening process was conducted by a single individual, relying on subjective judgment and perspective. Additionally, the manual screening of data introduces the possibility of human error.

4. RESULTS

In this section, we dive into the core characteristics of the studies we've included in our systematic review to provide insights into the types of documents, publication trends over the years, keyword usage, funding sources, publishers, and collaboration among authors that will help us better grasp the scope of our research.

4.1. DOCUMENT TYPE

As shown in Figure 3.1 – “Document type analysis” below, out of the 33 studies included in our systematic review, 23 were classified as grey literature, primarily comprising conference papers, while the remaining 10 were articles.

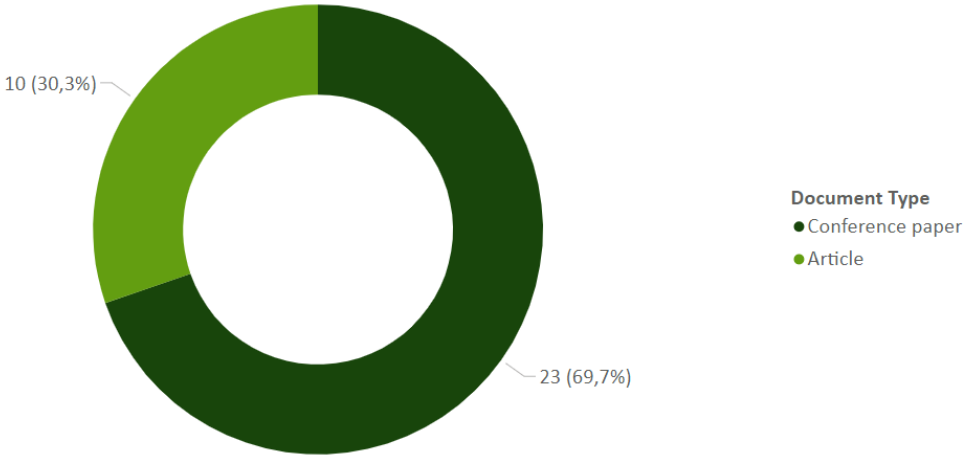


Figure 3.1 - Document type analysis

4.2. PUBLICATION

Using a histogram to visualize the distribution of publications over the years, as stated in Figure 4.1 – “Publication distribution over time”, we observed that 2023 had the highest number of studies, followed by 2022 and 2021. The oldest article in our dataset dates back to 2014, with the two most recent studies published in 2024.

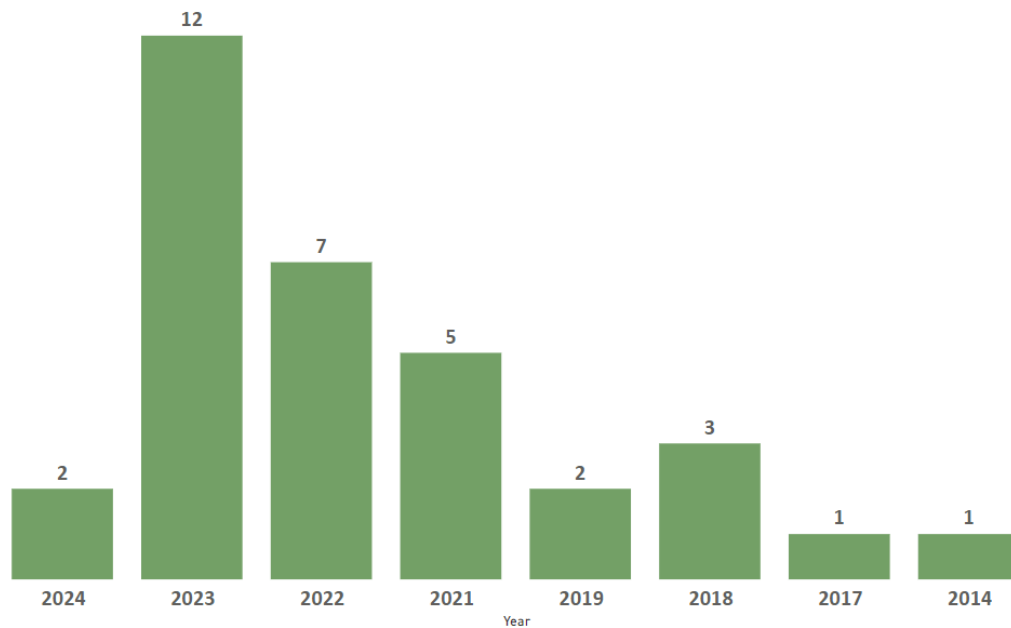


Figure 4.1 - Publication distribution over time

From the above, it is evident that the period from 2021 to 2023 witnessed a peak in research, with 2023 recording the highest number of studies. From our understanding, these results correlate significantly with the global impact of the COVID-19 pandemic, which began affecting worldwide activities in early 2020 and disclosed an urgency in developing efficient remote learning technologies and methodologies to support the learning curve of students.

In fact, this observation was also highlighted by Khan et al., (2022), Tzeng et al., (2023) and Monteiro et al., (2022). All these authors mentioned how the COVID-19 pandemic shaped and changed the learning environment.

4.3. KEYWORDS

Before stepping into the analysis of keywords, it is fundamental to first distinguish between two different terms: the "Authors keywords" and the "Index keywords". The "Author keywords" are defined by the researchers themselves based on the main topics they believe should be mentioned while the "Index keywords" are assigned by database indexers based on a controlled vocabulary and taxonomy to help standardize terms across publications.

Thus, for authors' keywords, and as shown in Figure 5.1 – "Authors keywords word cloud", the most frequently used terms were "Eye tracking", "Webcam" and "Webcam-based eye tracking". All other words and terms were used only once or twice, being more precise to the contours of each study.

4.4. FUNDING

Regarding the analysis of funding sources and as shown below in Figure 7.1 – “Funding analysis”, we can observe that the majority of studies (20) were not funded, while the remaining 13 studies received funding from various sources.

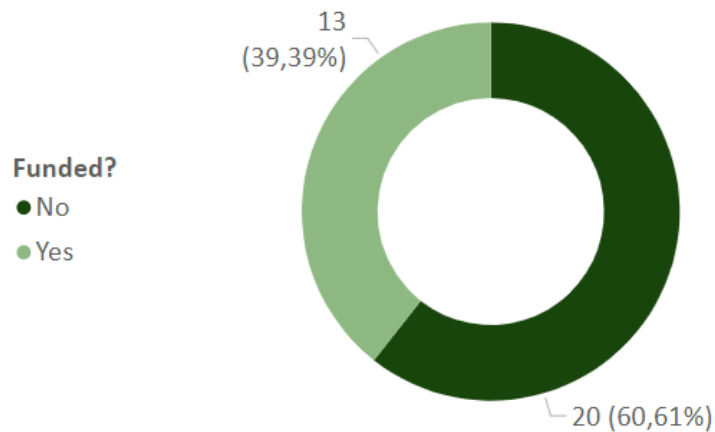


Figure 7.1 – Funding analysis

In this sequence, and regarding the studies that were funded, we can see from Table 2.1 – “Funding sponsors” below that the National Science Foundation was the only sponsor that funded more than one study.

Table 2.1 - Funding sponsors

Funding Sponsor	Documents
National Science Foundation	3
Arnold and Mabel Beckman Foundation	1
China Academy of Space Technology	1
Deutsche Forschungsgemeinschaft	1
Economic and Social Research Council	1
European Commission	1
Horizon 2020 Framework Programme	1
National Natural Science Foundation of China	1
Science and Technology Commission of Shanghai Municipality	1
Technion-Israel Institute of Technology	1
Tromsø Forskningsstiftelse	1
Tsinghua University	1
University of Illinois System	1
University of New Hampshire	1

4.5. PUBLISHER AND SOURCE

From Table 3.1 – “Publisher distribution”, we are able to observe that the Association for Computing Machinery (ACM) published the highest number of studies (12), followed by Springer Science and Business Media Deutschland GmbH (5), Institute of Electrical and Electronics Engineers Inc. (IEEE) (4) and Springer (4).

Table 3.1 – Publisher distribution

Publisher	Publications
Association for Computing Machinery	12
Springer Science and Business Media Deutschland GmbH	5
Institute of Electrical and Electronics Engineers Inc.	4
Springer	4
IEEE Computer Society	2
Blue Eyes Intelligence Engineering and Sciences Publication	1
Frontiers Media S.A.	1
Institute of Advanced Engineering and Science	1
Society for Judgment and Decision making	1
Springer New York LLC	1
Springer Verlag	1

Additionally, from Table 4.1 – “Publication sources” below we can see that most studies were published and discussed in international conferences focused on specific areas such as learning analytics, eye-tracking research, and intelligent user interfaces. In addition, this information strongly indicates that there is an increasing interest and ongoing developments in these specialized fields.

Table 4.1 - Publication sources

Publisher	Source
Association for Computing Machinery	13th International Conference on Learning Analytics and Knowledge: Towards Trustworthy Learning Analytics, LAK 2023
Association for Computing Machinery	14th International Conference on Education Technology and Computers, ICETC 2022
Association for Computing Machinery	15th Annual ACM Symposium on Eye Tracking Research and Applications, ETRA 2023

Publisher	Source
Association for Computing Machinery	2021 ACM Symposium on Eye Tracking Research and Applications, ETRA 2021
Association for Computing Machinery	2022 ACM Symposium on Eye Tracking Research and Applications, ETRA 2022
Association for Computing Machinery	2023 CHI Conference on Human Factors in Computing Systems, CHI 2023
Association for Computing Machinery	23rd ACM International Conference on Intelligent User Interfaces, IUI 2018
Association for Computing Machinery	24th ACM International Conference on Multimodal Interaction, ICMI 2022
Association for Computing Machinery	2nd ACM SIGIR Conference on Information Interaction and Retrieval, CHIIR 2017
Association for Computing Machinery	7th Joint International Conference on Data Science and Management of Data, CODS-COMAD 2024
Association for Computing Machinery	Proceedings of the ACM on Human-Computer Interaction
Blue Eyes Intelligence Engineering and Sciences Publication	International Journal of Recent Technology and Engineering
Frontiers Media S.A.	Frontiers in Psychology
IEEE Computer Society	13th IEEE Global Engineering Education Conference, EDUCON 2022
IEEE Computer Society	2014 19th IEEE International Conference on Automation, Quality and Testing, Robotics, AQTR 2014
Institute of Advanced Engineering and Science	IAES International Journal of Artificial Intelligence
Institute of Electrical and Electronics Engineers Inc.	2019 Portland International Conference on Management of Engineering and Technology, PICMET 2019
Institute of Electrical and Electronics Engineers Inc.	2023 International Conference on Consumer Electronics - Taiwan, ICCE-Taiwan 2023
Institute of Electrical and Electronics Engineers Inc.	2023 International Symposium on Networks, Computers and Communications, ISNCC 2023
Institute of Electrical and Electronics Engineers Inc.	2nd Asian Conference on Innovation in Technology, ASIANCON 2022
Society for Judgment and Decision making	Judgment and Decision Making
Springer	Behavior Research Methods
Springer New York LLC	
Springer Science and Business Media Deutschland GmbH	13th Annual Information Systems and Neuroscience, NeuroIS 2021
Springer Science and Business Media Deutschland GmbH	17th International Conference on Intelligent Tutoring Systems, ITS 2021

Publisher	Source
Springer Science and Business Media Deutschland GmbH	3rd International Conference on Artificial Intelligence in HCI, AI-HCI 2022 Held as Part of the 24th HCI International Conference, HCII 2022
Springer Science and Business Media Deutschland GmbH	6th International Workshop on PRedictive Intelligence In MEDicine, PRIME 2023
Springer Science and Business Media Deutschland GmbH	Proceedings of the 7th International Conference on Advances in Computing and Data Sciences, ICACDS 2023
Springer Verlag	20th International Conference on HCI, HCI International 2018

4.6. AUTHORS

From Figure 8.1 – “Number of papers per author for the top seven authors”, we concluded that seven authors collaborated in more than one study. Specifically, each author has been involved in two studies each, except for one author who has participated in three studies.

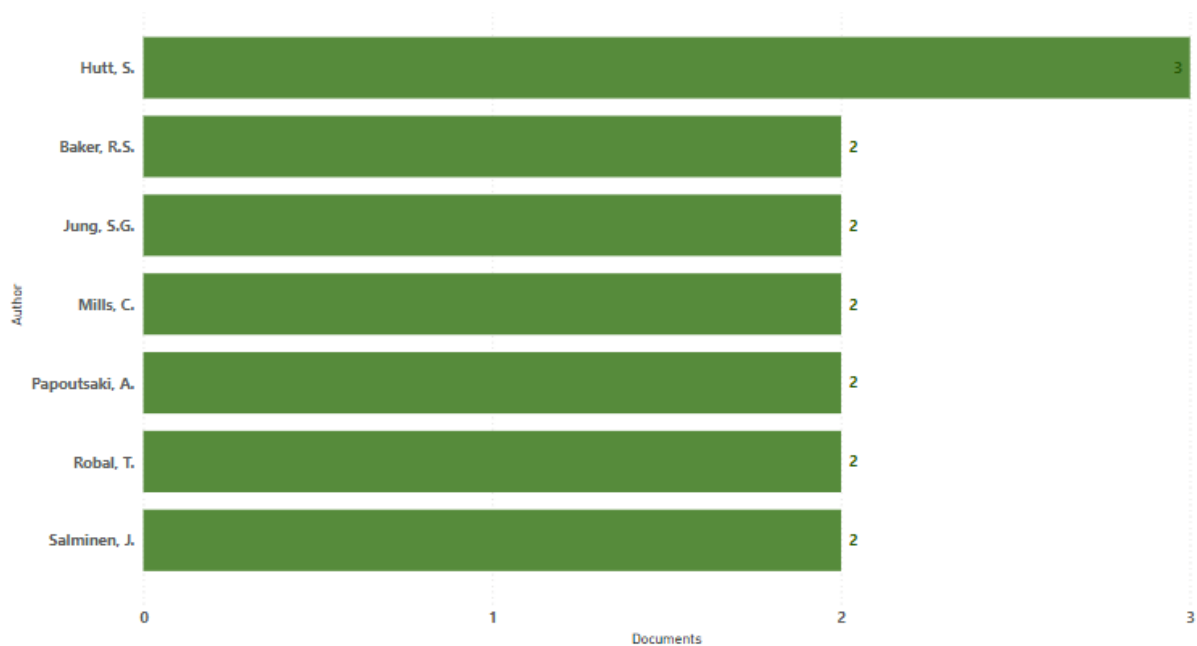


Figure 8.1 – Number of papers per author for the top seven authors

5. DISCUSSION

In this section, we deconstruct the studies included in our systematic review to get comprehensive deeper insights into topics such as sample procedures, participants included in the experiments, stimuli used, equipment deployed, methodology employed, major findings and potential limitations, and recommendations for future research.

As previously stated, our goal is to provide a comprehensive resource for researchers and academics, as well as a roadmap for future scholars investigating the use of webcam eye-tracking in educational research.

5.1. ADVANCED ALGORITHMS/HARDWARE ENHANCEMENTS

In the examination of sophisticated algorithms and hardware advancements, the selected studies show great progress in exploiting webcam eye-tracking technology to improve the educational experience, some directly, others in comparable environments.

Overall, all of these studies address several difficulties of user interaction, using cutting-edge approaches to give more insights into important topics like student attention and engagement.

By gathering over 11,000 photos from participants (individuals comfortable with online and offline shopping) for training convolutional auto-encoders (CAEs), Dharbaneshwer et al., (2024) explored the efficiency of CAEs and ridge regression models in creating a real-time eye-tracking system. Their experiment pointed strong potential in spotting customer attention on e-commerce platforms, emphasizing the appeal of central product placement for customers. Nonetheless, despite the success achieved, the system revealed some limitations, mainly related to individual eye variations and shifting lighting conditions, suggesting that future studies should focus on reducing the requirement for calibration by employing improved interpolation techniques.

On a distinct but similar venue, Gershov et al., (2023) moved inside the operating room and captured anaesthesiologists' activities via camera to better understand visual attention allocation during procedures. While the study did not specify the sampling strategy, it did use deep learning technologies, specifically YOLOv7 and ECIIN, to distinguish between regular and crucial task attention patterns, showing off as a promising and accurate route for recognizing on-focus behaviour. However, it is relevant to highlight that the study had a limited dataset, which may restrict the findings' generalizability. Additionally, there was a significant emphasis on certain surgical stages. In this sense, the authors recommend that future research might broaden this to include more clinical settings and longer-term studies evaluating patient safety and healthcare quality.

By employing techniques such as Haar cascade classifiers and Support Vector Machines complemented by advanced image segmentation methods, Deivanayagi et al., (2019) addressed the challenge of detecting and tracking eyes without specifying the stimuli or the number of participants. As a result, the authors showed that it's possible to track where the pupil moves accurately and in real-time, and even figure out emotions. However, they ran into challenges associated with potential false positives in categorizing eye states, especially when the surroundings looked different. Thus, as per our understanding, we might assume there is room for improving segmentation techniques and boosting the system's robustness.

Robal et al., (2018) offered a different and practical approach to online learning behaviour by recruiting graduate students and staff from TU Delft to simulate common online learning behaviours and evaluated two software-based frameworks. Unfortunately, their study indicated that tracking.js (TJS) significantly outperformed CLM combined with WebGazer in detecting attention, with Tobii hardware - as expected - providing the highest accuracy. Additionally, Robal et al., (2018) noted several challenges linked to delay times and accuracy affected by background noise, which led them to suggest that future studies should focus on applying these findings in large-scale MOOC environments and their impact on learning outcomes.

Overall, the above studies show a big step forward in how innovative technology and algorithms are moving to improve webcam eye trackers for education and related environments.

In general, these studies still open doors to potentially transform how students' engagement might be assessed and understood despite the limitations noted by the above authors.

5.2. EYE-TRACKING IN USER INTERACTION AND UX STUDIES

In this section, we intend to provide a critical synthesis of research that combines the fields of user experience (UX) with educational technology, using webcam eye-tracking to assess and improve student engagement in digital learning settings.

Wisiecka et al., (2022) experimented with 83 participants to crack attentional biases through emotional stimuli by examining eye movements in varying conditions of eye-tracking, which allowed them to discover a "happiness-superiority effect" where subjects were faster to locate happy faces. For this reason, the authors concluded that webcam eye-tracking can be considered a reliable tool in relation to traditional methods, although less accurate, suggesting exploring real-world applications and refining alerting systems to improve UX in live settings.

On a similar path, Akshay et al., (2023) experimented with 25 participants (20 to 35 years old) familiar with technology and online platforms. Here, the participants were instructed to navigate the UX of e-commerce sites to study product identification and advertisement engagement, which allowed the authors to uncover the dominance of centre-screen interest. Although promising, the authors observed some limitations associated with lighting

conditions and potential inaccuracies from webcam-based data collection, leading to suggestions for future research on improving calibration techniques to expand the utility of webcams to general and real-life situations.

In three insightful studies via Mechanical Turk, Papoutsaki et al., (2017) used webcams to track and predict gaze behaviour. From examining search result engagement to ad interaction and restricted focal views, the studies demonstrated the complex ways people interact. In addition, the research pointed out how well users adjust to the limits of interfaces and how differently they focus on search engine results, and even though working remotely can lead to quicker task completion and more distractions, the authors concluded that this approach gives a valuable look into how people behave and interact in real life. In this sense, the authors suggest that future studies might focus on improving gaze prediction algorithms for webcams, thereby improving the accuracy and application of eye-tracking in user studies.

Finally, Vaida, M.-F., & Pop, P. G., (2014) did not share how many people participated in their study, which integrated eye-tracking with a Content Management System to watch how students with different Enneatypes interacted in an e-learning setting. In this combined system, the authors were able to correlate eye-tracking data with varying learning styles, and findings suggested that forming groups based on personality traits works well, leading to better communication and more productivity in team projects considered in the study. Nevertheless, the authors believe there is scope for optimization given that the system's performance was only regarded as satisfactory. The authors also suggested the introduction of "Footprints," a new grouping algorithm that might refine the creation of groups, enhancing the learning and collaborative experience.

Overall, these studies represent a strong argument for combining eye-tracking and UX design, highlighting the importance of constantly improving and exploring their appliances to educational technology.

5.3. INTEGRATION WITH EDUCATIONAL PLATFORMS

In the domain of educational technology, integrating webcam eye-tracking technology with educational platforms has opened new avenues for assessing and enhancing student engagement and attention. This synthesis of research findings examines how such integrations have been explored to improve the educational experience.

Bogdan et al., (2023) used a robust sampling method at a local university. Participants took part in tasks that looked at how emotions and attention interact, doing these tasks face-to-face using infrared eye-tracking or online with webcam tracking. The studies used specially made images to explore the complicated relationship between emotions and where people look. They found, from a big group of participants in both studies, that online eye-tracking works well and can give results similar to in-person studies, even if the effects were smaller. One major limitation observed by the authors was that webcam tracking wasn't always precisely right in pinpointing locations, which could make the results of psychology studies

less accurate. As such, they suggested looking more into how reliable webcam tracking is for psychology research and figuring out how much this location error affects the outcomes of studies. Overall, the authors understand that this area of research highlights how eye-tracking can be helpful for educational websites, especially when moving from classroom to online learning. It also points out the importance of looking closely at the "centering bias" and how it affects studies on where people focus their attention.

Tzeng et al., (2023) did not get into the details about how they chose their study participants or how many there were. Instead, they focused on introducing a new system called the Artificial Intelligence Proctoring System (AIPS). Shortly, this system uses where people look and how their eyes move to spot unusual patterns when students take tests online. In this sequence, the study looked into making online exams fairer by trying out a technology called the Long Short-Term Memory (LSTM) model. Actually, Tzeng et al., (2023) found out that it's possible to tell the difference in how much and how often students of different skill levels pay attention using wearable eye trackers. However, they are pretty vague about the limitations encountered and suggestions for future works. From this, we can speculate that there's room to make AI proctoring systems more reliable and to dig deeper into the ethical questions these technologies bring up in schools and colleges.

To sum up, we believe that the above studies have not only demonstrated the significant contributions and stressed the challenges of applying webcam-based eye-tracking in educational platforms but also offered new routes for future research to improve online learning assessment methods and explore in more detail the relationships between students' attention, engagement, and emotional responses in digital learning spaces.

5.4. OPEN-SOURCE SOFTWARE (WEBGAZER)

From educational environments to cognitive and behavioural studies, and despite several limitations we will observe below, WebGazer has been doing impressive work in demonstrating its utility across diverse research contexts.

In 2023, Wong et al., (2023) proved that WebGazer could pick up on moments when learners with neurodiversity were having a tough time understanding something. As such, they emphasized how this technology could help create instant help specific to each person's needs. When thinking about students, especially students with learning-specific needs, it is true that this concept is promising and would create several benefits. However, compared to the accuracy of commercial eye trackers and real-world situations, Wong et al., (2023) stress that it is necessary to perform more studies to trust WebGazer's precision.

Moreover, Ding et al., (2023) achieved high accuracy in detecting unknown words in ESL reading material using GazeReader. Nonetheless, this experiment relied on a small dataset, which may impose limitations on extrapolating these results for real-world applications. Thus, Ding et al., (2023) recommend that future research explore testing the technology across

different computer systems. In addition, the authors also suggest contemplating more reading materials.

To support online learning and respect users' privacy, Khan et al., (2022) developed a new system for eye-tracking, the EXECUTE. Notably, EXECUTE achieved an accuracy of over 91% for classifying student attention. Yet, Khan et al., (2022) spotlighted some limitations associated with how often standard webcams record data and proposed future investigations into eye movement patterns to help spot ADHD early and create learning methods personalized to each person.

In a different scenario, Toreini et al., (2021) explored the difficulties people face with Terms and Conditions (TC) in the real world. With their experiment, they could observe increased attention to TC when participants received EyeTC-based feedback (EyeTC stands for the application developed in the experiment). Even though EyeTC proved valuable, the authors acknowledged that using mouse clicks instead of actual eye movement data could limit findings. Thus, Toreini et al., (2021) recommended using this technology in real-life situations to see how well it works.

Differently, targeting Russian native speakers as participants Prystauka et al., (2023), using WebGazer proved that webcam-based eye tracking can capture sentence processing effects in real-time with sufficient granularity, even for subtle linguistic phenomena. They pointed out differences in how they analysed their data, applying two approaches to computing statistical power, showing the need to keep improving how people do research like this in the future.

In addition, Van der Cruyssen et al., (2023) limited their experiment to English-speaking participants who did not wear glasses and had not participated in their pilot study, managing to replicate known cognitive phenomena like the cascade effect, the novelty preference and the visual world paradigm. However, distinct from their expectations, the results obtained were slightly weaker and not as pronounced as what has been observed in previous studies.

In their study, Calabrich et al., (2021) showed that adults with dyslexia have trouble connecting sight and sound, but using memory cues can help them learn from what they see and hear together. However, the authors stated the need for larger sample sizes to be able to actually go into the complex details of cross-modal binding, especially when looking at different types of dyslexia.

Similarly, Ikerman & El Mouatasim, (2023) gathered data from students across three primary schools in southern Morocco and found out that certain fonts designed for people with dyslexia can really help improve their reading. Yet, as it happened in the previous study, the authors note that the study had a small number of participants and used regular webcams, which aren't very precise, recommending that future research should keep these points in mind to improve the assessment of dyslexia.

Guan et al., (2022), in turn, it is focused on predicting reading comprehension. Similarly to other studies, this study also stressed possible limitations due to the small sample size of the experiment conducted, recommending larger-scale applications for future research. Additionally, the authors recognize potential fragilities in the study because the analysis related to text content was performed manually, which could be a constraint for scaling up the application of the platform.

Hutt et al., (2024) found a link between the predicted rates of task-unrelated thoughts (TUT) and the actual scores on comprehension tests with their study. Even though there was an actual effort to mitigate bias, the authors once again stated the need for larger datasets. Moreover, Hutt et al., (2024) also mentioned the need to employ a more complex feature engineering approach in future research.

Yang & Krajbich, (2021) combined WebGazer with studies on behaviour, tracking how people's looking patterns relate to their effective decisions. In the experiment, the authors faced challenges controlling the participants' surroundings. Consequently, they suggested that future research should try using smarter algorithms and more areas of interest to get more accurate results.

To decipher user engagement and obtain valuable insights, Jung et al., (2021) integrated a webcam-based eye-tracking module into an interactive persona system. Although their development sounds promising, the accuracy of tracking eyes online can vary a lot. As such, the authors recommended conducting more studies on adjusting these systems better and including tools for visualizing data right into the user interface.

(Robal, 2019) a study comparing WebGazer and Tobii pointed out a big issue with how accurately these systems could be set up. Indeed, the study suggested that letting users calibrate the system themselves could make it more accurate. Also, future research should work on fixing how the clmtrackr library in WebGazer tends to track faces too aggressively.

Finally, Semmelmann & Weigelt, (2018) looked into how accurate eye-tracking is for both fixation and pursuit tasks and found more inconsistency and errors in online data, suggesting that making tasks more gamified with automatic setup might make people more engaged and enhance eye-tracking accuracy online.

In light of the above, the previous studies provided a collection of shared limitations like accuracy, calibration, and the influence of environmental variables. In addition, and consistently, studies reflected on improving the technology, using it in more situations and with more types of people, and getting the most out of remote eye-tracking for learning and broader studies on people's attention and engagement.

5.5. PROPRIETARY/CUSTOM SOLUTIONS

Here, we will continue exploring different creative ways to use webcams to track where people look. From the studies we will look at, it is clear that these methods are starting to change how we employ webcam eye-tracking for specific tasks. As such, we will next expose how eye-tracking has been integrated into more sophisticated and domain-specific applications, plus other technologies associated with emotion recognition and advanced data processing.

In 2023, Rogers & Al Madi, (2023) developed software that uses a tool to track where people are looking and another part that understands emotions. Overall, they did this to determine how developers feel about different parts of the code. In this sense, the authors concluded that this approach allowed them to see what parts of the code caught developers' attention and how these parts made them feel, offering insights into the cognitive thinking and emotional processes involved in software development. Lastly, even though the study is still in its early stages, authors pointed to future large-scale applications with a plan to validate its eye-tracking module's accuracy in diverse working environments.

Hutt & D'Mello, (2022) chose to use data from three different learning settings to create a new algorithm capable of figuring out where someone is looking without the usual setup steps. Although their approach correlated well with traditional tracking methods and also performed well in predicting engagement and comprehension, it was not as strong at noticing quick eye movements as the authors hoped, which led them to recommend that future efforts should focus on improving the clustering approach and ensuring adaptability to various learning contexts.

Adiani et al., (2022) went on a distinct path from other authors and decided to test a webcam-based eye-tracking system within virtual job interview simulations, which showed a way to accurately identify areas of interest after setting up, as long as the person's head position didn't change. Also, it is relevant to mention that several challenges arose during their experiment, in particular associated with maintaining consistency and reflective interference from eyeglasses.

Monteiro et al., (2022) passed on providing details on findings or limitations in their study; however, they developed a system employing CNN and RNN algorithms for proctoring online exams - the "Proct-Xam", leveraging webcam feeds to ensure exam integrity. This system became a new development where eye-tracking could help ensure students stay honest during exams. Future improvements will probably aim to make its live monitoring more accurate and reliable.

S. Yang et al., (2023) developed a study on beginner programmers using real-time editing tools, which showed that these tools helped with awareness and significantly improved learning results and the way teaching evolved. Nevertheless, the authors admitted that eye-

tracking is not perfect and recommend that future research should look into how sharing where you're looking in real time can help online learning.

Anwar et al., (2018) developed the EREG system that demonstrated high accuracy in emotion recognition and eye gaze estimation. Like in other studies we have already explored, even though the system showcased a high correctness score, reliance exclusively on a webcam might present challenges associated with variable conditions.

To sum up, we understand that these studies have opened doors for future research to make eye-tracking more accurate and more robust in different conditions and blend it into more complex systems. Looking ahead, we understand that it will be likely to see these solutions become more common, easier to use, and better suited to what people specifically need, helping to connect technology closer to how we think and act.

5.6. STANDARD WEBCAMS WITHOUT SPECIAL HARDWARE

Despite certain technological constraints, the studies under this category offer a rich panorama of methodologies and insights into user engagement and cognitive processes.

Azem et al., (2023) introduced the METRIC system to measure user engagement remotely through basic webcam interaction and an eye-tracking JavaScript library. After their experiment, the authors reported that METRIC's range of functionalities helped analyse participant gaze patterns mainly due to its ability to handle both video and image materials. The fact that the system was seen as user-friendly was valued, but participants highlighted the need for more intuitive guidance for users, especially for setting up multiple user studies. In light of the above, the authors believe that the future direction for METRIC includes improving the accuracy of the remote eye tracking tool and incorporating more user study features such as audio recording and transcription.

Alternatively, Sodoké et al., (2021) focused on the medical field, specifically anaesthesiology. After recruiting clinicians with varying levels of expertise from the Université de Montréal, participants engaged with the CICO algorithm via a video simulation on the EyeLab platform. The authors found that a webcam-based Intelligent Tutoring System (ITS) could reinforce the learning stages of novice clinicians by simulating the implicit knowledge of expert visual attention. Nevertheless, this experiment was still affected by the lower precision of webcam-based eye-tracking compared to traditional lab-based equipment and the small sample size. Lastly, for future work, the authors intend to re-do the experiment with more participants to evaluate behavioural learning gains more extensively.

In essence, both these studies came to the same conclusion as other studies already scrutinized – employing standard webcams for eye-tracking offers significant benefits in accessibility and applicability across various settings. However, these solutions are still less precise than commercial and specialized eye-tracking hardware. Notwithstanding, as technology advances, we anticipate further improvements in webcam-based eye-tracking

methods and, who knows, these solutions may be accurate enough to be scalable in the near future.

5.7. WEB-BASED PLATFORMS AND REMOTE DATA COLLECTION

Using the Prolific platform, Sauter et al., (2022) explored how matching eye movements could indicate how much students were paying attention and predict their understanding of the material in video lessons. After exclusions associated with calibration and attention check performance, their experiment included 27 participants and sadly found no significant correlation between eye movement synchronicity and test scores. These results mean that although webcams can give us clues about how engaged learners are, there are important limitations, especially with how reliable the eye-tracking data is and how often it's collected, because of the limits of regular webcams.

Enhancing the precision of eye-tracking algorithms and validating attention measures in real-life settings - similarly to other studies, this is also the recommendation of Sauter et al., (2022).

In addition, the influence of a visible presenter on learning, a secondary aspect of the study, remains an intriguing avenue for future research, suggesting that such visual elements may affect user engagement and offer the potential for improving online educational practices.

To view the comprehensive overview table of all studies in this systematic review, please consult Appendix A – Systematic Review: Overview. This table serves as a consolidated repository, offering detailed insights into each study's sampling methods, number of participants, stimuli used, equipment employed, research methodologies, key findings, limitations, and suggestions for future research. By systematically categorizing this information, this table provides a structured summary that supports the deeper analyses presented above in this discussion chapter.

6. CONCLUSIONS AND FUTURE WORKS

The motto for the present study was the increasing integration of digital tools in education and the need for methods that accurately assess and support student engagement in these kinds of environments. As such, here we performed a systematic review that had as a central guiding research question: “How have studies used webcam eye-tracking technology to measure attention and engagement of students in educational settings, and what were the reported methodologies outcomes?”.

The main findings of this systematic review showed that webcam-based eye-tracking can be a viable alternative to more traditional methods of eye-tracking in real-time analysis of attention and engagement patterns, given it is more accessible and less invasive when it comes to monitoring students. In general, some authors have been using advanced algorithms or have developed hardware enhancements. Others have been exploring webcam eye-tracking in user interaction and UX studies. Others decided to integrate webcam eye-tracking in educational platforms or have constructed custom solutions. Nevertheless, most authors have been using WebGazer as the basis for their experiments and developments.

Furthermore, we understand that this systematic review will be relevant in supporting further studies and investigations on employing webcam eye-tracking technology in the academic world. Also, once this technology is proven stable enough for real-life situations, we believe that the potential practical applications could be significant given that educators will be able to use the extracted insights to adapt their instructional strategies to better align with the needs of students, which in turn will lead to a more enriched academic experience for the students.

However, we could also observe that most of the studies included in the systematic review reflected on how this technology is not yet stable enough to be scalable due to several limitations identified during experiments. Consistently, studies identified several issues and inconsistencies associated with data accuracy, the influence of external factors like lighting, and privacy concerns stressed by the participants of experiments.

In our opinion and based on our research, there are still aspects that need to be addressed and improved to reach the end of the road. Thus, based on our systematic review, future research should aim to improve the accuracy and reliability of webcam-based eye-tracking systems through technological progress and evaluate these systems' long-term effectiveness across diverse academic environments and types of students. The authors steadily stressed the need for larger sample sizes in experiments to extract insights closely related to real-life scenarios.

At the starting point of this thesis, our primary goal was to provide a broad and complete assessment of how studies have used webcam eye-tracking technology to measure the attention and engagement of students in academic settings. We consider we have achieved

this by conducting a systematic review employing PRISMA methodology, where we have identified and synthesized a diverse range of studies, contributing, this way, to future studies that wish to experiment with webcam-based eye-tracking in education settings. Our systematic review highlighted the versatility and potential of this kind of technology and in terms of practical applications, showed that there are several efforts on the field to effectively obtain more stable and effective results and get to a point where this technology can be deployed to production and to the real world. Overall, our systematic review provides a solid foundation for future advancements in the application of webcam eye-tracking technology in education. Researchers will now have a comprehensive database of studies conducted in this field, allowing them to understand more effectively which are the possible venues for further research and improvement.

Altogether, this thesis stresses the importance of technology in improving human experiences – in this case, the students' academic experience. Webcam-based eye-tracking technology offers promising capabilities in this regard, though its broader application will require ongoing refinement and research. The future is just one step away from the present, and technology will be a constant in everyone's life – the best we can do is to walk hand in hand with it.

REFERENCES

- Adiani, D., Qu, C., Gass, T., Gurram, S., LeMay, D., Bhusal, A., Sarkar, M., & Sarkar, N. (2022). Evaluation of Webcam-Based Eye Tracking for a Job Interview Training Platform: Preliminary Results. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13336 LNAI, 337–352. https://doi.org/10.1007/978-3-031-05643-7_22
- Akshay, S., Shukla, A., & Raman, V. K. (2023). UEye: Insights on User Interface Design Using Eye Movement Visualizations. *Communications in Computer and Information Science*, 1848 CCIS, 253–264. https://doi.org/10.1007/978-3-031-37940-6_21
- Anwar, S., Milanova, M., Abdulla, S., & Svetleff, Z. (2018). Emotion recognition and eye gaze estimation system: EREG. *Communications in Computer and Information Science*, 851, 364–371. https://doi.org/10.1007/978-3-319-92279-9_49
- Automation, Quality and Testing, Robotics, 2014 IEEE International Conference on.* (n.d.).
- Azem, J. Y., Salminen, J., Jung, S. G., & Jansen, B. J. (2023). Measuring Engagement Through Remote Interactions of Customers: Introducing METRIC. *2023 International Symposium on Networks, Computers and Communications, ISNCC 2023*. <https://doi.org/10.1109/ISNCC58260.2023.10323669>
- Bixler, R., & D’Mello, S. (2016). Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, 26(1), 33–68. <https://doi.org/10.1007/s11257-015-9167-1>
- Bogdan, P. C., Dolcos, S., Buetti, S., Lleras, A., & Dolcos, F. (2023). Investigating the suitability of online eye tracking for psychological research: Evidence from comparisons with in-person data using emotion–attention interaction tasks. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-023-02143-z>
- Bonifacci, P., Viroli, C., Vassura, C., Colombini, E., & Desideri, L. (2023). The relationship between mind wandering and reading comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, 30(1), 40–59. <https://doi.org/10.3758/s13423-022-02141-w>
- Calabrich, S. L., Oppenheim, G. M., & Jones, M. W. (2021). Audiovisual Learning in Dyslexic and Typical Adults: Modulating Influences of Location and Context Consistency. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.754610>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>

- Deivanayagi, S., Nandhini Sri, V. G., Kalai Priya, P., & Aarthi, G. (2019). Pupil detection algorithm based on feature extraction for eye gaze. *International Journal of Recent Technology and Engineering*, 8(2 Special issue 5), 73–76. <https://doi.org/10.35940/ijrte.B1016.0782S519>
- Dharbaneshwer, S. J., Gayathri Sowmya, G., Chauhan, S. S., Shekhawat, B. S., Kumar, L., & Ghosh, S. (2024). Real-time eye tracking using representation learning and regression. *ACM International Conference Proceeding Series*, 298–306. <https://doi.org/10.1145/3632410.3632416>
- Ding, J., Zhao, B., Huang, Y., Wang, Y., & Shi, Y. (2023, April 19). GazeReader: Detecting Unknown Word Using Webcam for English as a Second Language (ESL) Learners. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3544549.3585790>
- D’Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human Computer Studies*, 70(5), 377–398. <https://doi.org/10.1016/j.ijhcs.2012.01.004>
- Gershov, S., Mahameed, F., Raz, A., & Laufer, S. (2023). More Than Meets the Eye: Analyzing Anesthesiologists’ Visual Attention in the Operating Room Using Deep Learning Models. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 14277 LNCS, 253–264. https://doi.org/10.1007/978-3-031-46005-0_22
- Guan, X., Lei, C., & Feng, X. (2022). The Predictive Effect of Eye-Movement Behavior on Reading Performance in Online Reading Contexts. *ACM International Conference Proceeding Series*, 470–475. <https://doi.org/10.1145/3572549.3572624>
- Hutt, S., & D’Mello, S. K. (2022). Evaluating Calibration-free Webcam-based Eye Tracking for Gaze-based User Modeling. *ACM International Conference Proceeding Series*, 224–235. <https://doi.org/10.1145/3536221.3556580>
- Hutt, S., Wong, A., Papoutsaki, A., Baker, R. S., Gold, J. I., & Mills, C. (2023). Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-022-02040-x>
- Hutt, S., Wong, A., Papoutsaki, A., Baker, R. S., Gold, J. I., & Mills, C. (2024). Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*, 56(1), 1–17. <https://doi.org/10.3758/s13428-022-02040-x>
- Ikermane, M., & El Mouatasim, A. (2023). Dyslexia deep clustering using webcam-based eye tracking. *IAES International Journal of Artificial Intelligence*, 12(4), 1892–1900. <https://doi.org/10.11591/ijai.v12.i4.pp1892-1900>

- Jung, S. G., Salminen, J., & Jansen, B. (2021). Implementing Eye-Tracking for Persona Analytics. *Eye Tracking Research and Applications Symposium (ETRA), Part F169260*. <https://doi.org/10.1145/3450341.3458765>
- Khan, A. R., Khosravi, S., Hussain, S., Ghannam, R., Zoha, A., & Imran, M. A. (2022). EXECUTE: Exploring Eye Tracking to Support E-learning. *IEEE Global Engineering Education Conference, EDUCON, 2022-March*, 670–676. <https://doi.org/10.1109/EDUCON52537.2022.9766506>
- Kreitzer, R. J., & Sweet-Cushman, J. (2022). Evaluating Student Evaluations of Teaching: a Review of Measurement and Equity Bias in SETs and Recommendations for Ethical Reform. *Journal of Academic Ethics, 20*(1), 73–84. <https://doi.org/10.1007/s10805-021-09400-w>
- M. -F. Vaida and P. G. Pop, "Grouping strategy using Enneagram typologies," 2014 IEEE International Conference on Automation, Quality and Testing, Robotics, Cluj-Napoca, Romania, 2014, pp. 1-6, doi: 10.1109/AQTR.2014.6857844.
- Mills, C., Raffaelli, Q., Irving, Z. C., Stan, D., & Christoff, K. (2018). Is an off-task mind a freely-moving mind? Examining the relationship between different dimensions of thought. *Consciousness and Cognition, 58*, 20–33. <https://doi.org/10.1016/j.concog.2017.10.003>
- Monteiro, S., Bhate, R., Sharma, L., & Shaikh, P. (2022). Proct-Xam - AI Based Proctoring. *2022 2nd Asian Conference on Innovation in Technology, ASIANCON 2022*. <https://doi.org/10.1109/ASIANCON55314.2022.9908817>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. In *The BMJ* (Vol. 372). BMJ Publishing Group. <https://doi.org/10.1136/bmj.n71>
- Papoutsaki, A., Daskalova, N., Sangkloy, P., Huang, J., Laskey, J., & Hays, J. (n.d.). *WebGazer: Scalable Webcam Eye Tracking Using User Interactions*. <https://webgazer.cs.brown.edu>
- Papoutsaki, A., Laskey, J., & Huang, J. (2017). SearchGazer: Webcam eye tracking for remote studies of web search. *CHIIR 2017 - Proceedings of the 2017 Conference Human Information Interaction and Retrieval*, 17–26. <https://doi.org/10.1145/3020165.3020170>
- Prystauka, Y., Altmann, G. T. M., & Rothman, J. (2023). Online eye tracking and real-time sentence processing: On opportunities and efficacy for capturing psycholinguistic effects of different magnitudes and diversity. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-023-02176-4>

- Robal, T. (2019). Spontaneous webcam instance for user attention tracking. In: 2019 Portland International Conference on Management of Engineering and Technology (PICMET), pp. 1–8. IEEE
- Robal, T., Zhao, Y., Lofi, C., & Hauff, C. (2018). Webcam-based attention tracking in online learning: A feasibility study. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 189–197. <https://doi.org/10.1145/3172944.3172987>
- Rogers, T., & Al Madi, N. (2023, May 30). On the Pursuit of Developer Happiness: Webcam-Based Eye Tracking and Affect Recognition in the IDE. *Eye Tracking Research and Applications Symposium (ETRA)*. <https://doi.org/10.1145/3588015.3590129>
- Sauter, M., Hirzle, T., Wagner, T., Hummel, S., Rukzio, E., & Huckauf, A. (2022, June 8). Can Eye Movement Synchronicity Predict Test Performance with Unreliably-Sampled Data in an Online Learning Context? *Eye Tracking Research and Applications Symposium (ETRA)*. <https://doi.org/10.1145/3517031.3529239>
- Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, 50(2), 451–465. <https://doi.org/10.3758/s13428-017-0913-7>
- Sodoké, K., Nkambou, R., Tanoubi, I., & Dufresne, A. (2021). Toward a Webcam Based ITS to Enhance Novice Clinician Visual Situational Awareness. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12677 LNCS, 239–243. https://doi.org/10.1007/978-3-030-80421-3_26
- Toreini, P., Langner, M., Vogel, T., & Maedche, A. (2021). EyeTC: Attentive Terms and Conditions of Internet-Based Services with Webcam-Based Eye Tracking. *Lecture Notes in Information Systems and Organisation*, 52 LNISO, 112–119. https://doi.org/10.1007/978-3-030-88900-5_13
- Tzeng, J. W., Hsueh, C. Y., Lee, C. A., & Shih, W. Y. (2023). Identifying the Correlation Between Online Exam Answer Trajectory and Test Behavior Based on Artificial Intelligence and Eye Movement Detection Technology. *2023 International Conference on Consumer Electronics - Taiwan, ICCE-Taiwan 2023 - Proceedings*, 503–504. <https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226745>
- Van der Cruyssen, I., Ben-Shakhar, G., Pertzov, Y., Guy, N., Cabooter, Q., Gunschera, L. J., & Verschuere, B. (2023). The validation of online webcam-based eye-tracking: The replication of the cascade effect, the novelty preference, and the visual world paradigm. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-023-02221-2>
- Wang, Y., Lu, S., & Harter, D. (2021). Multi-Sensor Eye-Tracking Systems and Tools for Capturing Student Attention and Understanding Engagement in Learning: A Review. In

- IEEE Sensors Journal* (Vol. 21, Issue 20, pp. 22402–22413). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/JSEN.2021.3105706>
- Wisiecka, K., Krejtz, K., Krejtz, I., Sromek, D., Cellary, A., Lewandowska, B., & Duchowski, A. (2022, June 8). Comparison of Webcam and Remote Eye Tracking. *Eye Tracking Research and Applications Symposium (ETRA)*. <https://doi.org/10.1145/3517031.3529615>
- Wong, A. Y., Bryck, R. L., Baker, R. S., Hutt, S., & Mills, C. (2023). Using a Webcam Based Eye-tracker to Understand Students' Thought Patterns and Reading Behaviors in Neurodivergent Classrooms. *ACM International Conference Proceeding Series*, 453–463. <https://doi.org/10.1145/3576050.3576115>
- Yang, S., Poonawala, A. A., Jiang, T. S. A., & Schneider, B. (2023). Can Synchronous Code Editing and Awareness Tools Support Remote Tutoring? Effects on Learning and Teaching. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2). <https://doi.org/10.1145/3610177>
- Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, 16(6), 1485–1505. <https://doi.org/10.1017/s1930297500008512>

APPENDIX A – Systematic review: Overview

Advanced algorithms/hardware enhancements

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Dharbadeshwar et al., 2024)	Individuals aged between 18-50 years were recruited to participate in the study. This setup indicates that the sampling method did not employ random sampling, having the criteria to include volunteers who met the age criteria and were available for the study. Selection of participants who are in the age between 18-25 years and perform both online and offline shopping regularly.	Approximately 11,000 images are collected from participants aged 18-50 years for training the convolutional auto-encoders (CAEs). For both the User Experience Study and the Shopper Insights Study, about 100 participants aged 18-25 years are considered.	Participants were asked to gaze at various predefined regions on a screen, with images captured under varying lighting conditions. The stimuli included a calibration grid with predefined points for gaze mapping. Participants explored a web page of an e-commerce site after eye tracking calibration to capture the visual elements that attracted attention. Participants viewed a shelf containing products of various brands, and the system identified the products that gained the user's attention.	The study describes a real-time eye tracking system developed by Entropik Technologies Pvt. Ltd. The equipment used for the experimental setup and case studies includes: Standard Webcam, Desktop/Laptop Screens, Pretrained Convolutional Auto-Encoders, Ridge Regression Model, C++ SDK.	A real-time, non-intrusive eye tracking framework using representation learning and regression for gaze estimation. Utilization of convolutional auto-encoders (CAEs) to generate eye feature vectors and ridge regression for mapping these features to gaze points. Integration of compensation mechanisms to account for head pose variations and multimodal feature vectors including head pose and eye landmark corners. Calibration of the eye tracking system using predefined points on a screen, validation through comparison with mouse clicks and smooth-pursuit tasks, and benchmarking against a baseline model. Application of the eye tracking system in a User Experience Study to analyze web page engagement, and a Shopper Insights Study to determine product attention on a shelf display.	The authors developed a real-time eye tracking system using representation learning and regression, which requires only a limited number of calibration images. The system is non-intrusive, works with common webcams, and integrates compensation mechanisms for head pose variations. Regarding User Experience Study, the eye tracking system was used to analyze user engagement with a web page. The results showed that certain visual elements, like the H&M logo and category options, quickly captured the participants' attention. Regarding Shopper Insights Study, it was focused on identifying products on a shelf that attracted the most attention. The eye tracking system revealed that products placed at the center row of the shelf were the most viewed, indicating their higher consumer interest. The proposed eye tracking framework was validated through various tests and compared with a baseline model, showing mean pixel errors within an acceptable range, demonstrating its effectiveness for real-world applications. The system is adaptable to different platforms like mobiles and laptops.	The system must account for individual differences in eye appearance, such as eyelid shape and foveal offset, which can affect gaze estimation accuracy. Although the system requires fewer calibration images than other models, it still needs a certain number of images per calibration point to function correctly. The system includes mechanisms to compensate for head movements, but these may not fully account for all inaccuracies caused by changes in head pose during eye tracking. The system's performance can be influenced by varying lighting conditions, which can affect the accuracy of gaze estimation.	The authors suggest that future work should focus on reducing the number of calibration points used with advanced interpolation techniques or CNN-based architectures.
(Gershov et al., 2023)	The study does not provide specific details about the sampling method used. However, it mentions the collection of video recordings in the operating room and the use of these recordings to compare different visual behavioral patterns.	33	Real-life events and tasks that anesthesiologists engage with, such as monitoring patient vitals, managing anesthesia, and responding to critical incidents.	Webcam, StreamPix Software, BORIS Software, YOLOV7, Eye-Context Interaction Inference Network (ECIN), Spatiotemporal Gaze Architecture.	The study collected video recordings in the Operating Room, capturing anesthesiologists' behavior via monitor-mounted webcams, to understand their visual attention distribution during surgeries. The researchers employed advanced deep learning techniques to process the video data, aiming to minimize disturbance to the anesthesiologists' natural workflow. The study compared different visual behavioral patterns, distinguishing between baseline visual attention during uneventful periods and active phases or critical incidents. An eye-tracking method was used, replacing wearable eye-tracking devices with a webcam-based system for continuous data collection without disrupting the operating room environment.	The proposed framework was evaluated using a labeled simulation dataset, achieving an accuracy of 89.22% and an F1-score of 0.87, indicating high effectiveness in detecting onfocus behavior. Detailed analysis of anesthesia procedures showed that the framework could assess gaze frequency, length, and cumulative time, particularly during the induction phase of anesthesia, which is critical for patient safety.	The dataset used for evaluating different models' performance was relatively small. This limitation may restrict the generalization capability of machine learning models and could potentially lead to overfitting. Despite efforts to manually label events and monitor interactions, the study notes discrepancies between manually labeled events and those detected by the proposed framework. Specifically, inconsistencies were observed in the frequency, duration, and total time results for ventilator monitor interactions, partly attributed to camera angle variations. While overlapping manually labeled events with monitor interactions detected by the framework provided procedural context, the study recognizes that certain nuances may not have been captured comprehensively. For instance, the study focused on the induction phase of anesthesia procedures, but other phases and activities within the operating room environment were not extensively analyzed.	Further evaluate and enhance model performance in real-life clinical environments. Expand studies to real-life clinical settings to validate the framework's effectiveness. Conduct longitudinal studies to assess the long-term impact on healthcare quality and patient safety. Explore integration into clinical training and assessment programs. Investigate usability and workflow integration to minimize disruption in clinical settings.
(Deivanayagi et al., 2019)	The study does not provide specific details about the sampling method used.	The study does not provide specific information regarding the number of participants.	The study does not provide specific information regarding the stimuli used.	External Webcam, Raspberry Pi, Python, Pycharm, Hear cascade classifier, Support Vector Machines, Multi-label image segmentation method.	Detecting objects in the facial area using a sliding window technique, focusing on the middle area of the face image. Haar cascade classifier is used for locating the eye's area, followed by determination of the region of interest (ROI) of the eye. Images are pre-processed by converting color images into grayscale to simplify processing. Segmentation of the pupil of the eye is then performed using the Random Walker algorithm, allowing for accurate identification of the eye's center. Eye ball detection involves tracking the center of the iris using Harr's cascade in a video sequence. Machine Learning (ML) based approach is employed to classify eye states as open or closed, considering factors like image noise and motion blur. Feature extraction for detecting eye gaze. Inner eye corners serve as reference points for gaze tracking, and vectors connecting eye corners and iris center are used to calculate gaze position accurately. Haar features and a sliding window approach are employed for efficient detection. Classification of eye states and gaze tracking are performed using Support Vector Machines (SVMs), which define a linear boundary between classes for accurate classification.	Achieved precise real-time tracking of pupil movements during live eye gaze, ensuring accurate monitoring of eye behavior. Successfully detected facial landmarks and emotions using automated learning-free techniques, enhancing understanding of human behavior. Effectively identified eye ball movements, facilitating analysis of human behavior and mood recognition. Employed Support Vector Machines (SVMs) for accurate classification of eye states (open or closed), improving the system's reliability.	The algorithm for eye ball detection may return false positives, especially when the eyes are closed or under conditions like low contrast, image noise, and motion blur. While machine learning-based approaches improve eye state classification, they may be less accurate under adverse conditions, potentially reducing overall system reliability. Despite advancements in segmentation methods like the Random Walker algorithm, challenges remain in accurately segmenting the pupil of the eye, impacting the overall performance of the system.	The study does not provide specific details regarding future works.
(Robal et al., 2018)	Graduate students and staff members from TU Delft recruited via email lists. Participants did not receive compensation and the sample was diverse by having participants who wore glasses and contact lenses.	20	The study employed a benchmark suite of fifty tasks to simulate common behaviors of online learners like covering the camera for 2 seconds, looking straight up to the ceiling for 8 seconds and checking the phone for 10 seconds.	Tobii X2-30 Compact, Standard webcam, WebGazer, CLM (Kimmerdijs) - a face fitting library within WebGazer, tracking.js.	The study aimed to detect learners' loss of focus during online learning using webcam-based attention tracking. The study explored a simple alternative - detecting the presence or absence of a face - to identify loss of focus. Two software-based frameworks were evaluated: WebGazer.js (CLM) and tracking.js (TIS). Fifty realistic tasks were designed to simulate common learner behaviors during video lectures.	TIS outperformed CLM significantly in detecting both face hits and face misses. Tobii, a professional eye tracker, achieved the highest accuracy but is not practical for large-scale deployment. TIS and CLM introduced a delay of around 1-2 seconds compared to Tobii. Despite the delay, these software-based solutions are still viable for real-time attention tracking. TIS accuracy was affected by background noise; solid light backgrounds performed better than those with posterized photos. Using face presence/absence as a proxy, TIS provides a feasible solution for real-time attention tracking in Massive Open Online Course (MOOC) environments, although with slight delays compared to Tobii. CLM, on the other hand, had poor performance due to aggressive face fitting.	The approach has a delay of around 1-2 seconds between the instructed start time of a task and the first detection of face absence. While this delay is manageable, it is not instantaneous like the high-end hardware solution (Tobii). The accuracy of the software-based solution (TIS) is affected by noisy backgrounds. In particular, when the background contains posters or photos, the detection performance degrades. The CLM framework (used by WebGazer.js) tends to detect faces even when none are present due to its aggressive face-fitting algorithm. This results in false positives, impacting the overall accuracy. The software-based solutions rely on consumer-grade webcams and browser processing, which can vary significantly across different devices. Achieving consistent high detection quality across all tasks remains challenging.	Implement an attention tracker suitable for large-scale MOOC deployment based on the TIS framework. Investigating whether MOOC learners would be willing to accept and use such an attention detection tool, exploring the reasons why they would like or refuse to use such technology. Studying whether learners' acceptance of such tools indeed positively impacts their learning outcomes.

Eye-tracking in user interaction and UX studies

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Wisiecki et al., 2022)	Random assignment to one of three experimental conditions: webcam, remote, or integrated eye tracking.	83	Point Detection Task: Participants were asked to look at and click on points displayed on the screen, which appeared three times at one of five spots. Face-in-the-Crowd Task: Participants had to find and click on a face expressing a different emotion (happy or sad) from other neutral faces in a crowd matrix of varying sizes (2x2, 3x3, and 4x4). This task was requested to test attentional biases towards emotional stimuli.	RealEye, LOGITECH HD Pro C920 webcam, PsychEye 3, G9 eye tracker.	The study was structured as a mixed design experiment with three between-subjects conditions: webcam-based, remote, and integrated eye tracking. A total of 83 students participated in the study, divided into three groups corresponding to the three experimental conditions. Participants completed two tasks: a point detection task and an emotional visual search task - to evaluate the accuracy, precision, and validity of the webcam-based eye tracking against the remote eye tracker. The study analyzed location-based metrics like fixations and process-based metrics like ambient-focus attention dynamics to assess the performance of the different eye tracking methods.	The study found that webcam-based eye tracking had a higher measurement error compared to remote eye tracking and the integrated method. Despite this, the webcam method was still considered reliable. The time to first fixation on a target face was shorter for happy faces than sad faces across all recording conditions, supporting the "happiness-surgency effect". Larger crowds induced more focal attention than ambient attention during visual search tasks. This suggests that participants engaged in more detailed processing of visual stimuli in denser crowds.	The experiments were conducted in a controlled laboratory environment, which may not accurately represent real-world conditions where webcam eye tracking would be used. The accuracy of the webcam might be enhanced due to controlled factors like head position, which is not typically managed outside of a lab setting. While the study shows improvement in precision and accuracy due to advancements in hardware and webcam platforms, these factors are constantly evolving, and the study's results may not apply to future developments.	Replication of findings outside of the lab in real-world conditions to validate the results. Further exploration of real-time alerting systems for focal processing based on webcam eye tracking data. Investigation of applications in various fields, such as computer-supported learning or assistive technology, using the insights gained from webcam eye tracking.
(Ashay et al., 2023)	The study used a convenience sampling approach, selecting participants whose age ranged from 20 to 35. This age group was chosen due to their familiarity with technology and online platforms. Each participant underwent an initial calibration process to adapt the system to their screen size and lighting conditions.	25	Participants were asked to identify products on the website that were marked as sponsored. Participants were tasked with finding a specific product, such as the Reynolds OAS blue pen, within the product list page.	CASE-U 1000p webcam.	The study aims to understand user behavior related to User Interfaces (UI) using eye-tracking data. A cost-effective system was developed using a high-resolution webcam to record raw eye gaze data. Participants performed initial calibration to adapt the system to their screen size and lighting conditions. Participants interacted with an e-commerce website while their eye movements were recorded. Fixations (stable gaze points) were calculated based on Euclidean distance, time differences, and velocity thresholds. Fixations were used to generate visual heatmaps and scan paths, revealing user behavior on the test subject. The system identified UI components that users focused on the most, helping optimize UI design. The system can then inform ad placement, content positioning, and improve user experience.	The first product on the website consistently attracted the most attention. This finding suggests that placing advertisements near the first product can maximize user engagement. About 87% to 100% of fixations occurred in the center region of the screen. This information can guide UI designers in placing important components where users focus the most. By analyzing eye movements, designers can optimize UI components for user attention. For example, identifying areas where users look the most can inform the placement of newsletter sign-up forms or other critical elements.	Webcam data collection might be affected by environmental factors such as lighting conditions. The proposed system aims to be cost-effective by using webcams, but it sacrifices some accuracy. Fixations are essential for analysis, but false fixations can occur due to webcam focus issues or other factors. Preprocessing is applied to improve reliability, but false fixations may still impact results. Participants may exhibit different behavior based on their familiarity with technology. The study focused on participants aged 20-35, which might not represent all user demographics. While the system identifies regions of user attention, it does not directly suggest optimal placements for UI components. Designers must interpret the results and make informed decisions.	Enhance the initial calibration process to adapt to various screen sizes and lighting conditions more effectively. This will further improve tracking accuracy. Investigate other UI components beyond advertisements. For instance, focus on newsletter sign-up forms or blog snippets. By analyzing user gaze patterns, designers can strategically place these components for maximum attention. Collaborate with web designers to incorporate eye-tracking insights into webcam focus issues or other factors. Preprocessing is applied to improve reliability, but false fixations may still impact results. Conduct comparative studies between webcam-based and infrared (IR)-based eye trackers. Assess the trade-offs in accuracy, cost, and accessibility to determine the most suitable approach for specific contexts. Address eye detection challenges in low-light environments when using webcams. Explore techniques to improve accuracy even under suboptimal lighting conditions. Develop adaptive UIs that respond to real-time eye-tracking data. For example, adjust content placement based on user attention or dynamically change component positions to optimize engagement.
(Papoutsaki et al., 2017)	Recruiting crowd-workers through Amazon Mechanical Turk that were qualified based on having webcams and browsers that supported the getUserMedia() API. Restricted Focal View Result Examination Study: Initial number of participants was 47 but 11 were excluded. Final set was 36 participants.	Initial number of participants was 49 but 13 were excluded. Final set was 36 participants.	Participants performed search tasks using a search engine (e.g., Bing or Google). The SERPs contained a mix of organic search results, ads, and other elements. Participants examined these SERPs to find relevant information based on their search queries. As they view and interact with the SERPs, participants' gaze behavior is tracked by SearchGazer that predicts where participants are looking on the screen.	Standard Webcam, SearchGazer.	Result Examination Behavior Study: Investigate the effects of snippet length changes on web search behavior. 36 crowd-workers conducted search tasks on Bing SERPs. Replicated a study by Cutrell et al. with modifications. Using Bing instead of MSN Search. Participants performed 12 search tasks on unaltered SERPs. SearchGazer was added to predict gaze locations. Participants completed search tasks, providing answers or declaring inability to find information. Calibration steps were included. Ad Examination Behavior Study: Investigate user behavior regarding ads on search engine results pages (SERPs). 33 crowd-workers performed search tasks on SERPs. Replicated Bucher et al.'s study with modifications. Used 12 queries with varying ad quality. SearchGazer was used to predict gaze locations. Participants completed search tasks, examining SERPs with different ad qualities. Tasks were randomized, and ad quality blocks were assigned to participants. Restricted Focal View Result Examination Study: To assess user behavior using a tool that reveals one search result at a time on a SERP. 36 crowd-workers performed search tasks. Replicated Lagun and Agichtein's study using Google SERPs. Added SearchGazer for gaze prediction. Participants completed search tasks, clicking within the first page of results. Similar to the Result Examination Behavior Study.	Result Examination Behavior Study: Users tend to examine search results from top to bottom, with lower-ranked results attracting attention last. Fixation duration decreases with lower-ranked results, but attention is still spread across all results. Users typically view a few results above and below before clicking, with a slight bias towards higher-ranked results. Ad Examination Behavior Study: Most visual attention falls on top organic results, with some attention to top ads. Participants spend more time on SERPs for informational tasks than navigational tasks. Visual attention differs based on ad quality, with good ads receiving more attention, but results may vary due to study size. Restricted Focal View Result Examination Study: Users tend to spend more time on SERPs when viewing one result at a time, examining even lower-ranked results. Viewing and clickthrough rates decay linearly with lower-ranked results, but rates may vary between studies due to user experience differences. Restricted viewing leads to more careful examination of results and potentially higher clickthrough rates for lower-ranked results.	Result Examination Behavior Study: The authors acknowledge that the nature of remote crowd-workers using Amazon Mechanical Turk may lead to faster task completion and less diligent task performance compared to in-lab participants. Additionally, crowd workers may approach tasks differently due to English being a secondary language for some, potentially affecting the study outcomes. Ad Examination Behavior Study: The authors note that webcam eye tracking predictions may lack precision, leading to scattered inferred gaze positions along the x-axis. They also highlight that the reduced number of tasks (12 instead of 32) in their study might have diminished the effects that ads typically have on user behavior. Restricted Focal View Result Examination Study: The authors mention that the differences in user experience between in online study using a restricted focal view interface and traditional eye tracking studies may influence user behavior. They suggest that restricted focus viewing may lead to more thorough examination of results, potentially affecting clickthrough rates. Additionally, they observe that some informational tasks were vague, leading crowd-workers to click on the first few results instead of examining lower-ranked ones.	Developing an algorithm tailored for SearchGazer to identify fixations, similar to existing eye trackers, to provide a clearer picture of gaze behavior and enable replication of more fine-detailed studies. Conducting in-lab studies to extensively compare SearchGazer with physical eye trackers to better understand its behavior, strengths, and limitations. Refine SearchGazer's computer vision components by interacting the software as open source. They emphasize the modularity of SearchGazer, allowing different eye detectors, regression models, or relaxation-based learning algorithms to be swapped out for optimization. The authors discuss privacy risks associated with webcam use in online applications but suggest that users may accept these risks if clear benefits outweigh them. They envision scenarios where users could be compensated or offered incentives for sharing gaze behavior, such as an advertising-free search interface or additional gaze-based search features.
(Valda, M.-F., & Pop, P. G., 2014)	The solution considers a class of students.		Participants engage in practical activities related to software development and programming. The e-learning environment integrates eye tracking to observe students' reactions while following theoretical aspects. The content is presented through slides, and the eye tracking system detects areas of interest (e.g., text regions or images) where students focus their attention. Supplementary explanations are provided based on these observations.	Standard Webcam, Eye Tracking System, Content Management System (CMS), Machine Perception Toolbox (MPT) Framework.	Utilizing the RHETI test to determine students' Enneatypes, identifying their personality traits and tendencies. Implementing an e-learning platform with an eye tracking system to observe students' behavior during individual practical activities, analyzing their reactions to course material presentations. Processing eye tracking data to identify students' predominant patterns, such as visual, auditory, sensitive, and kinesthetic structures, and correlating these with Enneatypes. Formulating a grouping strategy based on typology compatibilities derived from Enneagram interpretations, aiming to maximize positive interactions between group members. Iteratively refining the grouping strategy by considering both Enneagram traits and NLP profiles, ensuring diverse and complementary representations within each group. Assessing the effectiveness of the grouping strategy in maximizing efficiency for team-based projects, with a focus on achieving balanced group compositions and promoting collaborative learning.	Groups formed based on Enneatypes and NLP profiles showed enhanced communication compared to classical teams. This communication included effective brainstorming, leading to increased productivity. Groups formed using the proposed strategy demonstrated higher levels of creativity and originality in their developed products. This creativity was fostered by the collaborative environment and diverse perspectives within the groups. Members of groups that utilized brainstorming methods showed significant improvement in combining knowledge with creativity, indicating a deeper understanding of how to integrate learning into practical applications. The grouping strategy based on Enneatypes and NLP profiles contributed to the self-development of students, as they learned to work collaboratively and leverage their individual strengths within a team setting. The use of the Enneagram methodology and NLP profiles in group formation led to positive group dynamics, resulting in more reliable and successful project outcomes. The effectiveness of the grouping strategy was further validated during a summer internship at a software company, where similar positive results were observed.	The correlations between Enneagram types and Jungian functions are acknowledged to be imperfect and not straightforward. The Enneagram test may introduce conflicts based on stereotypes and selective deformation, potentially affecting the accuracy of typology determination. The eye tracking system's performance is described as satisfactory but not optimal, indicating room for improvement. The grouping strategy based on Enneatypes and NLP profiles may not fully maximize the potential of all groups, as it relies on a specific interpretation of personality types.	Incorporating a new grouping algorithm called "Footprints", described as a semi-opportunistic grouping mechanism, to improve the creation of working groups.

Integration with educational platforms

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Bogdan et al., 2023)	<p>The sampling method described in the study involves recruiting participants from a local university for two separate studies on emotion-attention interaction tasks.</p> <p>The participants were selected to complete either an in-person version of the task, which involved infrared eye-tracking data collection, or an online version using webcam-based eye tracking.</p> <p>The study mentions specific exclusion criteria, such as responding to fewer than 75% of the emotional rating trials or having unusable eye-tracking data.</p> <p>Power analyses were conducted to determine the minimum number of participants needed for each study, and additional participants were recruited based on memory-related hypotheses beyond the current focus.</p> <p>The study also notes that the in-person data were collected pre-pandemic, while the online data were collected during the pandemic.</p>	<p>Study 1: The study initially recruited 139 participants. After exclusions, 34 participants completed the in-person version, and 87 completed the online version.</p> <p>Study 2: This study recruited 264 participants. After exclusions, 42 participants completed the in-person version, and 180 completed the online version.</p>	<p>The stimuli used in both Study 1 and Study 2 were composite images containing a negative or neutral foreground overlaid upon a neutral background. The difference from Study 1 to Study 2 was the inclusion of instructions on how to attend to the images.</p> <p>These images were utilized to examine the intersection between emotion and attention in participants.</p> <p>The foreground components were sourced from various image databases such as the International Affective Picture System (IAPS), the Geneva Affective Picture Database (GAPED), the Military Affective Picture System (MAPS), the Nencki Affective Picture System (NAPS), and the Emotional Picture Set (EmoPicS), as well as freely available online sources for the background components.</p> <p>The images were designed to be approximately 50% foreground and 50% background, with the emotional images being negatively valenced and arousing, while the neutral images were appropriately neutral and non-arousing.</p>	<p>The in-person data collection used the EyeLink 1000 system (Infrared Eye-Tracking System). The online data collection used standard webcam and WebGazer.</p>	<p>Study 1: Recruited 139 participants from a local university.</p> <p>Participants viewed a series of 90 composite images (60 negative and 30 neutral) and reported the intensity of their emotional reactions after viewing each image.</p> <p>Eye movements were recorded using an Infrared EyeLink 1000 system for in-person data collection and the WebGazer.js package for online data collection.</p> <p>The proportions of gaze points recorded in the foreground were submitted to paired t tests and multilevel linear regressions.</p> <p>Study 2: Recruited 264 participants from the local university.</p> <p>Similar to Study 1, but before each image, participants were cued to focus on either its foreground or background areas. After each image, participants reported their emotional rating.</p> <p>Eye-tracking procedures are the same as Study 1.</p> <p>Regarding analytic procedures, they are similar to Study 1 but also included additional analyses to investigate the effect of emotion on each focus condition and the effect of gaze on emotional ratings.</p>	<p>Study 1: Online vs In-Person Data: The study found that online eye-tracking data replicated seven of eight in-person results, with effect sizes being 52% [42%, 62%] the size of those seen in-person.</p> <p>Spatial Error Impact: It was observed that online eye tracking is biased towards recording more gaze points near the center of participants' screens, which can interfere with comparisons if not addressed.</p> <p>Study 2: Attentional Cue Effect: The study confirmed that participants could effectively control their attention based on cues, focusing on either the foreground or background as instructed.</p> <p>Emotion-Attention Interaction: Despite the attentional cues, subtle emotional attention-capture or attention-aversion effects were still present, indicating that emotional content can influence gaze patterns even when participants are instructed to focus elsewhere.</p>	<p>The authors noted that webcam-based eye tracking has increased spatial error compared to infrared systems, which can impact the precision of psychological research.</p> <p>Online eye tracking tends to record more gaze points near the center of the screen, which can interfere with analyses if not accounted for.</p> <p>Transitioning to online eye tracking led to a decrease in effect sizes, with online data replicating in-person results at about 52% of the size.</p> <p>The simulation revealed that spatial noise can cause a flip in attentional-aversion effects, indicating that online eye tracking may not accurately capture subtle gaze-related behaviors.</p>	<p>Investigate further into the feasibility and utility of webcam-based eye tracking for psychological research.</p> <p>Examine the impact of spatial error on aggregate-level quantities and how it affects the study of psychological phenomena.</p> <p>Assess the effect sizes when transitioning from in-person to online data collection and determine the necessary sample size for well-powered studies.</p> <p>Explore the "centering bias" in online eye tracking and its potential interference with psychological effect analysis.</p>
(Tzeng et al., 2023)	<p>The study does not provide specific details about the sampling method used.</p>	<p>The study does not provide specific details regarding the amount of participants.</p>	<p>Educational materials and academic tasks.</p>	<p>Wearable eye tracker. Video camera. Long Short Term Memory (LSTM) model.</p>	<p>The system incorporates eye movement detection technology to analyze students' answering behaviors, focus of sight, and the order of answering questions.</p> <p>A deep learning model based on Long Short Term Memory (LSTM) is designed to interpret the eye movement data and identify abnormal response behaviors.</p> <p>The experiment collects data on students' answering behavior and eye movement during tests. It uses wearable eye trackers and video cameras to gather large-scale data for analysis.</p>	<p>The system analyzes students' answering behaviors, focus of sight, and the order of answering questions to address issues arising from remote assessments.</p> <p>Four common eye movement indicators were used to understand the differences in attention and frequency among students of different achievement levels when answering questions.</p> <p>The study suggest that the AIPS can potentially improve the fairness of online assessments by identifying candidates with abnormal answering patterns through AI assistance</p>	<p>The study does not provide specific details regarding limitations.</p>	<p>The study does not provide specific details regarding future works.</p>

Open-source software (WebGazer)

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Wong et al., 2023)	Participants from three classes of an Education course at Lindemur College. The participants were neurodivergent students, defined as having a diagnosis of ADHD, autism, and/or learning disability. The selection strategy focused on the functional challenges of the students rather than specific diagnoses, aiming to capture a broad range of neurodivergence. The study did not employ random sampling.	43 (28 Male, 11 Female, 2 Non-binary, and 2 Prefer not to say)	An excerpt from "The Power of Habit" by Charles Duhigg, which examines habit formation and its correlation to academic success. Students read the text self-paced, one paragraph at a time, and answered periodic thought probes about their ongoing thought patterns. The study aimed to unobtrusively monitor students' gaze while learning from an online reading platform, focusing on their familiarity with the material and thought patterns such as mind wandering or "sticky" thoughts.	Standard webcam. WebGazer.	Gaze locations were collected using WebGazer, a JavaScript library that infers gaze locations using the user's webcam. It employs facial and eye detection algorithms to determine pupil location and represent the eye as an image patch. A ridge regression model was used to map pupil locations and eye features to gaze locations on the screen. WebGazer calculates x-y-z coordinates of gaze locations at a temporal interval of 500ms. Gaze measures such as the number of gazes, area of interest (AOI) gazes, and offscreen gazes were calculated for each paragraph read by the participants. AOI gazes measured attention within the text area, while offscreen gazes indicated disengagement. Proportions were calculated by dividing the relevant measure by the total gazes for that paragraph. These gaze measures were then used to analyze the relationships between gaze behavior and on-the-thoughts. The study validated the use of a free, open-source, webcam-based eye tracker to monitor reading behaviors and thought patterns in neurodivergent students within real-world educational settings. Eye-tracking measures were sensitive to moments when students experienced difficulty disengaging from their thoughts, indicating potential for real-time interventions. The eye tracker was able to detect differences in gaze behavior based on whether students had previously read the text, with those familiar with the text showing different reading patterns. The research highlighted the importance of designing adaptive educational technologies that cater to the unique needs of neurodivergent students, promoting a more inclusive educational environment.	Authors note that the webcam-based eye tracker, WebGazer, may be less accurate than traditional research-grade equipment, especially in ecologically valid settings. The study acknowledges that factors such as lighting conditions and participant variables (like eye color or wearing glasses) could affect the accuracy of WebGazer's gaze predictions. The absence of a comparison dataset with a traditional eye tracker means they couldn't statistically assess the validity of WebGazer's gaze location predictions. The authors mention the need for further research to assess the generalizability of their findings, particularly in comparison to neurotypical individuals. They also suggest exploring the impact of various moderating effects on the accuracy of WebGazer.	Conducting further research to better assess the accuracy of webcam-based eye trackers and compare them with traditional "gold standard" eye trackers. Understanding the limitations of WebGazer in ecological settings, including the impact of various moderating factors such as lighting conditions, eye color, and the use of glasses or contact lenses. Investigating whether the variation in gaze tracker accuracy meaningfully impacts user modeling and the effectiveness of interventions. Designing "tailored" interventions that account for the potential inaccuracies in gaze monitoring, ensuring no harmful effects if delivered at an incorrect time.	
(Ding et al., 2023)	Recruiting graduate students who were all second-language learners of English as participants. The selection criteria included the ability to use English in academic contexts and a mix of individuals who wore glasses and those who did not. The study did not employ random sampling.	12 (2 Male and 10 Female)	Use of TOEFL reading materials as stimuli for the participants. Participants read 12 articles per day for three days and spent four hours reading in total. These materials were selected for their academic relevance and to create a realistic setting for English as a Second Language (ESL) learners.	To propose Gazeheader, the following equipment was used: Standard webcam. WebGazer. Thinkpad X1 Carbon laptop (CPU: i7-12700H, 8 cores, 1.5 GHz, RAM: 32GB, Storage: 512GB).	The authors recruited 12 graduate students as participants, who were second language learners of English. They read 36 articles from TOEFL reading materials over three days, using a web PDF reader and WebGazer for gaze tracking. The gaze data collected was noisy and required de-noising using a moving average technique, re-sampling and linear interpolation. The gaze data was then aligned with the text data to match gaze and unknown words. A machine learning model was developed to predict unknown words during reading. It used LSTM layers to encode gaze and text positional data, and a transformer-based RoBERTa model to encode text information, knowledge-based features like term frequency, part of speech, and named entity recognition were also included to enhance performance. A user study was conducted to test the model's effectiveness and to explore design implications for ESL reading assistance tools. The study also involved a Vocabulary (Vocab) Test to assess participants' vocabulary levels.	Gazeheader achieved an accuracy of 98.09% in detecting unknown words for ESL learners using only a webcam. The method obtained an F1 score of 75.73%, indicating a balance between precision and recall in the model's predictions. The study identified areas for future design focus, such as proper nouns, multi-meaning words, and long, complex sentences. Positive user attitudes were observed towards eye-tracking reading assistance and vocabulary management features.	The dataset used for training the neural network is relatively small compared to the number of parameters, which might affect the model's performance. The inaccuracy of the webcam-based eye-tracking algorithm used (WebGazer) could limit the model's effectiveness, as it relies on the coordinates of clicks for calibration. Unknown words are sparse in the dataset, leading to an imbalance that could impact the F1 score of the model. While the method has potential, more diverse materials and user groups need to be included in future studies to enhance the model's generalizability.	More case studies to understand the reasons behind the model's F1 score and address the dataset's imbalance and size issues. Enhance the webcam-based eye-tracking algorithm to improve gaze contribution to performance. Test the method on different computer models and migrate the algorithm to various applications. More diverse materials to make the approach more generalizable. Include users at different stages of learning to broaden the study. Explore the application of their approach to other alphabetic languages and investigate its potential for non-alphabetic languages. Design a simple webcam-based tool that is efficient and usable for users at different levels, focusing on proper presentation of information needed during reading.
(Khan et al., 2022)	Small-scale experiment. The participants were selected to represent individuals engaging in a remote learning session. This setup was part of the development of a scalable, privacy-aware eye-tracking solution called DEECUTE, designed to support remote learning environments.	25	Short video lecture. This lecture served as the visual stimulus for raw gaze data collection during the experiment. The content of the lecture included various slides with pre-defined areas of interest (AOI) to facilitate data filtering and feature extraction related to gaze metrics. Participants were asked to watch this video and report distractions during their visual engagement with the content.	Standard webcam. WebGazer.	Using a webcam to capture and analyze raw gaze points, allowing natural head movement and scalability. Employing logistic regression, support vector machine, and polynomial regression to classify student attention with high accuracy. Developing a novel metric to quantify student engagement based on gaze metrics. Conducting a small-scale experiment with 25 participants to collect gaze data during a video lecture. Filtering raw data and extracting features like fixations, saccades, and eye blinks for attention classification.	The study presents an e-learning framework that uses eye-tracking data to profile students' learning behavior and attention during remote teaching sessions. A novel solution is proposed that uses a standard webcam for eye-tracking, allowing natural head movement and scalability. The study achieved over 93% accuracy in classifying student attention using machine learning models like logistic regression, support vector machine, and polynomial regression. The findings suggest potential applications in early detection and diagnosis of attention deficit hyperactivity disorder (ADHD) among students.	The proposed eye-tracking solution needs to be scalable for widespread deployment in remote learning environments. Achieving a high sampling rate with a standard webcam is challenging, which affects the accuracy of gaze data collection. The small dataset size may limit the generalization capability of machine learning models and could lead to overfitting.	Enhancing the study to identify anomalous gaze patterns for early detection and diagnosis of ADHD. Analyzing gaze movement to understand psychological states and assist individuals with ADHD. Developing tailored educational methods to support students with ADHD, including personalized assistance and accommodations. Conducting a large-scale remote learning study to evaluate visual perception and improve multimedia learning design.

Open-source software (WebGazer) (Cont.)

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Foroni et al., 2021)	Experimental design: two groups with manipulated attention feedback types, and a controlled laboratory experiment to evaluate the EyeTC system. The participants were randomly assigned to either the treatment or control group, and their interactions with the T&C were recorded and analyzed.	62	An experimental setup, named bogus experiment, where participants were asked to choose their favorite pictures among two options to track their eye movements to find the relationship between their choices and visual behavior.	Logitech Brio 4K Ultra HD webcam, WebGazer.	Creation of EyeTC Application, a neuro-adaptive system that tracks users' eye movements to determine if they are reading the T&C and provides feedback if they are not. A controlled laboratory experiment with two groups, where one group received EyeTC-based attention feedback and the other received general attention feedback. The effectiveness of EyeTC was measured by comparing the attention allocation and text comprehension between the two groups, using the duration of T&C reading and a declarative knowledge test.	The use of EyeTC led to users allocating more attention to reading terms and conditions (T&C), resulting in better text comprehension. When provided with EyeTC-based attention feedback, participants in the treatment group showed a significant increase in the time spent reading T&C during a wait phase compared to the control group. Participants using EyeTC performed better on a declarative knowledge test about the T&C content, indicating enhanced understanding. Despite the benefits, participants expressed privacy concerns regarding the sharing of eye movement data in real-world scenarios.	Webcam-based eye trackers are less accurate and precise compared to infrared eye trackers. Those kind of trackers are sensitive to movements, which required controlling for steady posture during the experiment. There is a possibility that EyeTC did not provide accurate highlighting for some participants. The evaluation focused on users' mouse click and survey results rather than eye movement data, to control for data noise from the webcam-based tracker.	Investigate the users' reactions to general highlighting of typical passages that people do not read, and the need for personalized adaptation of the system. Design and evaluate the EyeTC application using more accurate eye trackers to better understand how users process T&C and to support EyeTC in distinguishing between different reading behaviors. Check the effectiveness of EyeTC in real-world settings and conduct long-term studies to validate the results. Establish standards for integrating EyeTC either by T&C providers or in a way that users can install it to receive support. Extend the findings to create applications beyond attentive T&C, such as in e-learning or for companies to implement systems that enhance reading of important documents.
(Plytska et al., 2023)	A webcam-based eye tracking calibration procedure was implemented, and participants were given three attempts to pass it. Those who failed were excluded from the final analysis due to calibration failures and low sampling rates. The study targeted dominant native speakers of Russian and ensured that only personal computers were used for the experiment, excluding mobile phones.	253 participants but only 202 participants were included in the final analysis due to calibration failures and low sampling rates.	One set included 16 stimuli sets, each with a quadrant-based visual scene displaying four images and two minimal sentence pairs differing only in the verb. The sentences were designed to test the verb semantic constraint effect, where constraining verbs allowed prediction of the post-verbal object based on the verb alone, while non-constraining verbs did not. The other set consisted of 32 stimuli sets, each with a quadrant-based visual scene with four images and two minimal sentence pairs differing only in the color adjective. The scenes contained two pairs of different object types, with objects in the same pair differing only by color. The sentences had constraining verbs and were used to test the lexical interference effect, examining the influence of color adjectives on object identification.	GoRilla Experiment Builder platform, Standard webcam, WebGazer.	The research aims to replicate two psycholinguistic effects: the verb semantic constraint effect and the lexical interference effect, using the GoRilla Experiment Builder platform. Participants are recruited online, allowing for a more diverse sample that includes individuals who might not have access to laboratory settings. The paper discusses approaches to computing statistical power for mixed-effects models and the implications for determining the necessary sample size for future studies. The authors provide technical details and recommendations for implementing webcam-based eye-tracking studies, aiming to contribute to the method's development and reliability.	Demonstrates that webcam-based eye tracking can capture sentence processing effects in real-time with sufficient granularity, even for subtle linguistic phenomena. The study compared effects of different magnitudes, showing that not only robust effects but also smaller ones can be detected using online methods. This is significant for the field of psycholinguistics, as it expands the potential for diverse online research. The paper reports on two approaches to computing statistical power and discusses the differences in their outputs (80% power to detect the observed effect size can be achieved with 30 participants and 80% power to detect the observed effect size can be achieved with 20 participants). This contributes to the understanding of the necessary sample sizes and statistical considerations for future studies using similar online methodologies.	The discrepancy between simulation-based and resampling-based power analysis results, raising questions about the reliability of each method and the generalizability of their findings. The impact of sampling rate on the accuracy and resolution of online eye tracking, with lower rates potentially limiting the ability to capture fine-grained effects. Sound onset lag variability, which can affect the synchronization of eye movement data with audio playback, potentially leading to delay in experimental effects. Calibration challenges, with a relatively high number of participants failing calibration despite detailed instructions about lighting conditions. Timing delays in presentation sequences, which may be a limitation inherent to webcam-based eye tracking but could potentially be improved with technological advancements and experimental design adjustments.	Conducting more specific, dedicated research into the generalizability of different methods for power analysis. Explore how sampling rate impacts the ability to detect fine-grained language processing effects. Further investigation into the source of auditory lag issues, potentially exploring how different experimental packages, browsers, and operating systems affect timing variability. Providing more detailed instructions about lighting conditions and exploring the optimal frequency of recalibration. Suggest further research into technological advancements to reduce timing delays.
(Van der Cruyssen et al., 2023)	Recruiting participants through Prolific. The researchers aimed for a diverse sample, with eligibility restricted to English-speaking participants who had access to a computer with a functioning webcam, did not wear glasses during the experiment, and had not participated in the pilot study.	Initial Participants: 157 Final Participants: 136 Exclusion Reasons: Participants were excluded for showing no variation in estimated eye gaze across all trials, or having more than 50% of the measurement points falling outside any of the Areas of Interest (AOIs).	Study 1 - Cascade Effect: Pairs of faces were used to determine which one participants found most attractive. The faces were selected from the London Face Research database, ensuring minimal differences in average attractiveness ratings. The faces were matched for gender and ethnicity and had a maximum age difference of 4 years. Study 2 - Novelty Preference: Black and white, non-anxious oriented object images from the Snodgrass and Vandewort database were used. During the task, participants were first shown two identical images, followed by one familiar and one novel image to measure their attention to new stimuli. Study 3 - Visual World Paradigm: Common objects displayed on a screen, paired with sentences mentioning some of these objects. Participants' gaze was tracked to see if they looked more at the images of the words they heard in the utterances. The materials were constructed based on Huettig and Altmann (2005).	Standard webcam, WebGazer, EyeLink 1000 Pro.	Study 1 - Cascade Effect: To replicate the phenomenon where individuals prioritize the face they find most attractive before making a choice. Participants chose the most attractive face from a pair, with their gaze tracked to see if it favored the selected face before decision reporting. Online task with webcam-based eye-tracking using WebGazer's library. Study the likelihood of looking at the chosen face during the 100 ms before the decision was compared against chance level. Study 2 - Novelty Preference: Replicate the tendency of people to attend more to novel stimuli than familiar ones. Visual paired-comparison task with familiarization and test phases to track gaze toward novel stimuli. Online task with webcam-based eye-tracking. Viewing proportion toward novel stimuli was assessed against chance level. Study 3 - Visual World Paradigm: Compare online webcam-based eye-tracking with lab-based eye-tracking in replicating the effect where people look more at images of words they hear in sentences. Participants completed the task both online and in the lab, tracking gaze towards target words mentioned in sentences. Within-subject design with counterbalanced order for online and lab sessions. Understanding the proportion of fixations on target items compared between online and lab data.	Study 1 - Cascade Effect: The likelihood of participants looking at the face they eventually chose was 62% during the 100 ms before making a decision, which is significantly higher than the 50% chance level. The effect size was $d = 0.65$, indicating a medium to large effect, but smaller than the original study's effect. Study 2 - Novelty Preference: Participants looked at novel stimuli 57% of the time, which was significantly more than the 50% chance level. The effect size was $d = 0.46$, suggesting a moderate effect, again smaller than the original study's findings. Study 3 - Visual World Paradigm: In the online version, participants looked at the target items 52% of the time, significantly more than the control items. In the lab version, the fraction on target items was 71%, indicating a higher accuracy compared to the online method. The effect sizes were smaller in the online replication compared to the lab, with a significant difference between the two. Overall, these results suggest that while online webcam-based eye-tracking can replicate well-established effects, the precision and effect sizes tend to be smaller compared to traditional lab-based eye-tracking methods.	The effect sizes of the replicated studies were found to be 20-27% smaller than those obtained from lab-based eye-tracking. This shrinkage is attributed to the noisier data from online methods. Replicators typically show smaller effect sizes than original studies, often about 50% smaller. This phenomenon could also contribute to the smaller effect sizes observed in online studies. Changes in procedures between the original studies and the online replications could also affect the effect sizes.	Explore more complex experiments with longer durations and multiple Areas of Interest (AOI). Investigate calibration needs for extended studies or those with more than six AOIs. Utilize online eye-tracking to reach difficult-to-access populations like those at risk of Alzheimer's. Transform eye-tracking research by enabling remote data collection and large-scale studies, particularly during pandemics.

Open-source software (WebGazer) (Cont.)

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Catalbitch et al., 2021)	<p>Sample with adult readers with dyslexia and typical adult readers.</p> <p>Native speakers of British English, recruited through Bangor University and Proffix, with no history of psychiatric or neurological disease, visual or hearing impairments.</p> <p>Group membership (dyslexic or typical reader) was confirmed via a battery of literacy tests.</p> <p>Automatic exclusion from the experiment occurred if the participant's computer was left unattended mid-task, indicated by a lack of detected clicks in three consecutive trials.</p>	70 (35 readers with dyslexia and 35 typical readers)	Unfamiliar Mandarin Chinese characters and English-like pseudowords.	<p>Standard webcam.</p> <p>WebGazer.</p> <p>GoPro Experiment Builder.</p> <p>Desktop or Laptop Computer.</p> <p>Earphone/Headphone.</p>	<p>The study involved typical adult readers and adults with developmental dyslexia.</p> <p>Participants learned associations between Mandarin Chinese characters and English-like pseudowords over multiple exposures, simulating early stages of letter-to-sound mapping.</p> <p>The novel bindings were presented in varied or consistent screen locations and written varied or consistent contexts (co-occurring distractor items).</p> <p>The goal was to investigate how episodic memory cues, such as spatial and contextual properties, aid in binding acquisition and differ between dyslexic and typical readers.</p> <p>Recognition and recall of the bindings were tested during training and post-training tasks, with eye movements tracked to assess episodic memory retrieval.</p>	<p>Adults with dyslexia have persistent difficulties with cross-modal binding, which is the ability to link visual and auditory information.</p> <p>Both typical and dyslexic readers can use consistent contextual information from episodic memory to assist in audiovisual learning, but dyslexic readers are particularly reliant on this consistency.</p> <p>Dyslexic readers recognized and recalled significantly fewer of the novel cross-modal bindings compared to typical readers.</p> <p>The looking-at-sending behavior, which involves fixating on empty screen locations associated with previously presented stimuli, was generally linked to higher recognition error rates in both groups. This effect was more pronounced in later blocks for bindings encoded in a noncentral location.</p>	<p>The study acknowledges that a larger sample size might be necessary to robustly assess higher-order interactions, which could provide more comprehensive insights.</p> <p>The use of webcam-based eye-tracking technology is still in its infancy, and the study admits that this method may not be as reliable or accurate as more established in-lab eye-tracking equipment.</p> <p>A portion of the eye-tracking data was excluded due to suboptimal face detection values, which could potentially affect the overall findings.</p> <p>The study pre-registered a power analysis, but it suggests that the power might not be sufficient for the interaction of primary theoretical interest, indicating a limitation in the statistical robustness of the results.</p>	The study does not provide specific details regarding future works.
(Hermans & El Mouatadim, 2023)	<p>The study gathered data from students across three primary schools in southern Morocco. Students were selected from two different grades (4th and 5th) and were divided into sub-groups based on their Arabic reading abilities as assessed by their teachers (Low, Medium, Excellent).</p>	61	<p>The participants were given Arabic normal text materials with various configurations such as font sizes, character spacing, and font families to analyze their reading behavior.</p> <p>The tests were designed based on the "L'Alouette" dyslexia test and adapted to reflect the characteristics of the Arabic language, consisting of seven paragraphs and a total of 536 words without numerical expressions, acronyms, or foreign words.</p>	<p>14 inch Lenovo ThinkPad T490.</p> <p>Logitech C270 HD Webcam.</p> <p>HTML5/CSS3.</p> <p>Apache Server 2.4.51.</p> <p>WebGazer.</p>	<p>The study used a web-based application to present Arabic text materials with various configurations to analyze reading behavior.</p> <p>A consumer grade webcam was used to track the gaze points of participants as they read the text displayed on a computer screen.</p> <p>Participants were exposed to Arabic texts with different font sizes, character spacing, and font families to assess their reading behavior.</p> <p>A calibration procedure was conducted using the WebGazer library to ensure accurate gaze point calculation tailored to each subject.</p> <p>The Arabic test used in the experiment was based on the "L'Alouette" dyslexia test, adapted for the characteristics of the Arabic language.</p> <p>The study applied deep clustering techniques to reduce the dimensionality of the eye-tracking data and group subjects using unsupervised k-means clustering based on AutoEye reader output.</p>	<p>The use of Arabic dyslexic-friendly typefaces significantly improved reading performance, especially for students with low reading abilities.</p> <p>Dyslexic readers exhibited longer fixations, more fixations, and shorter saccades compared to non-dyslexic readers.</p> <p>Webcam-based eye-tracking techniques showed potential as an initial diagnostic tool to assess the risk of dyslexia in children.</p> <p>Deep clustering was effectively used to analyze the dataset, which helped in identifying different dyslexia traits among the subjects.</p>	<p>The dataset used for the study was relatively small, consisting of 61 subjects from three primary schools in southern Morocco. This limited sample size may affect the generalizability of the findings.</p> <p>The study relied on consumer-grade webcams, which may not be as precise or reliable as specialized eye-tracking equipment. This could impact the accuracy of the eye-tracking data collected.</p> <p>The study's results could be influenced by environmental variables like lighting conditions, which can affect the performance of webcam-based eye tracking.</p> <p>The Arabic test used in the study was based on the "L'Alouette" dyslexia test and adapted for Arabic characteristics. The specific choice of test and its properties might have an impact on the eye movement patterns observed.</p>	The study does not provide specific details regarding future works.
(Guan et al., 2022)	<p>The study involved higher education learners, native in Chinese and proficient in English as a second language. Participants were recruited voluntarily through electronic posters.</p> <p>They engaged in an online reading task using a platform capable of eye tracking via a webcam. Learners were grouped based on their scores on specific reading comprehension questions, aiming for a fine-grained analysis.</p>	32	CET 4 English reading comprehension passage, which included 5 multiple-choice questions.	<p>Standard webcam.</p> <p>WebGazer.</p> <p>PDF.js.</p>	<p>Eye movement data were collected through a webcam-based online reading platform while participants engaged in an English reading comprehension test.</p> <p>Participants read and answered multiple-choice questions. Their eye movements were tracked to predict reading performance based on scored and unscored responses to each question.</p> <p>Machine learning algorithms, including Decision Trees, Random Forest, Logistic Regression, Naive Bayes, and Support Vector Machine, were used to analyze the predictive effect of eye-movement behavior on reading performance.</p>	<p>Dividing learners into scored and unscored groups based on their performance on individual questions allowed for a more detailed analysis, leading to better prediction results.</p> <p>The combination of Regression Group (RG) and Other Indicator Group (OG) indicators yielded better prediction results across different questions.</p> <p>These findings suggest that selecting meaningful eye-movement indicators and appropriate algorithms based on the type of reading question can improve the prediction of reading performance.</p>	<p>The study had a relatively small sample size, which limits the abundance of data and may affect the generalizability of the findings.</p> <p>The analysis related to text content was conducted manually, which could be a constraint for scaling up the application of the platform in future research.</p>	<p>Integrating Natural Language Processing (NLP) algorithms to automate and dynamically track the association between behavior and cognition during reading.</p> <p>Expand the research for large-scale application, which would allow for more comprehensive data analysis.</p> <p>Develop methods for automated analysis related to text content.</p>
(Hutt et al., 2024)	<p>Study 1: A relatively homogenous group of students, predominantly white, aged 18-25, who participated for course credit.</p> <p>Study 2: A more diverse group of participants recruited through Proffix, with less than 20% white participants, aged 18-25, who were compensated for their participation.</p>	<p>Study 1: 195</p> <p>Study 2: 173</p>	<p>65 narrative stories taken from Cranford and Moss (2018).</p> <p>Each story consisted of three sentences and had three possible endings, with only one appropriate ending after reading all three sentences.</p> <p>The stimuli were delivered in two conditions: audio, where participants heard a reading of each sentence, and visual, where participants read each sentence presented on the screen.</p>	<p>Standard webcam.</p> <p>WebGazer.</p>	<p>Two distinct groups were involved in the study: university students and a diverse sample from Proffix, an online platform.</p> <p>The study used WebGazer, a webcam-based eye-tracking software, to collect gaze data during an online reading comprehension task.</p> <p>Participants engaged in a narrative anticipation task, reading short stories and making inferences about their endings.</p> <p>Gaze locations were inferred in real-time using participants' webcams, and the data was collected as they interacted with the task.</p> <p>Both global and local gaze features were calculated from the raw data to predict task-unrelated thought (TUT) and reading comprehension.</p> <p>Various classifiers were employed to relate gaze features to TUT probes and comprehension scores, using cross-validation methods for accuracy.</p>	<p>Found a positive correlation between predicted rates of TUT and actual comprehension levels, indicating that the eye-tracking data could predict comprehension levels.</p> <p>The study tested the generalizability of their models across two different populations, showing that the webcam-based eye tracker could provide accurate and precise gaze measurements for predicting TUT and reading comprehension.</p> <p>By using diverse samples, the study also aimed to address potential algorithmic bias in eye-tracking technology. The findings indicate no significant bias in the eye-tracking system that could have impacted its ability to collect understandable gaze data in both study populations.</p> <p>In addition, slicing analysis showed that detectors were mostly robust to individual differences in race/ethnicity and whether participants wore eyeglasses, with one exception. Additionally, they were resilient to conditional differences like lighting.</p>	<p>Webcam systems tend to be less accurate and precise than specialized video-based systems, especially when deployed in real-world conditions where factors like lighting and head position are not controlled.</p> <p>The sampling rate of the webcam-based eye tracker is variable and cannot be guaranteed, which can affect the quality of gaze data.</p> <p>There is an observed drift in the gaze recordings, indicating that calibration may not always remain stable throughout the experiment.</p> <p>The robustness of gaze tracking from raw data is limited, and the study did not correct for sampling rate or attempt to calculate fixation durations or identify saccades.</p>	<p>Investigate alternative task presentations and stimuli formats for improved ecological validity.</p> <p>Collect larger datasets to enable more complex data mining and machine learning.</p> <p>Consider more in-depth feature engineering processes with additional global and local features.</p> <p>Explore the mechanistic relationship between eye gaze and cognitive constructs using research-grade tracking systems and how this could translate to this more accessible tracking option.</p> <p>Validate automated detectors to reduce reliance on thought probes for measuring internal states like task-unrelated thoughts (TUT).</p>

Open-source software (WebGazer) (Cont.)

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(K. Yang & Kojanich, 2021)	Participants from Amazon Mturk located in the United States with a 95% or higher HIT approval rate and a laptop with a webcam.	135 but only 49 successfully passed the initial calibration and validation stage and completed the study.	Decision-making tasks conducted on Amazon Mturk where participants rated their desire for 70 snack food items on a discrete scale from 0 to 10, indicating their liking for each item. The binary choice task involved selecting their preferred food item out of pairs generated from the rated items.	Standard webcam. WebGazer. iPhon.	Investigate the feasibility of online web-based eye-tracking in behavioral research. Adapted the WebGazer JavaScript toolbox, originally designed for monitoring eye movements during internet browsing. Subjects underwent an initial calibration and validation process, adjusting their webcam position to optimize gaze tracking. Subjects rated 70 snack foods and then made binary choices between food items while being eye-tracked. Analysis performed regarding spatial resolution, temporal resolution, correlations between gaze and choice, and effects of individual dwell times, and last fixation bias.	Successfully integrated the WebGazer eye-tracking toolbox into a widely used JavaScript library for behavioral research (jQuery). This adaptation allowed for online eye-tracking, despite the challenges posed by the shift from in-lab to internet-based experiments. By adjusting the WebGazer code, the study achieved a consistent temporal resolution of approximately 245 ms throughout the experiment. This time resolution is comparable to many in-lab eye-tracking systems. Although WebGazer's spatial precision is slightly worse than in-lab eye-tracking, it still allows for the use of four to six areas of interest (AOIs) without significant degradation in data quality. The study replicated important relationships between gaze and choice observed in prior in-lab studies. These include correlations between dwell time and choice, the effects of individual dwell times, and the last fixation bias. The effect of choosing the last seen option may be smaller in online eye-tracking compared to lab-based studies.	The initial calibration and validation procedure can be time-consuming, especially in online studies. However, the authors have made efforts to reduce this time by adjusting the procedure and code. Online eye-tracking tools control over subjects' environment and hardware. Factors like lighting, webcam quality, and subject positioning can affect data quality. While online eye-tracking provides reasonable spatial precision, it may not match the level achieved in lab-based eye-tracking. The authors estimate that using four to six areas of interest (AOIs) is feasible without significant degradation in data quality, however it's still one unresolved issue. The time interval between gaze estimates can be limited by webcam sampling rates (typically around 30 Hz). Achieving higher temporal resolution may require in-lab eye-tracking.	The authors suggest further exploring the use of online eye-tracking in behavioral research. While this study demonstrates the feasibility of web-based eye-tracking, there is still room for improvement and optimization. Continue integrating web-based eye-tracking into existing experimental paradigms. Future work could explore other integration possibilities to make eye-tracking more accessible for researchers. Continue developing simple procedures for subjects to follow, especially considering the lack of control over subjects' environments in online studies. While the study achieved steady temporal resolution (24.85 ms) using WebGazer, further improvements could be made. Investigating alternative algorithms and methods to reduce computational demands may enhance temporal precision. Explore using more AOIs without significant degradation in data quality, however, they suggest that spatial precision should be considered when designing experiments. Future research should compare online eye-tracking results with different subject populations. Understanding potential differences between online and lab-based studies will help validate the approach.
(Jung et al., 2021)	The study does not provide specific details about the sampling method used. That's because the study focuses on implementing eye-tracking for persona analytics, and it discusses integrating webcam-based eye-tracking into an interactive persona system.	The study does not provide specific information regarding the number of participants.	The study does not provide specific information regarding the stimuli used.	Standard webcam. WebGazer.	Research Questions (RQs): Which personas were most/least viewed? What persona information was most/least viewed? What were the most/least common transitions between information elements? What was the average duration of viewing a persona? How long did User A view Persona X? The study integrates a webcam-based eye-tracking module into an interactive persona system. WebGazer.js is used to track users' gaze fixations on persona profiles. The system logs gaze movements during user interaction with personas. Metrics include time spent per persona, time spent on each information element, and transitions between elements.	Addressing the scarcity of behavioral user studies, especially when physical studies are difficult or impossible. A webcam-based eye-tracking module was integrated into an interactive persona system. This facilitated remote user studies, even during exceptional times when physical studies were not feasible. The system tracked 120 UI elements within the persona profile. Accuracy of online eye-tracking ranged from ~50-84%.	Online eye-tracking systems, while convenient, may impose quality challenges. Some users' gaze can be tracked better than others due to factors like lighting conditions and webcam limitations. The accuracy of online eye-tracking varies. Currently, researchers must download data reports and analyze them separately.	Accuracy can vary among users, so setting an accuracy threshold and excluding participants who don't meet it is advisable. Larger sample sizes can compensate for data loss due to quality issues. Multiple iterations of calibration may improve accuracy, but further study is needed. Future work could develop tables and visualizations directly within the persona system's UI for the purpose of data analysis.
(Robal, 2019)	The study does not provide specific details about the sampling method used.	20	50 benchmark tasks as cue cards in a random order. These tasks include actions such as "lean your head away from the screen", "stand up", and various body postures like tilting the body and head.	Tobi X2-30 Compact. Standard webcam. WebGazer.	Tobi X2-30 Compact and WebGazer.js were chosen for comparison. 20 participants performed 50 tasks in a controlled office environment using a Dell laptop with a built-in webcam and Tobi eye tracker. Tasks represented common web user behaviors to simulate real-life conditions. Gaze data was collected from both Tobi and WebGazer for analysis. Two main perspectives were considered: the accuracy of gaze tracking and the distance between predicted gaze points from Tobi and WebGazer.	Users are concerned about privacy when allowing webcam access to websites. The study emphasizes processing gaze data locally to address privacy concerns. Using webcams and local data processing ensures user privacy and eliminates the need for additional software installation, making the solution easily accessible. Without prior calibration, WebGazer's gaze prediction error rate was approximately 250 pixels, making it suitable for gaze direction establishment but less precise for tasks requiring high accuracy. On average, the distance between WebGazer's predicted gaze points and Tobi's detected gaze was around 400 pixels, suggesting a range of 400 pixels for valid gaze predictions. The cinetrack face tracking library used in WebGazer showed aggressiveness in predicting gaze even in the absence of valid eye gaze, resulting in false-positive predictions in 25% of cases.	The error rate for WebGazer doubles when it is used spontaneously without calibration. This limitation affects the precision of gaze tracking, making it less suitable for applications requiring high accuracy. Although WebGazer can provide gaze predictions, the majority of predicted gaze points remain within the range of 400 pixels from the actual gaze point. This limitation implies that WebGazer may not be suitable for tasks requiring precise gaze tracking. WebGazer's accuracy could potentially improve if users were allowed to self-calibrate during normal web browsing behavior. However, in this experiment, this option was excluded.	Allowing WebGazer to self-calibrate during normal web browsing behavior for users could potentially enhance its accuracy. Additional measures on the system side, such as validation of predicted gaze points, are recommended. This validation process could help improve the accuracy of spontaneous WebGazer instances for gaze prediction, particularly for use in intelligent systems like attentive user interfaces. Future research should explore ways to mitigate the issue of aggressiveness in the cinetrack face tracking library used within WebGazer, especially in scenarios where users may exhibit a lot of sudden movements in the camera viewport.
(Semmelmann & Weight, 2018)	Employed a combination of in-lab and online data collection. Participants were recruited at Ruhr-Universität Bochum and online participants recruited through crowdscience platform.	Initially, 30 in-lab participants were recruited for the study but one participant was excluded due to missing data. Thus, the final set for in-lab participants was 29. For the online data collection, a total of 84 consent form transmissions were received. After excluding incomplete data sets and multiple participations, the authors ended up with 28 complete online data sets.	Participants were shown instructions on how to position themselves before each block of tasks. Participants were provided with a live feed of their webcam. Participants received a short notice to try to stop moving their head once the block of tasks started.	Macbook Pro. Webcam. Experiment was programmed in HTML5, supported by jQuery 1.12.3 and PHP 5.3.3 on an Apache 2.4.18 server. WebGazer.	Participants fixated on a dot, and the accuracy of eye tracking was assessed during this task. Participants followed a moving target, and the stability of eye tracking during motion was evaluated. Participants freely viewed facial images, allowing analysis of gaze patterns. For in-lab settings, participants were seated in front of a MacBook Pro with a built-in webcam in order to obtain data. For online settings, the authors collected data through crowdscience platforms.	Participants exhibited a clear saccade (sapid eye movement) towards the target dot after its appearance. In in-lab data, users achieved an average offset of about 15% of screen size (approximately 3.94° visual angle), while online data had an 18% offset. Variance was higher in online data. Regarding the pursuit task, participants correctly followed the motion of the target stimulus. No significant difference in accuracy between in-lab and online data and pursuit speeds were faster than the actual target motion. Regarding the free-viewing tasks, participants predominantly fixated on the eyes of facial images and there was no significant accuracy difference between in-lab and online data.	The online data acquisition had a lower sampling rate (30s) compared to in-lab data. Lower fps might impact the temporal resolution of gaze data. Online participants took longer to complete the experiment compared to in-lab participants. Longer experimental sessions might affect participant engagement and fatigue. Webcam-based eye tracking relies on consumer-grade webcams, which may have limitations in terms of resolution and accuracy. Calibration procedures may be less precise in an online setting. The study was conducted in Germany, limiting the generalizability of findings to other cultural contexts. Online participants may differ from in-lab participants in terms of demographics and motivation. Microsaccades and measurement inaccuracies could contribute to higher pursuit speeds. Some deviations in pursuit speed might be due to system-related factors.	Investigate the optimal number of calibration and validation trials needed for high gaze estimation accuracy to minimize the time taken up by these phases and increase participant comfort. Implement gamification techniques in experimental paradigms to enhance participant engagement and reduce dropouts during studies. Explore methods to improve the absolute accuracy of gaze estimation, potentially through careful calibration, instructions, and pre-selection. Conduct experiments combining in-lab studies with web technology to differentiate between attentional effort and measurement inaccuracy, particularly regarding gaze estimation in different regions of the screen. Design experiments to investigate gaze resolution accuracy, especially considering the lower accuracy in the lower portion of the screen due to webcam positioning. Re-investigate automatic calibration through mouse movement and/or clicks to improve data acquisition and participant comfort during online eye-tracking studies.

Proprietary/custom solutions

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Rogers & Al Madi, 2023)	The study does not provide specific details about the sampling method or participant selection strategy.	The study does not provide specific information regarding the number of participants.	The study aims to understand the cognitive processes involved in software development by observing developers' emotional responses to specific source code elements. However, it does not specify the exact nature of the stimuli used in the study.	Standard webcam. Eye Tracking Module. Affect Recognition Module. GazeTracking Tool.	<p>Uses an unmodified webcam to capture video of developers interacting with code.</p> <p>Processes video frames through two modules, an eye-tracking module for gaze estimation and an affect recognition module for emotion inference.</p> <p>Employs a linear regression model using calibration points to predict the developer's gaze point on the screen.</p> <p>Implements a neural network trained on images of different emotions to detect the developer's emotional state.</p>	<p>The study presents a tool that uses an unmodified webcam for eye tracking and affect recognition during software development tasks.</p> <p>It combines gaze prediction, where the developer is looking on the screen, with emotion detection from facial expressions to understand developers' emotional states.</p> <p>The tool evolved through versions, from still-frame pupil recognition to real-time gaze prediction and emotion recognition.</p> <p>The research has potential applications for software development researchers, educators, and practitioners, such as improving understanding of cognitive processes and aiding in educational settings.</p>	<p>The authors aim to improve the accuracy of the eye tracking module, as webcam-based tracking might be less precise than dedicated eye trackers.</p>	<p>There is a goal to collect more data from volunteer developers in the field to understand the tool's benefits in remote and large-scale settings.</p> <p>The authors plan to conduct a controlled lab experiment to compare the accuracy of their tool with a research-grade eye tracker in software development tasks.</p>
(Hutt & D'Mello, 2022)	<p>The authors used pre-existing datasets for their research. The datasets they used were collected through specific sampling methods.</p> <p>These methods involved selecting participants from certain environments (like a university or a high school) and having them perform specific tasks (like reading or interacting with an Intelligent Tutoring System).</p>	The authors used pre-existing datasets for their research. Total number of participants in all datasets = 377.	<p>The authors used pre-existing datasets for their research.</p> <p>Dataset 1: Participants read educational text from the book "Soap Bubbles: Their Colors and the Forces that Mold Them," which was presented on-screen one page at a time.</p> <p>Dataset 2: Participants interacted with an Intelligent Tutoring System (ITS) called GuruTutor, which engaged students through natural language conversations and multimedia output.</p> <p>Dataset 3: Participants interacted with an Intelligent Tutoring System (ITS) called GuruTutor, which engaged students through natural language conversations and multimedia output.</p>	Open source software OpenFace. DBSCAN.	<p>The study introduces an unsupervised algorithm to approximate gaze fixation features from webcam data without the need for calibration. This method bypasses the traditional calibration process required for screen-based coordinates.</p> <p>Evaluation is conducted using three datasets that include different user interfaces, environments (laboratory and classroom), and comparison gaze trackers. This allows for testing the robustness of the approach across various settings.</p> <p>The study uses the DBSCAN clustering algorithm to derive fixation approximations from camera-centered gaze vectors. The algorithm operates without prior knowledge of user screen or calibration details, focusing on higher-order gaze features like fixations and saccades.</p> <p>The derived webcam-based gaze features are used to model user engagement and comprehension, demonstrating the potential for long-term, gaze-enabled user interfaces in real-world settings.</p>	<p>The study found that webcam-based gaze features generally correlated positively with those computed from traditional gaze trackers. Specifically, measures of central tendency of fixation durations showed moderate to strong correlations.</p> <p>The study demonstrated that fixation approximations from webcam data could predict comprehension with an AUROC of 0.77 and mind wandering with above-chance accuracy.</p> <p>The study mentions that saccades are important yet challenging to accurately capture with webcams, suggesting limitations as a sole replacement for PCCR trackers. However, they could offer an alternative in tasks with higher accuracy requirements, like reading comprehension.</p>	<p>The webcam data was collected at a lower frequency (approximately 12.5Hz) compared to the PCCR data (120Hz for Dataset 1 and 300Hz for Dataset 2), which is an order of magnitude difference. This lower sampling rate can lead to inaccuracies in gaze estimations.</p> <p>The method is more accurate for measures of central tendency of fixation durations but very inaccurate at measuring saccadic features, such as the proportion of horizontal saccades.</p> <p>There was data loss in both webcam and PCCR tracker data, with the webcam data showing more robustness but less accuracy compared to traditional gaze tracking.</p> <p>The correlations between webcam and PCCR-computed features were lower in real-world environments (Dataset 3) compared to laboratory settings (Datasets 1 and 2), potentially due to increased noise and distractions in less controlled environments.</p> <p>Particularly for mind wandering detection, the reliance on self-reported data for ground truth values can introduce subjectivity and affect the reliability of user modeling tasks.</p>	<p>Investigate how individual factors like wearing glasses or race/ethnicity may affect the accuracy of gaze estimations.</p> <p>Determine whether errors in gaze estimation are due to the methodology or the gaze detection of OpenFace, and explore ways to reduce these errors.</p> <p>Explore the use of deep-learning appearance-based gaze estimations combined with the clustering approach to potentially yield more accurate results.</p> <p>Examine if the hyperparameters used in the clustering approach can be preselected and automatically adjusted to enhance the methodology.</p> <p>Investigate the impact of increasing the webcam's sampling rate on estimation accuracy and consider the limitations of detecting detailed eye movements with traditional webcams.</p> <p>Address privacy issues by informing users about data collection and usage, and explore real-time feature extraction without video recording to reduce privacy concerns.</p>
(Adiani et al., 2022)	The study does not provide specific details about the sampling method used. Only mentions an experiment involving 12 participants, some with eyeglasses and some without, to assess the suitability of webcam-based eye tracking in virtual job interview simulations.	12	<p>A fixed image of a virtual environment was used for the experiment, which had regions of interest (ROIs) marked for the eye-tracking tasks.</p> <p>Participants engaged in a 9-point calibration task using a calibration interface, where they clicked on dots and their eye movements were tracked.</p> <p>Participants interacted with a validation component where they freely clicked on areas of an image, and the system tracked their gaze to detect if it fell within the correct ROI box.</p>	Standard webcam. Surface Pro 7. OpenCV and Dlib Libraries. Flask Application. PyAutoGUI.	<p>A webcam-based eye tracking algorithm is implemented and tested for its suitability in virtual job interview simulations, with a focus on detecting interview-relevant regions of interest (ROI).</p> <p>The study involved 12 participants, some with eyeglasses, to assess the impact of eyeglasses on calibration error and ROI detection, concluding that eyeglasses do not significantly affect the results.</p>	<p>The average calibration error across all participants was 169.4 pixels with a standard deviation of 76.4 pixels. The presence of eyeglasses did not significantly affect the mean calibration error.</p> <p>The study found that maintaining the same head position and distance during tasks after calibration is crucial for accurate Region of Interest (ROI) detection.</p> <p>The quality of eye images for feature extraction was influenced by the reflectiveness of eyeglasses, which could lead to higher mean prediction errors.</p> <p>For the webcam-based eye tracking method to provide accurate predictions, participants need to maintain the same position throughout the study.</p>	<p>The accuracy of the eye-tracking method is contingent on the participant maintaining the same head position and distance throughout the experiment. Any movement can significantly affect the calibration and the subsequent gaze prediction results.</p> <p>The presence of eyeglasses does not significantly affect calibration error, but the quality of eye images can be compromised by reflective eyeglasses, leading to higher mean prediction errors.</p> <p>The current method may not perform well under natural head movements, which is typical during a job interview, as participants may not be conscious of their level of movement.</p> <p>The results reported are preliminary and based on a small sample size, indicating that further testing is needed to validate the findings.</p>	<p>Test the webcam-based eye tracking method with an actual virtual interview simulation using their novel tool CIRVR to understand its performance under natural head movement while the participant speaks.</p> <p>Additional calibration on Regions of Interest (ROI) using the ROI task before free viewing of the interview interface, which may lead to improved ROI detection.</p> <p>Explore error correction methods such as Kalman Filter and Extended Kalman Filter on the gaze prediction value to improve the gaze prediction results.</p> <p>Investigate the impact of higher frame rates (>6 images per click) on the variation in prediction results.</p>

Proprietary/custom solutions (Cont.)

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Monteiro et al., 2022)	The study does not provide specific details about the sampling method used.	The study does not provide specific information regarding the number of participants.	The study does not provide specific information regarding the stimuli used.	Standard webcam. CNN/RNN Algorithms. WebRTC. Flask. OpenCV and Dlib Libraries. Caffe Model.	<p>The study employs Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to monitor students during exams. These algorithms analyze video feeds to detect suspicious behaviors and ensure candidates' authenticity.</p> <p>Key features include tracking eye movements, lip movements, and detecting additional people in the frame. This is to prevent cheating and ensure that the student is alone and focused during the exam.</p> <p>The proposed system includes functionalities like screen sharing monitoring, sending warning messages, and the ability to terminate an exam if cheating is detected.</p> <p>The system is developed using WebRTC for real-time video communication, Twilio API for cloud communications, and Flask, a web framework for building the application.</p> <p>The study acknowledges challenges in online learning, such as user authentication and maintaining exam integrity. It aims to overcome these with the proposed proctoring system.</p>	<p>The study is more focused on the implementation phase and does not provide much information regarding findings.</p> <p>Nevertheless, findings suggest that the "Proct-Xam" system could potentially enhance the supervision of online exams, making them more secure and trustworthy.</p>	The study does not provide specific details regarding limitations.	The study does not provide specific details regarding future works.
(Yang et al., 2023)	The study involved English-speaking adult learners who were novice programmers recruited from various academic disciplines. Experienced tutors with teaching experience in computer science and proficiency in Python were selected. Each tutor conducted three learning sessions under different conditions, randomized to control for learning effects. Learners were chosen based on criteria ensuring they were beginner programmers, with no prior text or block-based programming experience, to generalize findings to novice learners.	39 learners and 13 tutors	<p>10-minute video from Udacity introducing basic Python programming concepts.</p> <p>A browser-based notebook hosted on Deepnote containing instructional material and exercises focused on control flow (if, elif, else statements) and for-loops, adapted from a popular online course.</p>	<p>Chrome Extension. Deepnote Notebook. Webcam Eye-Tracking using GazeCloudAPI.</p>	<p>Three conditions were tested: screensharing, synchronous editing, and synchronous editing with awareness tools.</p> <p>A Chrome extension was developed to enable real-time sharing of gaze and cursor data.</p> <p>Tutors taught learners in a remote setting using Zoom, working through a Deepnote notebook with instructional material.</p> <p>Data was gathered through pre- and post-tests, post-study surveys, instructor exit interviews, and session video analysis.</p> <p>The study used thematic analysis, reference and typing coding, and statistical analysis to assess learning gains and the impact of the tools on teaching and learning.</p>	<p>The combination of synchronous editing with awareness tools significantly improved learning outcomes compared to screensharing alone.</p> <p>The awareness tools enhanced tutors' abilities to communicate with students, track their understanding, and establish a greater sense of presence.</p> <p>Despite some inaccuracies, webcam-based gaze sharing was found to be beneficial for tutors to develop awareness of students' actions and foster a sense of presence.</p>	<p>The gaze-tracking accuracy was not as precise as a dedicated sensor, with an average error of 121.93 pixels, which could affect the reliability of gaze-related findings.</p> <p>The study is experimental, and its findings may not generalize to all remote tutoring settings or subjects beyond computer science.</p> <p>The learners were beginner programmers with no prior experience, which limits the applicability of the results to novice learners only.</p> <p>Tutors were instructed to keep their teaching consistent across conditions, which may not reflect the natural variability in teaching styles.</p> <p>The use of a Chrome extension and specific tools like Deepnote may have influenced the results based on the participants' familiarity with these technologies.</p>	<p>Suggests that future studies should continue to evaluate the effectiveness of various tools in remote learning environments.</p> <p>Explore the design and implementation of new awareness mechanisms to improve distance learning.</p> <p>Future research could focus on refining Gaze-Sharing Technology to make it more accessible and effective for educational purposes.</p> <p>Further exploration of the instructional benefits of real-time gaze-sharing using webcam eye-tracking.</p>
(Anwar et al., 2018)	The study does not provide specific details about the sampling method used. However, the study aims to develop an Emotion Recognition and Eye Gaze Estimation (EREGE) system and mentions that the system is tested on a group of people for eye gaze estimation accuracy.	5	The study does not provide specific information regarding the stimuli used.	<p>Active Shape Model (ASM) Tracker. Support Vector Machine (SVM) Classifier. Orthography and Scaling with Iterations (POSIT). Standard webcam.</p>	<p>Implements a Support Vector Machine (SVM)-based classifier to recognize seven emotions (neutral, happiness, sadness, anger, disgust, fear, and surprise).</p> <p>Creates a head model using Active Shape Model (ASM) and Pose from Orthography and Scaling with Iterations (POSIT) algorithms for head tracking and position estimation.</p> <p>Determines eye gaze direction based on the head model.</p>	<p>The proposed system achieved a correctness score of 93%.</p>	<p>Relies solely on a PC's webcam which may affect accuracy and robustness, especially in challenging lighting conditions or with low-quality cameras.</p> <p>Variability in eye movement patterns and head motion can introduce errors.</p> <p>The system assumes that both eyes are in the same distance from the camera, neglecting depth differences. This simplification may lead to inaccuracies, especially when users are at varying distances from the camera.</p> <p>EREGE's advantage lies in its simplicity but lacks the sophistication of commercial solutions.</p>	<p>The EREG system can be further explored for monitoring health by recording gaze direction.</p> <p>It can also be used to track emotional behavior in autistic children by analyzing their eye movements and recognizing facial expressions.</p> <p>EREGE can serve as a tool to analyze how customers view advertisements and which products catch their attention in online or public stores.</p> <p>EREGE exemplifies successful HCI. Further research can enhance its usability, accuracy, and real-time performance. Investigating user experience and interaction information would be beneficial.</p>

Standard webcams without special hardware

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Azem et al., 2023)	The study does not provide specific details about the sampling method used. However, it mentions that the authors interviewed a researcher who had used the METRIC system for a user study.	1	The study does not provide specific details about the nature of the stimuli used. It focuses on describing the METRIC system, a tool for measuring engagement through remote interactions of customers, and its capabilities for collecting, measuring, analyzing, and reporting user engagement with online systems	Standard webcam. Eye-tracking JavaScript library but does not mention a specific one. PupilCore Core tracker for physical eye tracking in lab settings.	METRIC collects data through direct user-computer interaction, such as mouse clicks, eye gaze, and mouse movements, as well as pre- and post-study surveys. The system supports both on-site and remote user studies, utilizing various tools and techniques like online surveys, interviews, and usability testing to gather qualitative and quantitative data. METRIC is evaluated using the System Usability Scale (SUS) and user interviews to assess its usability and functionality. It provides sophisticated reporting features that display metrics over time, page views, sessions, and user interactions in the form of heatmaps and tree diagrams.	The user appreciates METRIC's ease of use and its range of functionalities, particularly the ability to use both video and image materials for user studies, track mouse clicks, and use two different eye-tracking systems. The inclusion of Areas of Interest (AOIs) was found helpful in analyzing participant gaze patterns. The user found it convenient that data can be easily exported in Excel CSV format. Overall, METRIC was perceived as user-friendly, with easy navigation and clear button and feature locations. While METRIC was deemed user-friendly, the user suggests a feature for easier creation of multiple user studies with similar settings to enhance scalability. Suggestions were made for the inclusion of guiding features, such as pop-up explanations when hovering over features, to aid less experienced users. While features were well integrated, clearer explanations were recommended, especially for first-time users, regarding how different integrations work and understanding METRIC's full potential.	While the website's format and functionality resemble other websites, understanding METRIC's full potential may require additional guidance, suggesting a potential learning curve for users. Clearer explanations were suggested for first-time users regarding how different integrations work and what kind of data the system outputs. The absence of a feature for easier creation of multiple user studies with similar settings could pose scalability challenges, particularly for users managing multiple studies simultaneously.	Enhancing the accuracy of the remote eye tracking tool, which is already deployed and functional. Introducing additional user study tool features such as audio recording and transcription. Incorporating active user interface testing features, including experiments in collaborative work settings.
(Sodoké et al., 2021)	Recruiting a group of clinicians from the Department of Anaesthesiology at the Université de Montréal. The selection criteria for participants are: Experts: Defined as hospital staff members with more than 8 years of experience. Novices: Identified as resident clinicians within the first or second year of their residency program.	24 (12 experts and 12 novices)	Participants were asked to visualize a video simulation within a webcam-based online platform called EyeLab. The simulation was based on the CICO (Cannot Intubate/Cannot Oxygenate) algorithm from the Difficult Airway Society, which is used to manage unanticipated difficult intubation in adults. During the simulation, participants had to observe and verbalize their observations in identified zones or AOIs to perform their clinical reasoning, following a think-aloud protocol.	Standard webcam.	The study involved 12 expert clinicians with over 8 years of experience and 12 novice clinicians in their first or second year of residency from the Université de Montréal's Department of Anesthesiology. Eye tracking technology was used, specifically a webcam-based platform called EyeLab, to collect eye gaze data as participants engaged with a video simulation. Participants were asked to watch a video simulation of a clinical scenario and verbalize their observations in real-time using a think-aloud protocol. The simulation was based on the CICO (Cannot Intubate/Cannot Oxygenate) algorithm. The eye gaze data, represented as spatiotemporal sequences, were analyzed to compare fixation counts, durations, and scanpaths between novices and experts. Techniques like heatmaps, Smith-Waterman algorithm, and deep convolutional LSTM were used for comparative and classification analyses.	The insights from the comparative analyses were used to build a webcam-based Intelligent Tutoring System (ITS) aimed at reinforcing the learning stages of novice clinicians by providing cues from experts' implicit knowledge in visual attention. Using the Smith-Waterman algorithm, the study compared the scanpaths of clinicians, finding that experts had more consistent and similar eye movements, which could be used to construct a typical expert scanpath for training purposes. EyeLab, a web application developed for the study, captures learners' eye movements via webcam and provides tutoring services by replaying video recordings with visual traces of their scanpath, offering metacognitive support to improve clinical reasoning skills. These findings suggest that the ITS and EyeLab can effectively support novice clinicians in developing visual situational awareness, a critical skill in fields like anesthesiology.	The use of webcams for eye tracking is a novel approach that may not be as precise or reliable as traditional laboratory-based eye tracking devices like the Tobii TX-300. The study involved a relatively small sample of 12 experts and 12 novices, which may limit the generalizability of the findings. The ITS (Intelligent Tutoring System) and eye tracking were performed using an online platform, which may not perfectly replicate the conditions of a real clinical environment. There is no formal domain knowledge about eye movements during clinical reasoning, which means the machine learning techniques used to extract scanpath patterns are based on limited data.	The authors plan to conduct a larger scale experiment with more participants and clinical cases to measure behavioral learning gains and learning persistence. The authors also aim to add an emotional state to the learner model by using facial expressions.

Web-based platforms and remote data collection

Article Reference	Sampling Method	Number of Participants	Stimuli	Equipment	Methodology	Key Findings	Limitations	Future Works
(Sauter et al., 2022)	The study recruited students as participants via the Prolific platform.	Initially 89 participants, but after exclusions, the final set of participants included 27 individuals. Exclusions were based on not managing to do the initial eye tracking calibration and attention check performance.	The study used a set of six videos, including both conference presentations and e-learning videos. Four videos were recorded for an online psychology conference, featuring both male and female presenters. Two of these videos had a visible presenter, while the other two did not. Additionally, two e-learning videos were selected for comparison. Participants watched three out of the six videos in a randomized order. After each video, they answered comprehension questions related to the content. An attention check was included to ensure participants' engagement with the videos.	Standard webcam. Labvanced platform, which includes a cloud-based webcam-based eye tracking module.	Webcam-based eye tracking data was collected while participants watched educational videos. Labvanced, a cloud-based solution, was used for eye tracking. The algorithm estimated gaze points on the screen based on deep learning. Eye movement synchronicity was calculated as a weighted inter-subject correlation (ISC) of gaze position and velocity. The measure aimed to capture attention based on eye movement patterns. The study explored whether eye movement synchronicity could predict test performance.	The average confidence in the gaze samples was around 72%. The eye-tracking sampling rate varied among participants from 4 Hz to 13 Hz (mean: 7 Hz, SD: 20 Hz), depending on their device's technical properties. The study did not find a clear correlation between eye movement synchronicity and test performance. Specifically, the synchronicity measure based on eye movements did not consistently predict participants' performance in comprehension quizzes.	The study relies on webcam-based eye tracking, which suffers from varying spatial and temporal resolutions due to the use of standard webcams. The low and inconsistent sampling rates (typically max. 30Hz) compared to laboratory eye-tracking systems (1000Hz) pose challenges in data quality and reliability. The study highlights the difficulty of obtaining reliable correlations between eye movement synchronicity and test performance using non-ideal data. The study attempted to reproduce earlier findings that eye movement synchronicity predicts performance in a comprehension quiz. However, they were unable to replicate these results with unreliably-sampled data, highlighting the challenges faced by webcam-based eye tracking in practice. Although preliminary, the study suggests that the presence of a visible presenter (especially in the corner of the screen) might influence test performance. The estimated confidence in gaze samples was around 72%. Cleaning the data for unreliable samples would result in a loss of nearly a quarter of trials.	Enhancing the reliability and accuracy of webcam-based eye tracking algorithms. Addressing challenges related to varying spatial and temporal resolutions is crucial. Additionally, understanding how confidence values are estimated in gaze samples is essential for informed data cleaning. Investigate whether attention measures based on eye movement synchronicity can be applied at the individual student level. While average values provide insights, adapting algorithms to work for each student is essential for personalized learning experiences. Explore the impact of presenter visibility on test performance. Preliminary findings suggest that having a visible presenter (especially in the corner of the screen) might influence attention and subsequent learning outcomes. Further experimental manipulation is needed to validate this effect. Validate attention measures in real-life scenarios, especially for students using standard hardware with potentially low temporal resolution. Building adaptive learning algorithms based on eye movement data remains challenging but promising.



NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa