Use Cases and Limits of Webcam eye tracking for Cartography Research

Master Thesis

for the attainment of the Master's degree "Master of Science", abbreviated "MSc"

submitted within the University Master Program for Further Education
"Geographical Information Science & Systems – (UNIGIS MSc)"
at the Department of Geoinformatics - Z_GIS,
Faculty of Digital and Analytical Sciences,
University of Salzburg

submitted by

Rahimeh Gharibpour Poshtiri

Supervisor:
Dr. Merve Keskin

May, 2025

Science Pledge

By my signature below, I certify that my thesis is entirely the result of my own work. I have cited all sources I have used in my thesis and I have always indicated their origin.

Munich, 26.05.2025	Rahimeh Gharibpour Poshtiri
Place, Date	Signature

To all the women and girls of my homeland

I speak out of the deep night,
Out of the deep darkness.
Your courage breaks the silence,
Not only for you, but also for all who long for light.
Your voices, your steps, your dreams,
They are the revolution.
This is for you,
The soul of a freer Iran.

Abstract

Eye tracking has gained much attention as a method for studying map cognition, providing valuable insights into how users process spatial information. This thesis presents webcam-based eye tracking as a low cost, scalable solution for cartographic research, and specifically investigates the usability of maps and user experience while using cartographic products. This study uses a subset of stimuli and tasks from a prior lab-based experiment (CartoGAZE, 2023) which presented a spatial memory experiment consisting of recognizing road and hydrographic features on 2D static Google road maps. Map users' cognitive load was assessed through behavioral metrics such as response times and success rates, as well as eye tracking metrics including average fixation duration, the number of fixations per second, average saccade length. The results of the webcam eye tracking experiment were compared with the results from lab-based studies. Thirty-five participants took part in an online experiment, with data from 28 participants included in the analysis after applying a 70% calibration accuracy threshold. The findings indicate that webcam-based eye tracking can replicate general attentional pattern such as average fixation duration, the number of fixations per second, average saccade length. However, the method demonstrated limitations in precision and data quality, primarily due to lower sampling rates (15–20 Hz compared to 250Hz) and environmental variability. To account for reduced gaze estimation accuracy, larger Areas of Interest (AoIs) must be used. Technical issues such as differences in devices, participant movement, and inconsistent calibration limited the accuracy of the data. However, the ease of use and wide availability of WebGazer.js make it a useful tool for expanding access to cartographic research, especially in large online studies. This study suggests several ways to improve webcam-based eye tracking including better calibration and real-time feedback. These practices can help improve how maps are used and support future research in more natural, everyday settings.

List of Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA: Analysis of Variance

AoI: Area of Interest

ET: Eye Tracking

MED: Median

MS: Milliseconds

S: Second

SD: Standard Deviation

RT: Response Time

N: Sample Size

Chapter 1:
General Introduction

1.1. Motivation

Among other processes, both in the sphere of arts and in terms of science, cartography has known a great deal of advancement ever since the onset of digital technology. More so than ever, researchers look for innovative methods that would allow for better cartographic design, ease of use, and user experience.

Additionally, the latest developments in deep learning have ramped up the progress of webcam-based eye tracking. This paves the way for more precise tracking of eye movements, unlike conventional eye tracking systems using infrared (IR) based cameras, which send and receive signals to the eyes. Webcam-based eye tracking relies on deep learning-based gaze estimation methods, face recognition, and other computer vision techniques. Though promising, it has yet to address the lower precision against IR systems and many more difficulties involving the performance outcome, such as user movement and camera quality.

Eye tracking is generally understood to be the process that provides interesting information about people's focus of vision and their cognitive processes: how and where a person is looking, and for how long. Webcam-based eye tracking on the other hand, represents the use of regular optical cameras, present, for example, in personal devices such as laptops and even smartphones, making use of specialized hardware.

The distinction between these two technologies should be highlighted: where both try to obtain eye movement information, webcam tracking lacks the accuracy of, reliability, and data quality such that it is not as useful for such applications that require a lot of precise measuring, like reading a map.

The primary motivation for this thesis is to investigate whether webcam-based eye tracking can replicate the results of traditional lab-based eye tracking studies, specifically in the context of map usability and user experience. The main research question is

Can webcam-based eye tracking effectively replicate and scale the results of lab-based eye tracking studies in the context of map reading?

This particular question has not so far been extensively discussed in literature, specifically on map reading studies. eye tracking methodologies offer valuable insights into map reading by objectively revealing where users focus their attention and how long they fixate on specific areas, while also uncovering patterns of attention and strategies for processing information.

This exploratory research tries to establish whether webcam-based systems are able to produce any similar level of data and insight into that of professional eye tracking systems. Since human computer interactions and tasks require a fair amount of running back and forth between desktop and handheld maps eye tracking methodologies do enable data gathering in natural environments such as the home and office. However, its usefulness in understanding real-world behaviors with respect to maps is not quite certain, with a host of limitations that the authors will elaborate on. Challenges such as varying camera quality, user movement, and calibration accuracy can affect data reliability and validity. Addressing these issues becomes paramount in effectively leveraging the technology for the betterment of map design and user experience. Further, webcam-based eye tracking is an economical and user-friendly alternative to advanced systems that require specialized hardware. With a webcam readily available on any computer and mobile device, this tech aims at democratizing cartographic research and enabling researchers with limited resources to conduct usability and user experience studies. However, webcam-based eye tracking faces its own set of challenges. Indeed, variability in camera quality, user movement, and calibration accuracy can contribute serious flaws to reliability when it comes to the data.

In addition to those points, eye- tracking data in connection with map reading should be inferred by quite a fair knowledge of interpreting both cartographic principles and cognitive processes. This thesis will analyze the use cases and limitations of webcam-based eye tracking as a cartography research tool by introducing open-source software. It seeks to provide a comprehensive overview of the technology's potential, as well as the challenges involved in improving cartographic design and user interaction.

1.2. Research objectives

The main objective of this thesis is to explore the potential and limitations of using webcam-based eye tracking in cartography research, focusing on its ability to replicate traditional lab-based studies and enhance our understanding of how users interact with maps. Specifically, this research aims to investigate how webcam-based eye tracking can be applied to study map usability and user experience in more naturalistic settings, as well as assess the technical challenges involved. This includes identifying and reviewing a set of free software solutions for webcam-based eye tracking and then analyzing how effective they are as compared to traditional lab-based systems.

Another important goal is to evaluate the scalability of webcam-based eye tracking for online studies, in examining how novice users recall and employ visual information from digital maps. This study will compare usability metrics both quantitative and qualitative collected through lab-based and webcam-based methods, and will explore the associated technical and methodological challenges. In the final step, this thesis will propose improvements or best practices for optimizing map design based on the findings. These objectives lead to several research questions that will be addressed throughout the thesis.

Research objective 1:

Evaluating webcam eye tracking software solutions.

RQ 1: What are the best free webcam eye tracking tools for map reading?
RQ 2: How accurate and practical are these tools for map reading tasks?

Research objective 2:

Comparing webcam-based eye tracking for online studies and lab-based eye trackers.

RQ 3: How do user performance and data quality differ between the two methods?
RQ 4: Can webcam-based eye tracking replicate lab-based results, especially for online studies, particularly in evaluating novice users' interactions with 2D static maps?

Research objective 3:

Identifying limitations, challenges, and developing best practices.

RQ 5: What technical and methodological challenges arise when using webcam-based eye tracking how do factors like environmental variables and device differences affect data quality?
RQ 6: What best practices and recommendations can improve to best design a map reading experiment with webcam eye tracking?

1.3. Structure of the thesis

This thesis consists of six different chapters, each offering in depth information on the research into webcam-based eye tracking for cartography studies.

The following chapter reviews existing literature, exploring key studies and concepts in cartography eye tracking, and human computer interaction, comparing lab-based and webcam-based approaches in map related research and identifying their current shortcomings.

The third chapter outlines the research design and methodology used in this study, including how free webcam-based eye tracking software was chosen, and how online experiments were designed, detailing data collection, participant selection, and analysis of usability and eye tracking results.

The discussion chapter presents the experimental results, focusing on usability aspects like task duration, error rates, and user interactions, supported by statistical analysis, and addresses challenges such as error visualization in webcam-based tracking.

The subsequent section discusses the findings, comparing webcam-based tracking to lab systems, highlighting its potential for cartography research despite precision limitations, and reflecting on methodological constraints, like limited task diversity and varying user environments, while considering broader implications for the field.

The final chapter revisits the research questions, summarizes the comparison between tracking methods, acknowledges the promise and drawbacks of webcam-based systems, and proposes future research directions to tackle challenges and enhance its application.

Figure 1.1 shows the detailed structure of this master's thesis for further clarity.

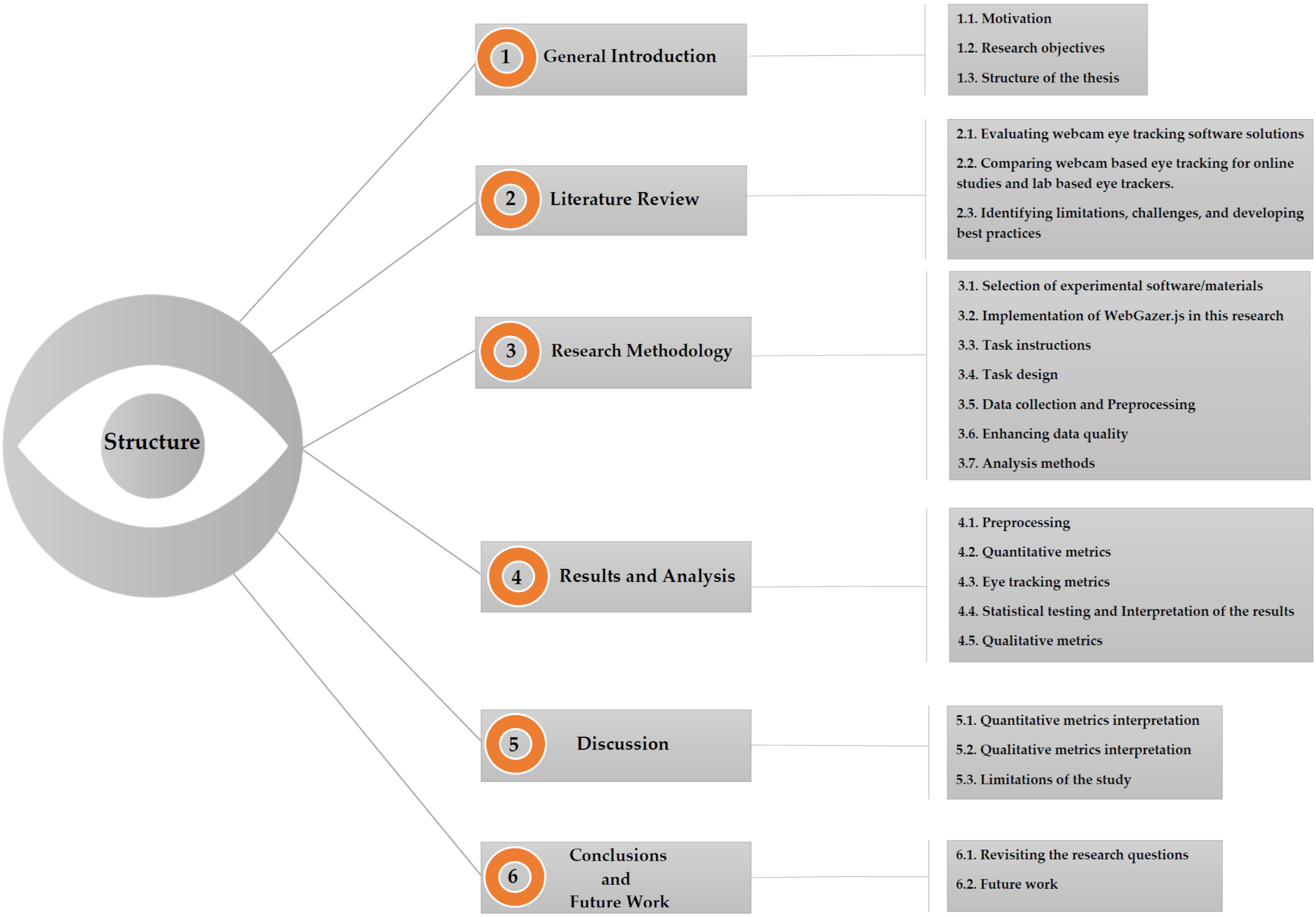


Figure 1.1. Structure of the master thesis

Chapter 2:

Literature Review

Over the last decade, eye movement analysis has significantly advanced our understanding of visual attention processes in various types of maps. Given the growing need to examine user interactions with realistic cartographic products such as static, animated, interactive, and multimedia maps a substantial body of experimental research has been conducted to explore different aspects of the map reading process. Eye tracking has become a key method in spatial cognition, geographic information science (GIScience), and cartography, offering valuable insights into how users engage with maps and geographic data Kiefer et al. (2017). The emergence of webcam-based eye tracking now provides researchers with cost effective and scalable tools, enabling large-scale online studies that are particularly beneficial for cartographic research. This literature review examines both traditional laboratory based and modern webcam-based eye tracking technologies, with a focus on free software solutions, their effectiveness, scalability, and limitations in the context of cartographic studies.

2.1. Evaluating webcam eye tracking software solutions

In recent years, free webcam-based eye tracking software has emerged as a compelling alternative to traditional lab-based systems, particularly for cartographic and behavioral research. Tools such as WebGazer.js, GazeRecorder, and RealEye have gained traction due to their accessibility and cost effectiveness. Among these, WebGazer.js stands out for its open-source framework and scalability, making it ideal for large-scale studies. This JavaScript library leverages standard webcams to estimate gaze in real-time, with self-calibration based on user interactions and seamless web page integration Papoutsaki et al. (2016). These features make it especially valuable in cartographic research, where it enables the evaluation of user interaction and comprehension without the need for specialized hardware. By contrast, GazeRecorder and RealEye offer greater precision in gaze tracking, which suits them to detailed analyses, such as fine-grained cartographic studies. However, their higher technical demands and associated costs limit their scalability for broad online research. For studies prioritizing accuracy over participant reach, these tools may be preferable, though they lack the flexibility of WebGazer.js. Despite its lower accuracy compared to commercial infrared systems, WebGazer.js remains a powerful tool for large-scale research. (Semmelmann and Weigelt, 2018) highlight its value in online studies, noting that while webcam-based systems may not rival lab grade precision, their accessibility democratizes research.

As technology evolves, the accuracy and reliability of these tools are expected to improve, expanding their applications in analyzing user interaction and spatial cognition within cartographic contexts. The effectiveness of eye tracking depends on a few key factors: accuracy (how well it tracks where someone is actually looking), precision (how consistent the results are over time), temporal resolution (how often it records gaze data), and scalability (how easily it can handle many participants).

Papoutsaki et al. (2016) describe WebGazer.js as a tool that uses regular webcams to track eye movements of website users in real-time, noting its usefulness and good performance. Recent studies highlight the differences between webcam-based and lab-based infrared systems. Wisiecka et al. (2022) point out that webcam tools have more errors but still give results that match what researchers expect, especially in large-scale studies about behavior and maps. Yang et al. (2021) show that with proper setup, webcam systems like WebGazer.js can be accurate enough to study how people interact with maps online, linking gaze patterns to decision making. On the other hand, traditional infrared trackers are better at precision, have fewer errors, and capture data quickly, making them perfect for detailed, controlled studies Myrte Vos et al. (2022). However, they are expensive and harder to use for big online projects. For advanced studies on map usability, webcam-based tools like WebGazer.js are a practical option. They may not be as reliable as infrared systems, but their affordability and ability to work with many people make them popular Wisiecka et al. (2022). This balance makes them a helpful choice for researchers studying how different groups read maps. The webcam tools like WebGazer.js, GazeRecorder,and RealEye each have strengths for map reading research. WebGazer.js stands out for being easy to use with large groups, making it great for broad online studies, while GazeRecorder and RealEye offer better precision for specific tasks. Their usefulness depends on the research goals, big picture usability studies benefit from scalable tools, while detailed analysis needs ones that are more precise. As these webcam, tools improve; they could close the gap with lab-based systems, providing a flexible, low cost option for cartographic research.

2.2. Comparing webcam-based eye tracking for online studies and lab-based eye trackers

Webcam-based eye tracking has transformed online cartographic research by offering a scalable, cost effective alternative to lab-based systems. Tools like WebGazer.js enable large-scale studies with diverse participants, a key advantage over the controlled but limited settings of infrared trackers (Krassanakis & Cybulski, 2021).

These systems support research into cognitive and perceptual processes in map reading, refining map design and geographic data representation through broad data collection. Lab-based infrared systems, such as the SMI RED250, provide superior accuracy and precision due to their controlled environments and advanced hardware. Wisiecka et al. (2022) report low error rates (e.g., <0.5° visual angle) and high reliability, ideal for detailed gaze analysis over complex map features.
In contrast, webcam-based tools like WebGazer.js exhibit higher error rates (e.g., 1–2° visual angle) due to variable lighting, camera quality, and participant movement Papoutsaki et al. (2016). However, Hassoumi et al. (2019) note that improved calibration techniques are narrowing this gap, making webcam systems increasingly viable for less precision critical tasks. Lab-based systems excel in capturing fine-grained gaze patterns, revealing how user expertise influences performance. For example, Keskin et al. (2023) used the SMI RED250 to investigate how expert and novice map users focus on features such as hydrographic areas and road junctions. Their findings demonstrate that factors like map design, task difficulty, and user expertise significantly influence visual attention and usability outcomes. Specifically, polygonal features like water bodies were easy to learn and remember across different user groups. Webcam-based tracking, while less precise, supports broader usability studies.
(Yang and Krajbich, 2023) demonstrate its effectiveness in naturalistic settings, though data quality suffers from latency and hardware inconsistencies Robal et al. (2018). Novice users may perform similarly across systems for simple tasks, but complex map reading reveals webcam limitations in consistency. Webcam-based systems can partially replicate lab-based insights, particularly for high-level usability trends in online studies. (Krassanakis & Cybulski, 2021) argue that they capture general perceptual patterns in map reading, such as attention to salient features, aligning with lab findings. For novice users interacting with 2D static maps, webcam tracking identifies broad gaze behaviors (e.g., focus on water bodies), but it struggles with the precision needed for detailed cognitive analysis Myrte Vos et al. (2022). While not a full substitute, advancements in tools like WebGazer.js suggest growing potential to approximate lab results in scalable, real-world contexts. Webcam-based eye tracking offers unmatched scalability and accessibility for online cartographic studies, though it lags behind lab-based systems in accuracy and data quality. Lab trackers remain more accurate, but webcam tools are improving. User performance varies by task complexity, with lab systems excelling in controlled settings and webcam systems enabling naturalistic insights despite lower data quality.

Webcam tracking can replicate general usability findings for novice users with 2D static maps but falls short of lab precision for detailed replication. As technology advances, webcam systems may bridge these gaps, balancing scalability and reliability for future research.

2.3. Identifying limitations, challenges, and developing best practices

As highlighted by (Krassanakis & Cybulski, 2021), variability in device quality, particularly in webcams, presents a significant challenge for eye tracking. The accuracy of gaze tracking is closely tied to the webcam's quality, with differences in resolution, frame rate, and lens quality across devices potentially leading to inconsistent data collection. Environmental factors such as lighting and participant movement further impact accuracy (Semmelmann1 & Weigelt, 2017). For instance, inconsistent lighting can disrupt pupil detection algorithms, resulting in inaccurate gaze estimates. Moreover, even minor head movements can alter perceived gaze direction, especially since many webcam-based eye trackers, including popular ones like WebGazer.js, lack effective head movement compensation (Papoutsaki et al, 2016). These factors can degrade data quality by increasing noise and reducing the precision of fixation and saccade measurements. Methodologically, calibration poses a challenge in uncontrolled settings. Accurate calibration requires participants to fixate on specific points, but distractions, fatigue, or unclear instructions often lead to misalignment, undermining data reliability (Krassanakis & Cybulski, 2021). Additionally, webcam systems' lower sampling rates (e.g., 15–30 Hz compared to 250 Hz in infrared trackers) limit temporal resolution, missing rapid eye movements critical for map reading tasks like visual search Yang et al. (2021). Other factors, such as participant diversity (e.g., glasses, age-related vision differences) and internet latency in online studies, further introduce variability and delays, affecting the consistency and usability of collected data Myrte Vos et al. (2022). Given the limitations and potential of webcam-based eye tracking , several best practices can enhance its effectiveness in cartographic research. First, providing clear calibration guidelines and instructions for participants is essential to improve data quality. Accurate calibration is particularly crucial in webcam-based systems, which are sensitive to variations in devices and environmental conditions Semmelmann et al. (2017). Explicit and thorough calibration processes, akin to those used in controlled lab settings, are vital for minimizing errors and ensuring reliable data collection (Krassanakis & Cybulski, 2021).

However, webcam-based systems are more susceptible to variations in lighting, participant positioning, and device differences compared to specialized lab-based eye trackers (Semmelmann et al, 2017, Yang & Krajbich, 2021).

Second, researchers can optimize study design by concentrating on key map features that are known to attract attention, such as landmarks and hydrographic areas, as demonstrated by Keskin et al. (2023). Incorporating these attention-grabbing features can make it easier to analyze user interaction with maps.

Additionally, integrating peripheral vision analysis and refining feedback loops during calibration can provide deeper insights into the cognitive processes involved in map reading and navigation Kiefer et al. (2017) (Yang & Krajbich, 2021).

To account for the lower spatial resolution of webcam-based systems, it is recommended to use larger Areas of Interest (AoIs) around key features. This ensures that fixations on important elements arecaptured accurately, even in the presence of minor gaze estimation errors (Kiefer et al, 2017, Keskin et al, 2023).

Eye tracking also offers valuable insights into map design. It reveals how users interact with maps, identifying which symbols are most easily recognized, how color choices influence attention, and how layout affects search patterns. For instance, eye-tracking studies have been used to understand how users perceive and remember specific map features, such as landmarks (Krassanakis & Cybulski, 2021). These findings allow cartographers to refine map features to enhance information retrieval and spatial comprehension for both novice and expert users Keskin et al. (2023).

Chapter 3:
Research Methodology

The growing use of webcam-based eye tracking technologies has opened new avenues for studying user interactions across various fields, including cartography. Eye tracking provides valuable insights into how users interpret maps, focus on geographic features, and engage with cartographic designs. Traditionally eye tracking studies relied on expensive, lab-based equipment, which limited scalability and participant diversity. However, the development of webcam-based eye tracking solutions has made it possible to conduct remote, large-scale studies, enabling researchers to gather gaze data from users in real-world environments. Webcam based eye tracking tools have been validated in numerous fields, with research showing that gaze data gathered remotely can be comparable in quality to data collected in controlled laboratory environments.

However, the specific application of these technologies to cartography remains underexplored. By employing open-source tools such as WebGazer.js, this study seeks to assess the feasibility of using webcam-based eye tracking to analyze how users engage with maps in online settings. This approach facilitates the exploration of web based cartographic applications on a larger scale, providing valuable insights into user behavior and interaction patterns. In particular, WebGazer.js reflects the goal of optimizing for scalability, accessibility, and ease of integration into online environments, making it an ideal choice for research into web-based map interfaces. This study outlines the experimental software, rationale for tool selection, and the methodological steps involved in implementing WebGazer.js for data collection in the context of cartographic studies. By examining how participants visually engage with maps, this research contributes to a better understanding of user behavior and informs the design of more intuitive and effective cartographic systems. Ultimately, this study evaluates how well these technologies capture eye movements during map interpretation tasks, offering a new methodological approach for cartographic usability studies.

3.1. Selection of experimental software/materials

This thesis evaluates three different software solutions for webcam-based eye tracking: WebGazer.js, GazeRecorder and RealEye . To choose the software; a pilot test was conducted for each to determine which one corresponded best to the study. Each of these tools offers different strengths and weaknesses, which are critical to consider in cartography research.

-WebGazer.js: WebGazer.js is a JavaScript library, an open-source, browser based tool and easily incorporated into most websites. It is designed for real-time, eye tracking using standard webcams. Its accessibility and ease of use make it suitable for large-scale online studies. However, its accuracy can be influenced by environmental variables like lighting and participant movement, which may affect its reliability in detailed cartographic analysis

-GazeRecorder: GazeRecorder offers a user-friendly interface for webcam, eye tracking and records gaze data that can be analyzed offline. While it provides a higher level of accuracy compared to WebGazer.js, it requires software installation, making it less scalable for remote or online studies.

-RealEye: RealEye is a commercial solution that uses cloud-based webcam eye tracking, allowing researchers to collect data without the need for software installation. This tool provides a balance between ease of use and data accuracy, but it is subscription model could limit access for researchers with limited budgets. After comparing these tools, WebGazer.js was selected for this thesis due to its flexibility for large-scale online cartography studies, aligning with the objective of evaluating scalable webcam based eye tracking solutions. To enhance the clarity of the experimental design, this study employs a mixed factorial design matrix similar to Keskin's (2020) lab-based study. Specifically, we recreate three tasks of that study, which involves road networks and hydrographical features, to aid in visually organizing the components of the research methodology. The Table 3.1. Vaníček et al. (submitted) provides a comparative perspective on WebGazer.js and RealEye,which aligns with the objectives of this study.

Table 3.1.Preliminary comparison of evaluation results for the three (modified from Vaníček et al., 2025).

	◆ WebGazer (static map exp., open-source)	• GazeRecorder (desktop-based, open-source)	• RealEye (interactive map exp., com.)
Initial Setup	Self-hosted web solution (requires programming skills).	Desktop application; requires installation on user's machine, some setup for webcam integration.	Hosted web platform; quick setup, minimal coding needed for basic use.
Calibration or Gaze Accuracy	No built-in calibration; accuracy depends on proper implementation and external factors like lighting, webcam quality, and user positioning. Precise calibration is essential to reduce errors.	Calibration required; accuracy depends on webcam quality and user environment, similar to WebGazer.	Built-in calibration module; still influenced by user environment factors (lighting, webcam position, etc.).
Stimuli & Exper. Design Flex.	Best for static maps with high customization, but requires significant effort for interactive tasks and may face accuracy challenges.	Flexible for static and interactive stimuli; desktop nature allows broader experiment design.	Suited for static maps with easier setup; interactive tasks possible but limited in flexibility.
Data Quality & Particip. Excl. Rates	High-quality data with careful setup, but may have higher exclusions due to calibration issues or poor webcam quality. Managing large data files can be challenging.	Good data quality with proper setup; exclusions possible due to webcam or environment variability.	Generally reliable data with fewer exclusions; however, accuracy may decline for low-quality webcams or poor environments.
Ease of Use	Requires intermediate to advanced technical skills for custom integrations and data handling.	Moderate ease; requires basic technical setup but simpler than WebGazer for non-programmers.	User-friendly web interface; lower technical barrier for most standard tasks.
Export Options	Raw data export supports detailed processing but requires custom scripts or tools like Python (Matplotlib, Seaborn), Excel, Tableau, or R for visualisations and reports.	Exports raw data (e.g., CSV); visualization requires external tools like Python or Excel.	Offers built-in export (in CSV format). In addition, there is a possibility of exporting output videos with heatmap visualisations.
Additional Metrics Logging	It can be implemented via custom scripts; higher programming effort.	Custom metrics possible with script integration; moderate programming effort.	It can be implemented via custom scripts for own web pages; programming skills are needed.
Cost and Scalability	No license fees; you must consider development resources and potential server costs.	Free (open-source); scalability limited by local machine resources, no subscription costs.	Subscription-based model; fees vary by plan but provide swift scalability and dedicated support.

	WebGazer (static map exp., open-source)	GazeRecorder (desktop-based, open-source)	RealEye (interactive map exp., com.)
Technical Support	Community support via forums, documentation, and GitHub, with options for self-troubleshooting or third party help.	Community-driven support (forums, GitHub); no formal dedicated support.	Dedicated customer support, faster response times and personalised help for higher-tier plans. Regular updates are included.
Participant Recruitment Integration	No built-in recruitment tools; relies on external platforms or manual methods like email, social media, or physical notices.	No built-in recruitment; relies on external methods (e.g., email, social media).	Offers participant recruitment tools and integration with higher plans, simplifying participant sourcing and management.
Updates and Maintenance	Relies on the OS community for updates; users manage updates and fixes manually.	Open-source community updates; manual maintenance required by user.	Maintained by Real Eye with automatic updates included in the subscription.
Data Security and Privacy	Data security depends on hosting; users ensure GDPR compliance.	Local storage on user's machine; security depends on user's setup and compliance efforts.	Managed platform with GDPR-compliant data security.
Performance Issues	Prone to interruptions or delays in complex self-hosted setups.	Dependent on local hardware; potential lag with low-spec machines or high data volumes.	Rare performance issues due to the supported black box environment.
Best Suited for…	Developers, researchers, or teams that need maximum flexibility and are comfortable with programming and self-hosting.	Researchers or small teams with desktop preference, moderate technical skills, and cost constraints	Non-technical users or teams needing quick, reliable hosted solutions, but with a subscription fee.

3.1.1. Reason for choosing the software

WebGazer.js was chosen as the primary software for this thesis due to several factors that align with the research objectives. It is an open-source, browser based tool that tracks eye movements in real-time using standard webcams. This makes it easy to modify and use in any web based setting, which is ideal for large-scale remote online studies, such as testing how usable maps are in real-world scenarios. Unlike other tools such as GazeRecorder or RealEye, WebGazer.js does not require participants to install software. This feature facilitates participant recruitment and supports efficient remote data collection, making it well suited for large-scale online studies. Participants can join the study using their own devices without any need for downloads or setup.

Another key benefit is that WebGazer.js is free since it is open-source, making it perfect for researchers with limited budgets. In contrast, RealEye offers slightly better accuracy but comes with subscription costs that can add up for long term or large projects. GazeRecorder might also be a bit more accurate and allows offline data analysis, but it requires installation, which limits its use for remote studies.

WebGazer.js strikes a good balance between accuracy and ease of use, making it the best choice for this research, which focuses on collecting real-time data for online map studies on a large scale. While its accuracy can be affected by things like lighting or movement, it still provides solid tracking for this type of online experiment. When considering accessibility, cost, and flexibility, WebGazer.js stands out as the right tool for evaluating how users interact with and interpret maps in real-life situations. WebGazer.js tracks where a person looks on the screen by analyzing their eye movements. It depends on outside tools to spot facial features, especially the eyes, and figure out their positions. To locate the pupil (the black center of the eye), WebGazer.js assumes the iris (the colored part) is darker than the white area around it, called the sclera, and that the pupil is in the middle of the iris. Turning the pupil's position into a spot on the screen is complicated. It involves complex math with many factors, depending on how the user's head is tilted or turned compared to the camera and screen. To match eye movements to specific screen locations, WebGazer.js uses a method called Ridge Regression. This takes a detailed 120-part description of the eye's features and links it to X and Y coordinates on the screen each time the user clicks to set it up. As the user keeps interacting with the system, WebGazer.js gets better at guessing where they are looking, improving its accuracy over time. This ability to refine itself while being easy to use and accessible makes WebGazer.js a strong fit for this research.

3.2. Implementation of WebGazer.js in this research

Before implementing the use of WebGazer.js, it is crucial to address and mitigate privacy concerns associated with webcam use for gaze monitoring. Participants should be informed of the following key points to ensure their comfort and privacy:

Local data processing: WebGazer.js processes all webcam images locally on the participant's device. No webcam footage is transmitted or stored beyond the participant's own device.

Data transmission: The only information transmitted from the participant's device consists of the output coordinates generated by the WebGazer.js algorithm. This data includes the horizontal (x) and vertical (y) positions, along with the corresponding time (t), which collectively represent the estimated point of gaze at each moment.

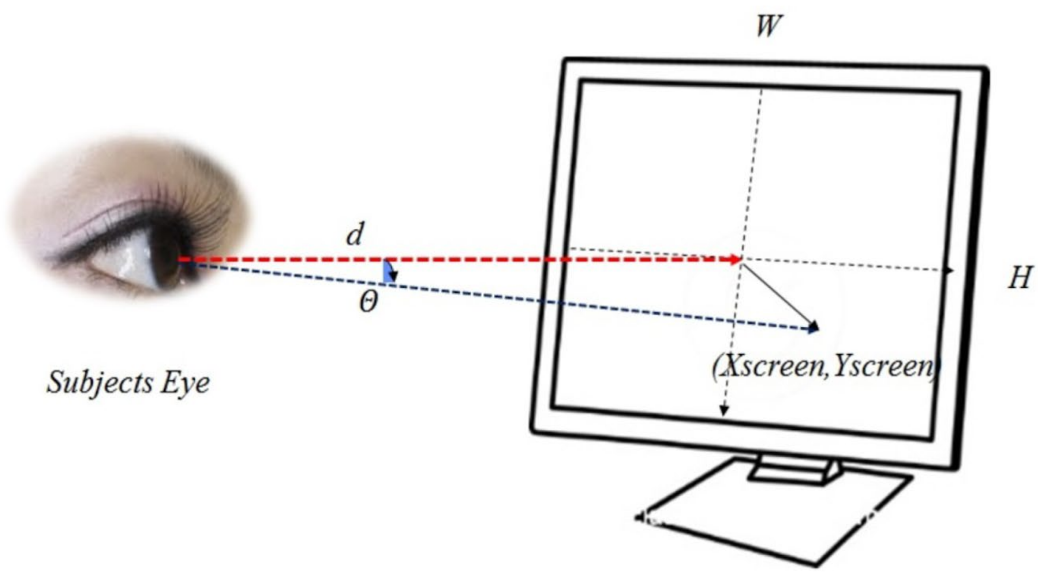


Figure 3.1: The arrangement for gaze recording.

Live video feed: In this study, participants saw themselves live on screen prior to the calibration procedure, accompanied by calibration guidance. This feature helped them position their heads correctly. For privacy-sensitive participants, the live video feed could be disabled by turning off the "show video" option during, eye tracking on surveys, and the face mesh was turned off to minimize user distraction. However, the "show face overlay" and "show face feedback box" options remained enabled to facilitate accurate calibration. While this adjustment may have slightly affected calibration accuracy, it effectively addressed privacy concerns at the start of the experiment.

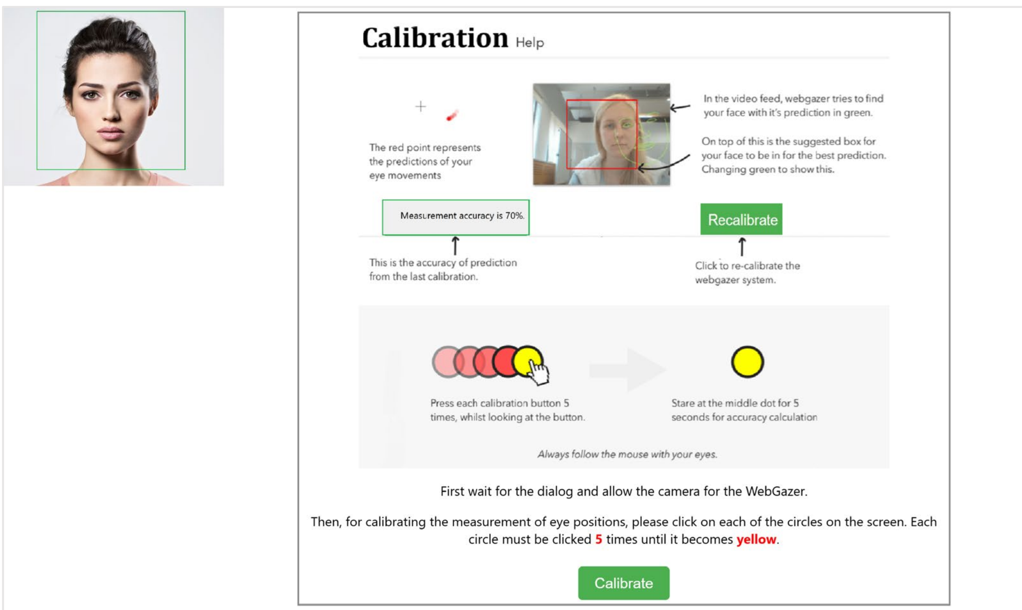


Figure 3.2.Calibration assistance screen.

Calibration steps. Participants completed an initial calibration before starting the task. Instructions were provided to guide them through the calibration process. Once participants were correctly seated, with the webcam positioned appropriately and the experiment screen selected, they clicked the "Calibrate" button to begin. A video feed appeared in the top left corner of the screen. Participants used this feed to adjust their position and center their face within the green box displayed on the screen (Figure 3.3).

- Participants clicked each of the nine calibration points five times, until they turned yellow.
- In the next step, participants were instructed not to move the mouse and to focus on the center dot for 5 seconds. This step allowed for calculating the accuracy of predictions and validated the calibration process.
- If the calibration results were unsatisfactory, participants could restart the calibration process by pressing the “Recalibrate” button.
- Calibration was considered successful if measurement accuracy exceeded 50%. Once calibrated, participants proceeded to the training task to familiarize themselves with the main task.

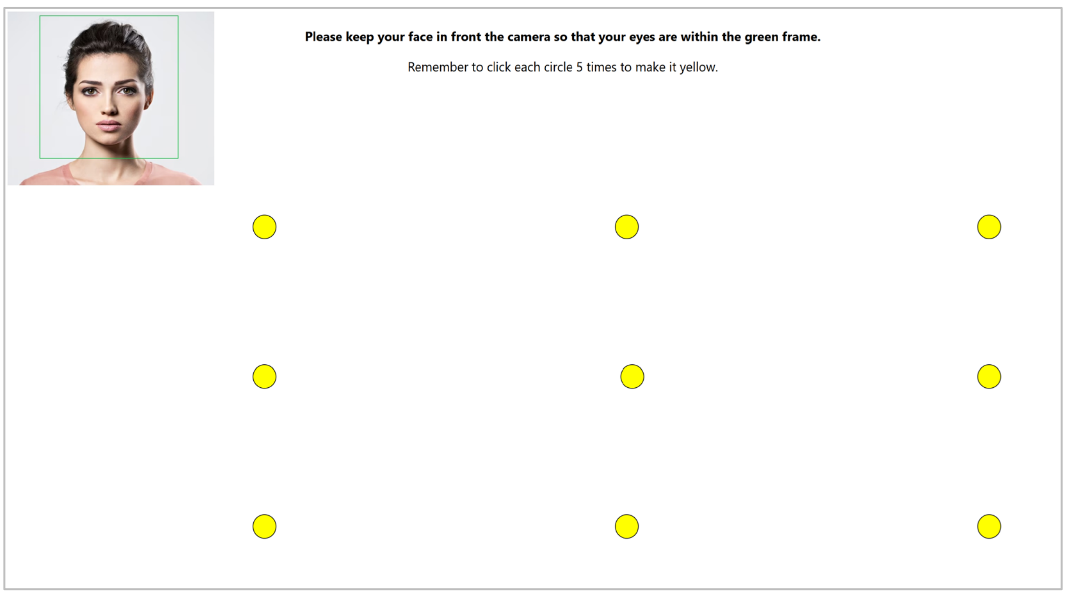


Figure 3.3. Calibration process steps for accurate gaze tracking.

Process of calibration in WebGazer. WebGazer.js does not natively provide calibration data in degrees of visual angle, which is the standard unit used in eye tracking research. Instead, it uses self-calibration (a process where the system observes how web visitors interact with the webpage and learns to map eye features to specific positions on the screen).The system outputs gaze estimates as pixel coordinates relative to the screen. The calibration accuracy is quantified as a percentage, calculated through the following process:

This process begins by enabling WebGazer.js to store gaze prediction points for 5 seconds, allowing it to collect sample gaze data as the participant interacts with various calibration points. After this interval, WebGazer.js retrieves the stored gaze points, which include the $x_{i,\ pred}$ and $y_{i,pred}$ are the *i*th predicted gaze coordinates, and *width* and *height* denote the display dimensions.
Note that d_i is chosen so that its value lies between 0 and 1 when the fixation point coordinates are equal to the central point of the display; that is, x_{true} = width / 2, y_{true} = height / 2.

$$d_i = \sqrt{\frac{1}{2}\left(\left(\frac{x_{i,pred} - x_{true}}{width/2}\right)^2 + \left(\frac{y_{i,pred} - y_{true}}{height/2}\right)^2\right)}$$

The final value of d is obtained by averaging over 50 predicted points. This averaged normalized distance is essentially the statistical error of the prediction, obtained empirically. The precision is then computed accordingly as:

$$\text{Precision} = 1 - \overline{d}$$

and reported as a percentage.
The system then evaluates whether the calculated precision meets the required standard. If accuracy is below the predefined threshold, the participant is prompted to recalibrate. If the accuracy is sufficient, the calibration process is concluded. In cases where recalibration is necessary, WebGazer.js clears the previous data and restarts the calibration process to improve accuracy.

3.3. Task instructions

After participants were asked to sign an informed consent form and the eye tracker was accurately calibrated, they proceeded through a series of trials. Each trial consisted of two stages: encoding (learning) (Figure 3.4) and decoding (recognition) (Figure 3.5).

Part I: Map study (encoding stage):

In the first part, the user was presented with a map on the screen and was asked to observe its general structure. The goal was not to memorize every detail but to focus on key structural elements, such as recognizing roads and bodies of water. The user had 7 seconds to study the map. Encoding refers to the process of transferring sensory input into memory, transforming information from working memory into long-term knowledge. Once the user had reviewed the map, it automatically disappeared after 7 seconds, and they proceeded to the second part of the task (Figure 3.4).

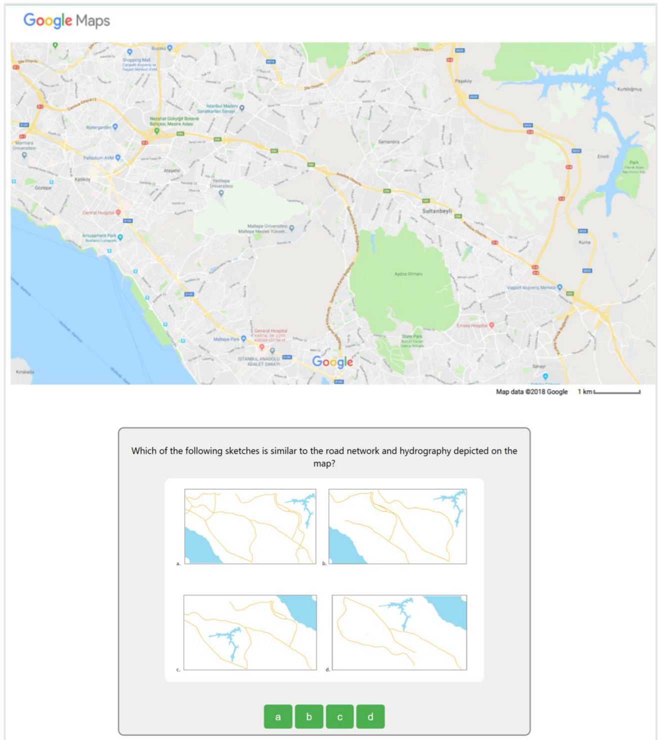


Figure 3.4.Task structure of experiment (encoding stage).

Part II: Multiple-choice graphical answer screen (decoding stage):

After the map disappeared, a survey screen appeared where the user was asked questions about the structural elements of the map previously shown. Users were required to select the correct answer by clicking the corresponding letter (e.g., a, b, c, d) that matches the correct option. The decoding stage involved retrieving previously encoded information, requiring participants to recall stored details and bring them back into working memory (Figure 3.5).

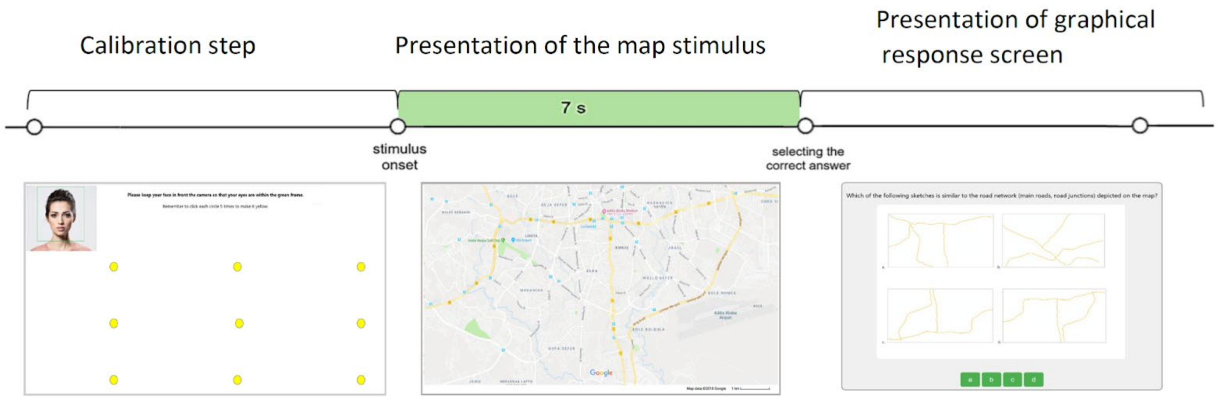


Figure 3.5. Experiment procedure (decoding stage).

3.4. Task design

Training task structure. To ensure higher data quality in this study, a training task preceded each main task. During the training, users were shown a map related to the upcoming main task and were given 7 seconds to study it. Afterward, a question about key map elements such as roads or hydrography appeared on the screen. Users had to select the correct answer by clicking on the corresponding letter (e.g., a, b, c, or d) as quickly as possible. Upon completing the training task, users given access to the correct, to help them become more familiar with the process of the main task. They then proceeded to the main task, where they answered similar questions based on the Subsequent map stimuli.

Main task structure. This study consists of three tasks. In each task, participants are asked to study a map stimulus under a free viewing condition for seven seconds. During the task period, participants are required to complete an experimental task. After viewing the map stimulus, four graphical response panels appear, each displaying skeleton maps that include only the main structuring elements of interest. Participants are then instructed to select the panel with the correct skeleton map corresponding to the stimulus they had just studied.

Tasks and stimuli. In this study, three memorability tasks were used, focusing on the recognition of different map landmarks, with two stimuli presented across the tasks. Each task contained 11 maps, with one map in each task dedicated to a training exercise. To streamline the study following the pilot phase, the number of stimuli was reduced. Each task featured different maps to avoid repetition, which also helped mitigate the learning effect.

The structure of the tasks is as follows:

Task 1: 11 Roads (roads and road junctions)

Task 2: 11 Hydrography (major lakes and rivers)

Task 3: 11 the combination of Roads and Hydrography

To ensure the study remained focused on the most attention grabbing elements, hydrography and road junctions were selected as the primary stimuli. This decision was based on prior research Keskin et al. (2023), which identified these features as the most visually salient in map-based tasks. Additionally, given the constraints of an online experiment, the study duration was limited to a maximum of 20 minutes. The original experiment lasted 2.5 hours, which was impractical in an online setting. To address this, only a portion of the original experiment was replicated, carefully shortened based on insights from previous research. This approach ensured meaningful results while maintaining participant engagement and data quality.

Figure 3.6. Example stimulus and experiment Task1: Roads, Task2: Hydrography, Task3: Roads and Hydrography.

Post-experiment questionnaire. At the end of all tasks, a brief Google form provided to collect basic demographic information, including

- Age
- Gender
- What is the highest level of education you have completed?
- What did you study/graduate in?
- Do you have a background in cartography?
- Please specify your profession.
- What kinds of maps do you mostly use?
- How confident are you about reading maps?
- How often do you use Google Maps?
- Do you think Google Maps is easy to use?
- What was your strategy to remember map landmarks?
- Any thoughts or comments about the experiment are welcome!
- Email (optional, for updates on results).

This step provides better insight into participants' backgrounds and how these may influence their performance in the tasks.

3.5. Data collection and preprocessing

Materials. All stimuli were selected from 2D static snapshots of Google road maps, sourced from the CartoGAZE dataset (https://doi.org/10.7910/DVN/ONIAZI)

Participants. A total of 35 participants initially took part in the study; however, only those who achieved a calibration accuracy threshold of 70% were selected, resulting in a final group of 28 participants. This group included a balanced mix of 14 female and 14 male participants, with 15 classified as novices and 13 as experts. The novice participants, with an average age of 35, had educational backgrounds ranging from undergraduate degrees to master's degrees across various fields, including engineering & technology, humanities & social sciences, health & medicine, and business & management. In contrast, the expert group, with an average age of 40, consisted of participants with at least a master's degree in geography or agricultural & environmental sciences. In this study, all participants were combined into one group instead of distinguishing between experts and novices. Figure 3.7 illustrates the participants' characteristics.

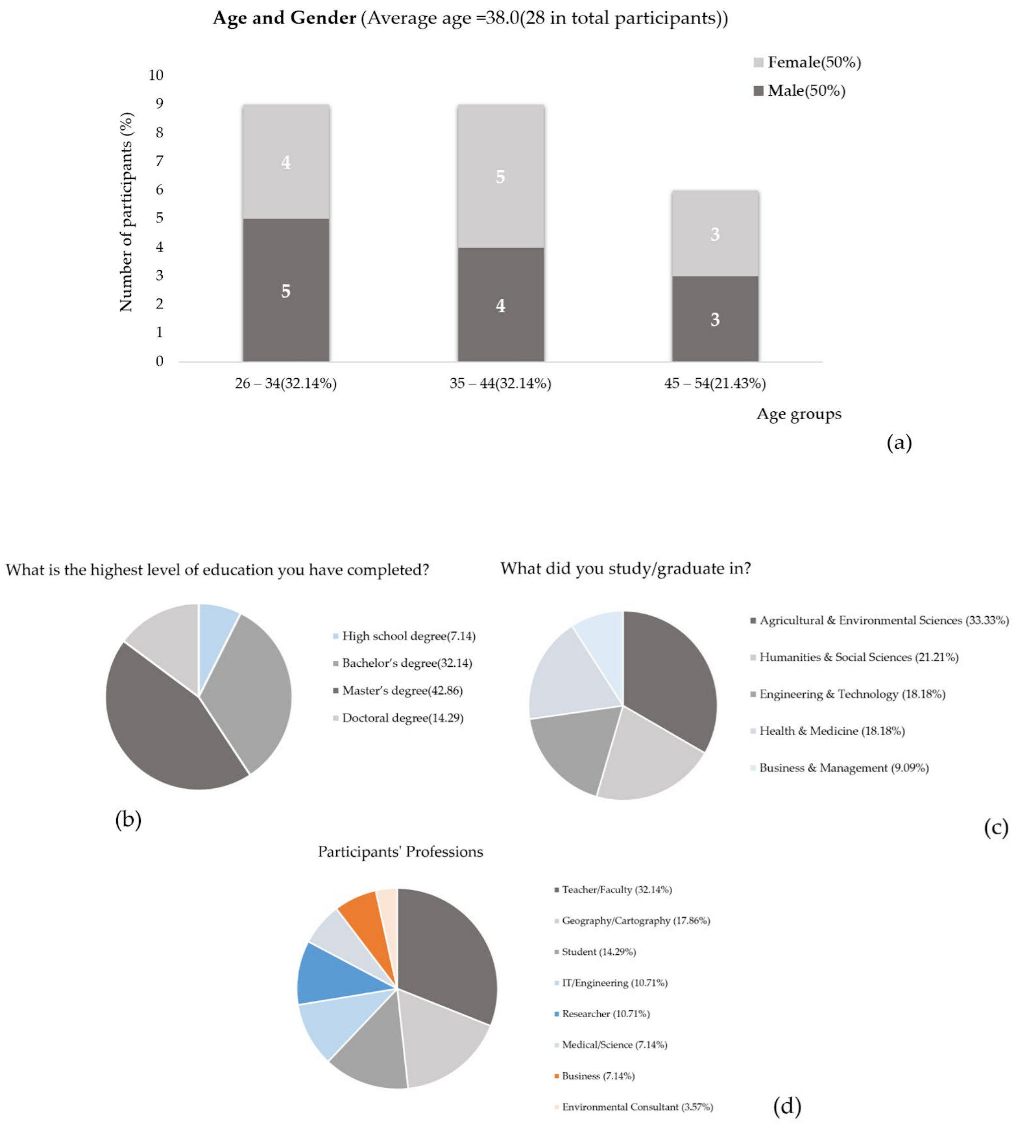


Figure 3.7. Participant's characteristics.

Apparatus and recording. This study was conducted using an online webcam-based tracking system to monitor participants' gaze behavior remotely. Participants received the experiment link via email and social media, along with a brief orientation script explaining the procedure. They used their own personal computers equipped with standard webcams. The eye tracking software, WebGazer.js was embedded within the web interface and recorded gaze data in real-time as participants interacted with the map stimuli. All participants were required to have a functional webcam and a stable internet connection.

The WebGazer.js tool estimated gaze position in real-time, tracking participants' gaze as they viewed the map stimuli and responded to the tasks. To ensure optimal tracking accuracy, participants first signed a consent form and then received the following instructions:

- Close any unnecessary programs or applications on their computer.
- Sit at an appropriate distance from the camera (50-60 cm), ensuring their eyes are clearly visible.
- Use only their eyes to look around the screen, minimizing head movement.
- Position lighting in front of them, rather than behind, so clearly capture the participant's face.

Structure and submission of raw data .To save the raw data effectively, it is first necessary to convert the data into a structured format before transmission. The system uses JSON (JavaScript Object Notation) for this purpose. JSON is a versatile, lightweight format that is human-readable while remaining easily parsed by machines.
It is commonly used in web-based applications because of its simplicity and compatibility with modern systems. The decision to use JSON ensures that the data can be transferred efficiently over the internet. As part of this step, the raw data is serialized into a JSON string. This serialized format is essential for ensuring that the data can be reliably transmitted via web-based protocols. To maintain transparency and accuracy, the prepared JSON data is logged to the system. This logging enables verification of the data structure before it is sent offering a clear view of what will be transmitted. Once the raw data is serialized, it needs to be sent to a cloud-based service for storage or further processing. The system employs Google Apps Script, a cloud platform that integrates with Google services and supports efficient data storage. The raw data is submitted to a predefined endpoint URL, which represents the destination server where the data will be processed and stored. To send the structured data to the cloud, the system relies on an HTTP POST request. This method is one of the most widely used approaches for securely transmitting data to a server. In addition to cloud submission, the system incorporates an optional feature for local storage of raw data. It also supports additional analysis or offline access to the data when needed. By employing both cloud and local storage, the system enhances data reliability, security and flexibility, to ensure that data is consistently available for future use or further processing. This organized way of handling raw data keeps the system strong, able to grow and safe from losing data.

3.6. Enhancing data quality

To enhance data quality during, eye tracking on maps two main validation methods are applied that guide user positioning and improve the reliability of collected data.

Eye feedback box validation. The eye feedback box provides real-time visual feedback based on the user's positioning within the camera frame (Figure 3.8). It checks the eye position every 3 seconds to confirm accurate alignment. The validation helps ensure only valid data is captured by providing color-coded borders indicating the eye tracking status:

- ***Green border***: eye is correctly positioned; data is valid.
- ***Red border***: eye is out of frame; prompts repositioning.
- ***Black border:*** Indicates a neutral or uninitialized state, typically used when eye detection has not started or is inactive.

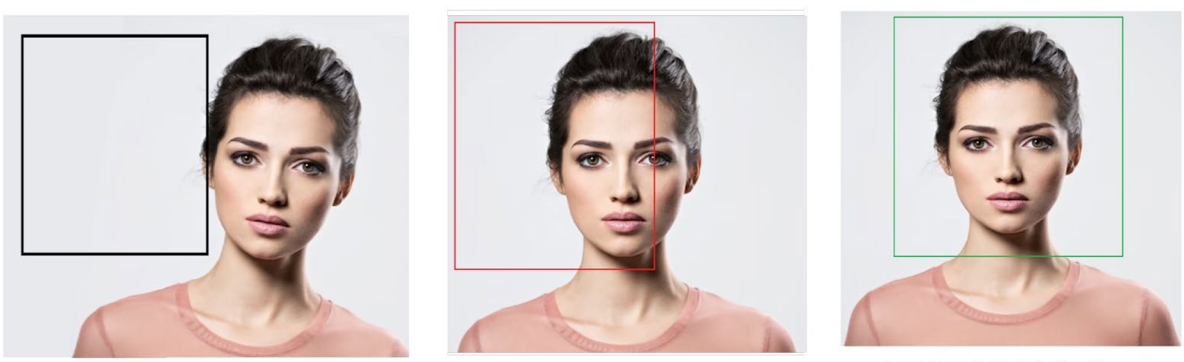


Figure 3.8 .Structure of eye position check within validation box.

Alert for out of frame detection. The alert function shows a clear warning if the user's face exits the camera frame, prompting repositioning for accurate eye tracking. This temporary alert message appears at the top of the screen and automatically fades after a few seconds once the user returns to the correct position (Figure 3.9).

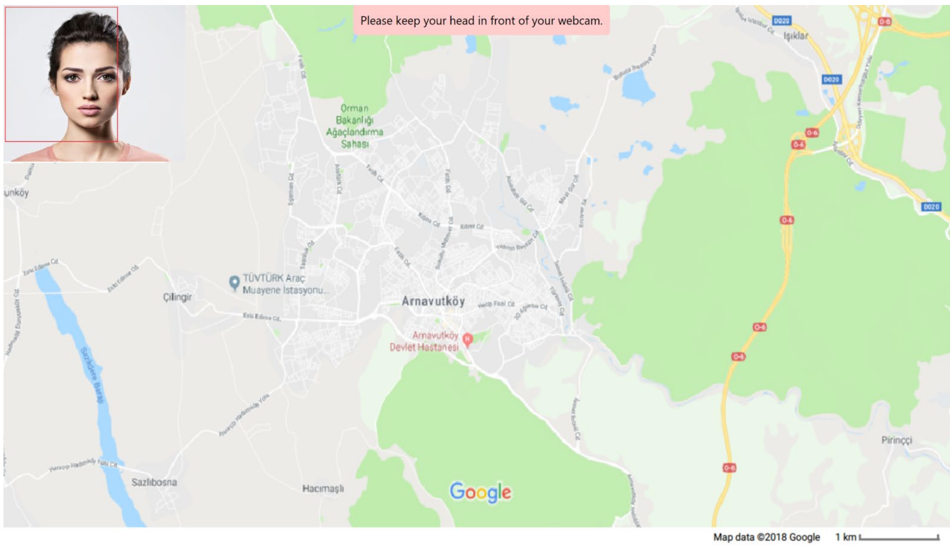


Figure 3.9. Alert for out of frame detection.

3.7. Analysis methods

With the aim of this thesis in mind, the study wanted to answer:

The main research question

Can we effectively replicate and scale lab-based, eye tracking studies online using webcam-based eye tracking technology for map reading?

To achieve this question, webcam-based eye tracking tools, specifically WebGazer.js were used to collect data. A carefully designed experiment was conducted to ensure meaningful and interpretable results, providing a comprehensive understanding of complex behavioral and cognitive processes. After a thorough inspection of the scan paths and considering the low quality of recorded gaze samples, seven participants were excluded from both behavioral and eye tracking data analysis. The final dataset used for analysis adhered to a calibration accuracy threshold of 70%, ensuring that only high quality data were included. The Metrics Used in the Study is that:

3.7.1. Quantitative metrics

-Behavioral measures

- ***Response time***: The duration elapsed from the presentation of a stimulus or task to initiation or completion of a participant's response. Typically measured in milliseconds or seconds, it reflects the speed of cognitive processing, decision-making, and motor execution. Response time varies with task complexity, participant expertise, and interface design, with shorter times indicating faster processing or higher efficiency, commonly used in cognitive psychology, human-computer interaction, and usability studies.

- ***Success rate (Accuracy)***: The proportion of tasks a participant completes correctly out of the total tasks attempted. It quantifies performance effectiveness, reflecting task comprehension and the usability of interfaces or systems. Widely used in usability testing and cognitive studies, higher success rates indicate better understanding and intuitive design, while lower rates may highlight difficulties or inefficiencies.

-Eye tracking metrics

- ***Fixation***: A brief period during which the eyes remain relatively still, focusing on a specific point or object in the visual field to process visual information. Fixations typically last 200–300 milliseconds, though durations can range from 100 to 500 milliseconds depending on the task, cognitive demands, and stimulus complexity. They occur between saccades (rapid eye movements) and are essential for visual perception, attention, and information processing, with neural control involving areas like the superior colliculus and visual cortex.

- ***Average fixation duration***: The mean time spent per fixation, on visual elements, typically ranges from 200–300 milliseconds but can vary. Longer fixation durations (e.g., 300–500 ms or more) often indicate increased cognitive processing, such as deeper analysis of complex stimuli, or difficulty in interpreting information, such as unfamiliar or ambiguous content. This is supported by eye tracking research, which links longer fixations to higher cognitive load or greater attentional engagement with visual elements.

- ***Number of fixations per second:*** The number of fixations per second refers to how often the eyes pause (fixate) to process visual information within a given time window. A higher fixation rate (e.g., 4–5 per second) often indicates rapid scanning behavior, such as skimming text or exploring a complex scene. Conversely, a lower fixation rate (e.g., 1–2 per second) typically suggests deeper cognitive processing of fewer visual elements, such as when analyzing detailed or complex stimuli. This relationship is supported by; eye tracking studies, which show fixation frequency varies with task demands, attention, and cognitive load.

- ***Saccades:*** Rapid ballistic eye movements between fixations, allowing the visual system to shift gaze from one object or location to another. These movements, typically lasting 20–100 milliseconds, are involuntary and occur several times per second to scan the visual environment, facilitating visual perception and attention. Saccades are characterized by high velocity (up to 900°/second), minimal latency (100–300 milliseconds), and precise coordination between the eyes, driven by neural circuits in the brainstem and cortex, particularly the superior colliculus and frontal eye fields.

- ***Average saccade length****:* The mean distance traveled by the eyes during a saccade, typically measured in degrees of visual angle. It ranges from 1° to 10° in natural viewing conditions, with smaller saccades (1°–2°) common in tasks like reading and larger saccades (5°–10°) occurring during visual search or scene exploration. Saccade length reflects visual attention allocation and task demands, influenced by stimulus layout and cognitive processing, and is controlled by neural mechanisms in the superior colliculus and frontal eye fields.

3.7.2. Qualitative Metrics

-Visual Analysis

- ***Scanpath visualizations (via Gazealytics software)***: A graphical representation of the sequence of fixations and saccades. Scanpaths illustrate the order and pattern of visual attention and are used to interpret viewing strategies, compare user behavior, and highlight areas of interest or confusion, with linear or erratic patterns reflecting task efficiency or cognitive demands.

3.7.3. Statistical testing

To study the significance of the findings, the data were first analyzed in SPSS to assess whether it followed a normal distribution. Based on the results, parametric tests, such as ANOVA (Analysis of Variance), and non-parametric alternatives to repeated measures ANOVA were applied. These statistical tests were used to determine whether there were significant differences in metrics such as fixation duration, the number of fixations per second, and average saccade length across various experimental conditions.

3.7.4. Post-test questionnaire analysis

To better interpret and analyze, eye tracking data, as well as participants' experiences, and to identify potential areas for improvement in the experimental design for future studies, participants' demographic and background information (e.g., age, gender, educational level, field of study, cartographic experience, and profession) was collected through a pre-test questionnaire. Additional information about their map usage habits, confidence in reading maps, and frequency of Google Maps usage was also gathered. Participants further provided details about their strategies for remembering map landmarks and shared their feedback on the experiment through a post-test questionnaire. To collect information regarding participants' learning strategies, the post-test questionnaire included a task where they were asked to rank the experiment tasks from "easy to remember" to "hard to remember" based on their experience. Participants were also encouraged to share any additional thoughts or suggestions about the study through open-ended responses.

Table 3.2. Summary of Experiment Design.

Main research questions & goals	Can webcam-based, eye tracking effectively replicate and scale the results of lab-based eye tracking studies in the context of map reading? *Goal:* To examine the effect of task type on user behavior in map-reading tasks.
Task description	• **Preparation stage:** Instructions for the experiment, ET calibration, presentation of the training task. • **Map study:** Three memorability tasks focusing on the recognition of different map landmarks: *Task 1:* Roads (roads and road junctions) *Task 2:* Hydrography (major lakes and rivers) *Task 3:* The combination of roads and hydrography Each task included 11 trials (i.e., One map in each task was dedicated to a training task). • **Multiple-choice graphical answer screen:** Second, a question related to some main structuring map elements (i.e. roads and hydrography areas) will appear on the screen for 7 seconds. • **Post-experiment questionnaire:** Answer the same question for the following 33 map stimuli in the Task and basic demographic information.
Participants	28 participants: 14(females) & 14 (males) Age range: 26-50 Average age 38 years (SD = 7.5)
Apparatus	WebGazer.js was used to develop the necessary code for the experimental design.
Data processing procedures	• Selected data based on a calibration threshold of 70%. • Converted raw gaze data (x- and y-coordinates with timestamps) into two formats. • Cleaned data using gazealytics software (to explore the lower quality of recorded gaze samples from participants).
Analysis	- **Quantitative metrics for estimating cognitive load** • Behavioral measures: Response time, Success rate. • Eye tracking metrics: Average fixation duration, Number of fixations per second, Average saccade length. - **Qualitative metrics** • Scanpath visualizations of eye movements across the map using gazealytics software. - **Statistical testing** • Descriptive statistics summarize key trends in the data. • Applied both descriptive and statistical analyses to interpret the results. • Inferential statistics assess the significance of observed effects and relationships.

Chapter 4:
Results and Analysis

4.1. Preprocessing

To ensure the accuracy and reliability of the collected data, Gazealytics software was used to clean and analyze eye-tracking data. This all-in-one toolkit processes, filters, and checks eye-tracking information to maintain consistency and high quality (Chen et al., 2023). Only data with calibration accuracy above 70% were used, which is slightly lower than that of lab setting studies but still sufficient for online gaze tracking. The raw gaze data was carefully checked to identify and remove issues such as missing information, outliers, or data disruptions caused by excessive head movement or technical glitches like calibration slipping. The Gazealytics algorithm automatically identifies fixations (when the eyes stop on something), saccades (quick eye movements), and blinks keeping everything consistent and reducing errors. The study focused on several core metrics: such as average fixation duration (how long participants looked at specific points), the number of fixations per second (indicating visual scanning frequency), average saccade length (the distance between fixations), and behavioral indicators like response time (how quickly participants responded to tasks) and task success rate. This preprocessing step allowed us to detect and correct potential issues, ensuring that the data was reliable and valid for analyzing both behavior and eye movements

.

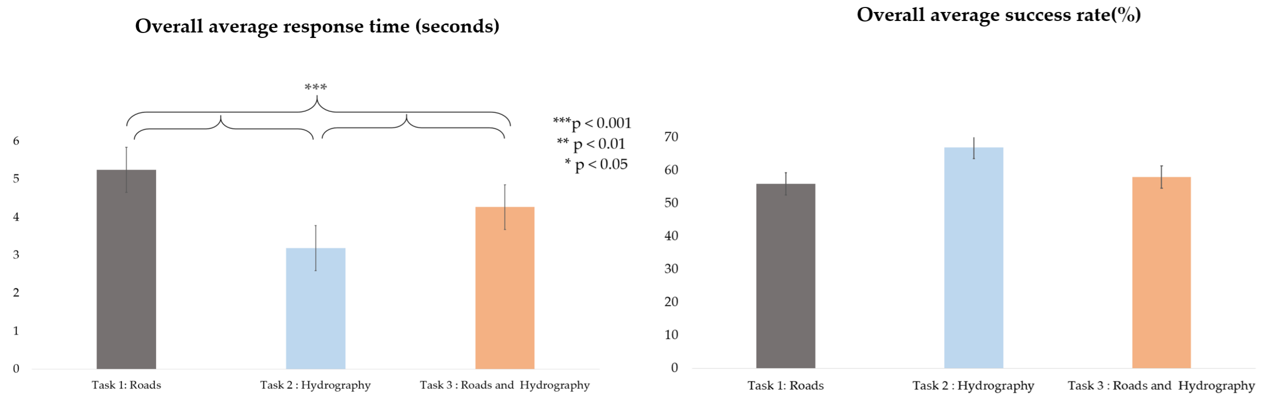
4.2. Quantitative metrics

4.2.1. Behavioral measures

Response time and success rate. To evaluate participant performance and assess task difficulty, response time (RT) and success rate (accuracy) based on correct answers from participants (Figure 4.1) were analyzed. The overall average response time across all tasks was 4.2 seconds (N = 758, MED = 4.2 s, SD =0.8 s), with an overall average success rate of 60% (SD = 18.6%, range = 32.3%–87.3%). These metrics provide a baseline for understanding participant performance across the tasks.

Task 2 demonstrated the highest success rate of 67% (SD = 18.6%) and the shortest response time of 3.1 seconds (N = 255, MED = 3.1 s, SD = 0.2 s). In contrast, Task 1 exhibited the lowest success rate of 56% (SD = 12.0%) and the longest response time of 5.2 seconds (N = 247, MED = 5.2 s, SD = 0.2 s). This indicates that participants made more errors and may have struggled more with accuracy while taking slightly longer to process the Road task.

Task 3 showed a moderate success rate of 58% (SD = 16.3%) and a response time of 4.2 seconds (N = 256, MED = 4.2 s, SD = 0.2 s). This suggests that although participants maintained a reasonable pace, the task's complexity slightly affected both speed and accuracy compared to the individual tasks.



(a) (b)

Figure 4.1. (a) Response times (error bars indicate standard deviation); (b) Success rates (error bars indicate standard deviation).

Normality testing using Kolmogorov-Smirnov test revealed that the success rate data was normally distributed across all three tasks (all $p > 0.05$). A repeated measure ANOVA (a parametric test) was subsequently conducted to compare success rates across tasks. The analysis revealed no significant differences between the three tasks (all $p > 0.05$). Response time data failed to meet normality assumptions across all three tasks. Therefore, a non-parametric Independent-Samples Kruskal-Wallis test was employed to examine response time differences across tasks. The analysis yielded significant differences, χ^2 (2) = 673.020, $p < .001$. Post-hoc pairwise comparisons using Bonferroni correction revealed significant differences between all task pairs: Task 2 vs. Task 1 ($p < .001$), Task 3 vs. Task 1 ($p < .001$) and Task 2 vs. Task 3 ($p < .001$).

4.3. Eye tracking metrics

Average fixation duration. To investigate how long participants focused on the entire stimuli and to understand task specific differences, the average fixation duration was calculated for all three tasks: Road (Task1), Hydrography (Task2), and Road + Hydrography (Task 3). The overall average fixation duration across all tasks was 239 ms (N = 758, MED =231 ms, SD =38.4 ms). This measure serves as an important indicator of the cognitive processing demands associated with each task, as longer fixation durations often reflect increased cognitive effort or task complexity.

As illustrated in Figure 4.2, Task 1 (Roads) recorded an average fixation duration of 250 ms (N = 247, MED = 245 ms, SD = 37.6 ms). Task 2 (Hydrography) showed an average fixation duration of 226 ms (N = 255, MED = 218 ms, SD = 31.2 ms). Task 3 (Roads and Hydrography) had an average fixation duration of approximately 242 ms (N = 256, MED = 239 ms, SD = 41.5 ms).

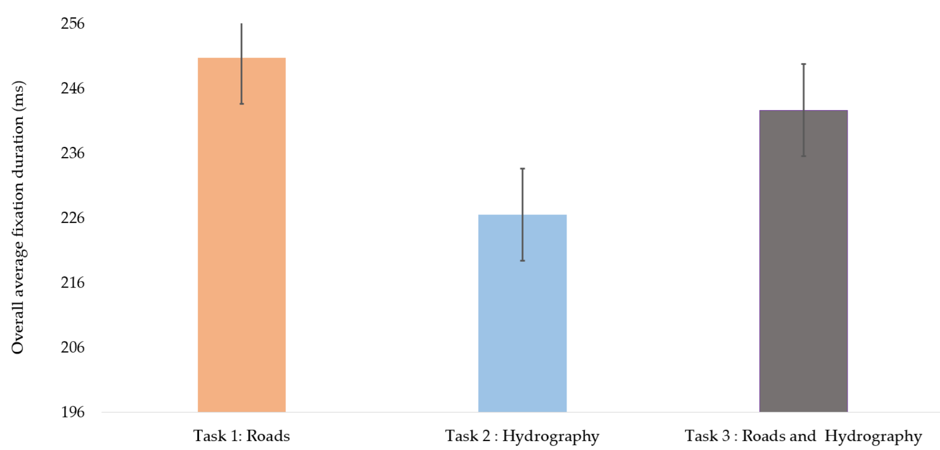


Figure 4.2. Average fixation duration (ms) (error bars indicate standard deviation).

Number of fixations per second. To study the distribution of visual attention across tasks, and the average number of Fixations per Second was analyzed for the three task conditions: Road, Hydrography, and Road + Hydrography.

The overall average across all conditions was 3.3 fixations per Second (N = 758, MED = 3.2 s, SD = 1.1 s), reflecting a balanced attentional demand across tasks. This metric sheds light on participants' visual search strategies and the intensity of attention necessary for task completion. As shown in Figure 4.3, Task 1 exhibited an average of 2.6 fixations per second (N = 247, MED = 2.7 s, SD = 0.7 s).In comparison, Task 2 recorded an average of 4.3 fixations per second (N = 255, MED = 4.1 s, SD = 1.1 s). Task 3 demonstrated an average fixation rate of 3.1 fixations per second (N = 256, MED = 3.2 s, SD = 0.9 s).

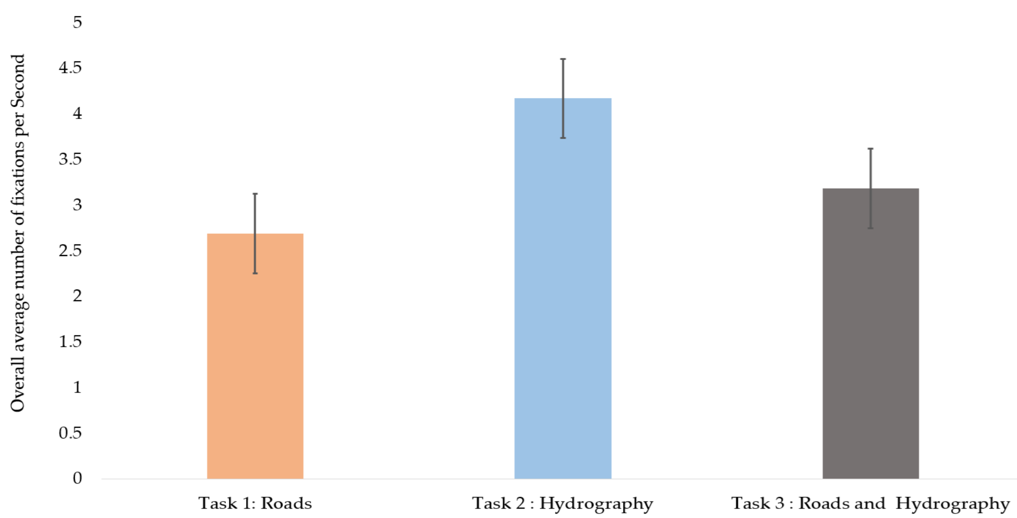


Figure 4.3.Number of fixations per second (error bars indicate standard deviation).

It is a complementary metric to fixation duration and allows for internal validity of the experiment design, as they are generally inversely proportional.

Average saccade length. To further analyze the cognitive load associated with each task, average saccade length (=amplitude) was calculated to quantify the spatial distance covered by eye movements. The overall average saccade length across all tasks was approximately 3.5° (N = 758, MED = 3.4°, SD = 0.7°), providing a baseline for understanding the spatial extent of visual attention across conditions. This metric indicates the nature of participants' visual search strategies: longer saccades suggest a less organized search across the entire image, while shorter saccades reflect a more targeted, less chaotic search between nearby focal points (Keskin, 2020). As shown in Figure 4.4, Task 1 exhibited the longest average saccade length at 3.7° (N = 247, MED = 3.5°, SD = 0.8°). Task 2 recorded the shortest average saccade length at 3.4° (N = 255, MED = 3.3°, SD = 0.5°). Task 3 showed an average saccade length of 3.6° (N = 256, MED = 3.5°, SD = 0.7°).

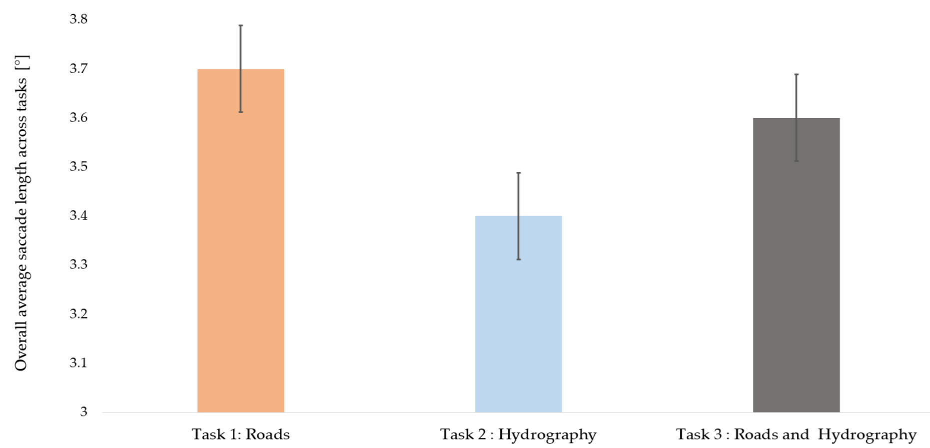


Figure 4.4. Average saccade length [°] (error bars indicate standard deviation).

Overall, the data in Table 4.1 reveals significant differences in cognitive processing demands across the three tasks. Task 1 was identified as the most complex, requiring deeper cognitive engagement and extended visual exploration. This is supported by the post-task questionnaires Figure 4.5, where participants rated Task 1 as the most difficult (83% hard, 17% easy). The objective performance metrics also align with this perception, as task 1 had the lowest success rate (56%) and the longest response time 5.2 seconds. These results indi-cate that Task 1 was indeed challenging, requiring substantial cognitive effort and re-sulting in lower accuracy and slower responses.

The objective performance metrics also align with this perception, as task 1 had the lowest success rate (56%) and the longest response time 5.2 seconds. These results indicate that Task 1 was indeed challenging, requiring substantial cognitive effort and resulting in lower accuracy and slower responses.
In contrast, Task 2 was unanimously rated as the easiest in the post-test questionnaires (0% hard, 100% easy). This perception is strongly supported by the objective performance metrics, as it had the highest success rate (67%) and the shortest response time 4.2 seconds. These findings reflect the task's relatively lower cognitive demand, allowing participants to respond more quickly and accurately.
The task 3 was perceived as moderately difficult in the post-test questionnaires (75% hard, 25% easy). This aligns with its intermediate performance metrics: a success rate of 58% and a response time of 3.1 seconds. The added complexity of multitasking likely contributed to the slight decrease in performance compared to the individual Task2, but it was still more efficient than the Task 1 alone. These findings suggest that subjective perceptions of difficulty, as captured in the post-test questionnaires, generally align with objective performance measures such as success rate and response time.

Table 4.1. Summary of analysis data.

Tasks	Average fixation duration (ms)	Number of fixations per second	Average saccade length [°]	Average response time (seconds)	Success rate (%)
Task 1: (Roads)	250	2.6	3.7	5.2	56
Task 2: (Hydrography)	226	4.1	3.4	3.1	67
Task 3: (Roads and Hydrography)	242	3.1	3.6	4.2	58

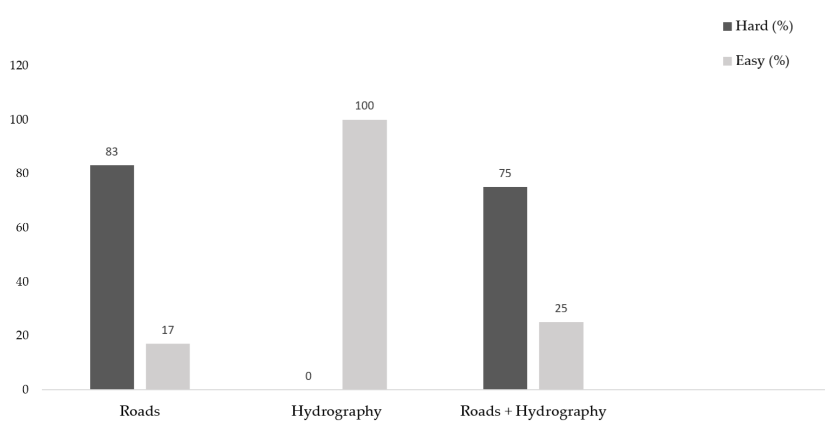


Figure 4.5.Post-test questionnaire results for participants' perceived difficulty order.

4.4. Statistical testing & Interpretation of the results

To study the distributional properties of the eye tracking data and evaluate the statistical significance of the observed results, a comprehensive series of statistical analyses were conducted. The dataset encompassed three distinct experimental tasks, each characterized by specific eye tracking metrics, including average fixation duration (ms), the number of fixations per second, and average saccade length [°]. The table below shows the summary of statistical analysis.

Table 4.2.Summary of statistical analysis of visual attention in different tasks.

Issue	Summary
Tasks	- Task1: (Roads) - Task2: (Hydrography) - Task3: (Roads and Hydrography)
Participants (After Data Cleaning in Gazealytic) **Stimuli & Tasks**	30 participants × 10 stimuli = 300 trials per task - Task 1(Road): 247 trials - Task 2(Hydrography): 255 trials - Task 3(Roads and Hydrography): 256 trials
Metrics	- Average fixation duration - The number of fixations per second - Average saccade length
Analysis	- Normality tests (To assess the distribution of the data.) - Descriptive statistics (To summarize the central tendencies and variability of the metrics.) - Interactions: The **non-parametric Friedman test** was conducted to compare differences across the three tasks after participant removal in the preprocessing stage: - N (sample size)= Task1(Roads): 225 trials - N (sample size)= Task 2(Hydrography): 225 trials - N (sample size)= Task3(Roads and Hydrography): 225 trials - Post-hoc pairwise comparisons (Pairwise comparisons were conducted to identify specific differences between tasks.)

4.4.1. Normality tests and ANOVA

Descriptive statistics. After cleaning the data in Gazealytics software, the Kolmogorov- Smirnov test was employed to assess the normality of the data distribution for each task. The results indicated that the data for all tasks deviated significantly from a normal distribution ($p < 0.05$), violating the assumption of normality required for parametric statistical tests.

The table below presents the descriptive statistics for each task along with the

Sample sizes:

Task 1 (Roads): 247 trials

Task 2 (Hydrography): 255 trials

Task 3 (Roads and Hydrography): 256 trials

Table 4.3 .Descriptive statistics in all three tasks.

Tasks	Eye Tracking Metrics	N	Mean	Median	Standard Deviation (SD)
	Average fixation duration (ms)	247	250	245	37.6
Task 1: (Roads)	Number of fixations per second	247	2.6	2.7	0.7
	Average saccade length [°]	247	3.7	3.5	0.8
	Average fixation duration (ms)	255	226	218	31.2
Task 2: (Hydrography)	Number of fixations per second	255	4.3	4.1	1.1
	Average saccade length [°]	255	3.4	3.3	0.5
	Average fixation duration (ms)	256	242	239	41.5
Task 3: (Roads and Hydrography)	Number of fixations per second	256	3.1	3.2	0.9
	Average saccade length [°]	256	3.6	3.5	0.7

In this study, the dataset consisted of participants who completed all three tasks. However, due to incomplete data from some participants, only the subset of 225 participants who completed all three-task conditions were included in the repeated measures analysis. This ensured consistent and comparable task conditions for the Friedman test.

Normality tests. For assessing the normality of the data distribution, the Kolmogorov-Smirnov (K-S) test was employed, which is suitable for the dataset ($n = 225$).

This test is often preferred for larger datasets as it is less sensitive to minor deviations from normality. The results indicated significant deviations from normality ($p < .05$) in most cases, except for some variables in Task 3 (roads and hydrography).

Average fixation duration:

- Task 1 (Road): K-S $D(225) = 0.087$, $p < .001$ (Not normally distributed)
- Task 2 (Hydrography): K-S $D(225) = 0.181$, $p < .001$ (Not normally distributed)
- Task 3 (Roads and Hydrography): K-S $D(225) = 0.043$, $p = .200$ (Normally distributed)

Number of fixations per second:

- Task 1 (Road): K-S $D(225) = 0.077$, $p = .003$ (Not normally distributed)
- Task 2 (Hydrography): K-S $D(225) = 0.096$, $p < .001$ (Not normally distributed)
- Task 3 (Roads and Hydrography): K-S $D(225) = 0.056$, $p = .080$ (Normally distributed)

Average saccade length:

- Task 1 (Road): K-S $D(225) = 0.158$, $p < .001$ (Not normally distributed)
- Task 2 (Hydrography): K-S $D(225) = 0.084$, $p = .001$ (Not normally distributed)
- Task 3 (Roads and Hydrography): K-S $D(225) = 0.117$, $p < .001$ (Not normally distributed)

Since Tasks 1 and 2 exhibited significant deviations from normality in all variables ($P < .05$), non-parametric statistical methods were applied. In Task 3, average fixation duration and number of fixations per second did not show significant deviations from normality ($P > .05$); however, for consistency across tasks, on-parametric approaches were used.

Interactions. Initial analyses using the Friedman test revealed statistically significant differences across all eye tracking metrics, including average fixation duration, number of fixations per second, and average saccade length ($p < 0.05$). The Friedman test confirmed significant variations in eye tracking metrics among the three tasks, as detailed in Table 4.4. To assess effect size, Kendall's W was used which measures the strength of association between the dependent and independent variables in non-parametric tests like the Friedman test.

According to Tomczak and Tomczak (2014), Kendall's W is a suitable and widely recommended measure for evaluating effect size in the context of the Friedman test, which serves as a non-parametric alternative to repeated-measures ANOVA. The Kendall's W values obtained in this study ranged from small effect (e.g., 0.105) to large effect (e.g., 0.73), reflecting varying levels of association across the tasks (Table 4.4).

These values can be interpreted as follows:

- Small: 0.1 to 0.3
- Medium: 0.3 to 0.5
- Large: 0.5 to 1.0

Furthermore, to assess the statistical power of the independent variable and dependent metric, G*Power software was used. The analysis indicated that the power was high, suggesting that the analyses were highly reliable in detecting any significant effects.

Table 4.4. Main effects of independent variables and task performance analysis.

Independent variable	Dependent metric	Friedman test(alternative to repeated measures ANOVA)	Significance
Task type(3x) (n = 225)	Average fixation duration (ms)	Chi-Square(χ^2) =32.6, w= 0.731,F= 360, $p < 0.001$, power = 1.000	***
Task type(3x) (n = 225)	The number of fixations per second	Chi-Square (χ^2) = 26.1, w = 0. 581, F= 270, $p < 0.001$, power = 1.000	***
Task type(3x) (n = 225)	Average saccade length[°]	Chi-Square (χ^2) = 4.81 w = 0.105, F=376 , $p < 0.001$, power = 0.940	***

Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

4.4.2. Post-Hoc comparisons

Pairwise comparison. Subsequent to the significant findings from the Friedman test, post-hoc analyses were conducted using pairwise comparisons with the Nemenyi test. This approach enabled a detailed examination of the specific differences between task pairs, providing deeper insights into the cognitive demands associated with each condition. Cliff's Delta was used for this purpose, with effect sizes categorized as follows:

Small: |δ|≥0.147 l

Medium: |δ|≥0.330 l

Large: |δ|≥0.474 l

Additionally, to assess the statistical power of the comparisons, these effect sizes were integrated into calculations using G*Power. This ensured a precise evaluation of the study's sensitivity in detecting true differences between conditions, reinforcing the statistical and practical significance of the observed results. The results of these pairwise comparisons are summarized in Table 4.5.

Table 4.5. Post hoc pairwise comparisons with tasks and eye tracking metrics.

Post-Hoc comparison	Dependent metric	Friedman test (alternative to repeated measures ANOVA)	Cliff's Delta	Significance
Task 1(Roads) vs Task2(Hydrography)	Average fixation duration (ms)	F= 0.533, $p < 0.001$, power =0.999	0.548	***
Task1(Roads) vs Task 2(Roads and Hydrography)	Average fixation duration (ms)	F = 0.200, p = 0.34, power =0.452	0.252	Not significant
Task 2(Hydrography) vs Task 3(Roads and Hydrography)	Average fixation duration (ms)	F = 0.333 $p < 0.001$, power = 0.892	0.382	***
Task1(Roads) vs Task 2(Hydrography)	Number of fixations per second	F= 0.447, $p < 0.001$, power =0.991	0.480	***
Task 1(Roads) vs Task3(Roads and Hydrography)	Number of fixations per second	F= 0.93, p = 0.322, power =0.571	0.124	Not significant
Task 2(Hydrography) vs Task 3(Roads and Hydrography)	Number of fixations per second	F = 0.353, $p < 0.001$, power =0.925	0.400	***
Task1(Roads) vs Task 2(Hydrography)	Average saccade length[°]	F= 0.582, $p < 0.001$, power =0.999	0.584	***
Task 1(Roads) vs Task 3(Roads and Hydrography)	Average saccade length[°]	F = 0.044, p = 0.63, power =0.393	0.101	Not significant
Task 2(Hydrography) vs Task3(Roads and Hydrography)	Average saccade length[°]	F = 0.538, $p < 0.001$, power =0.999	0.551	***

Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

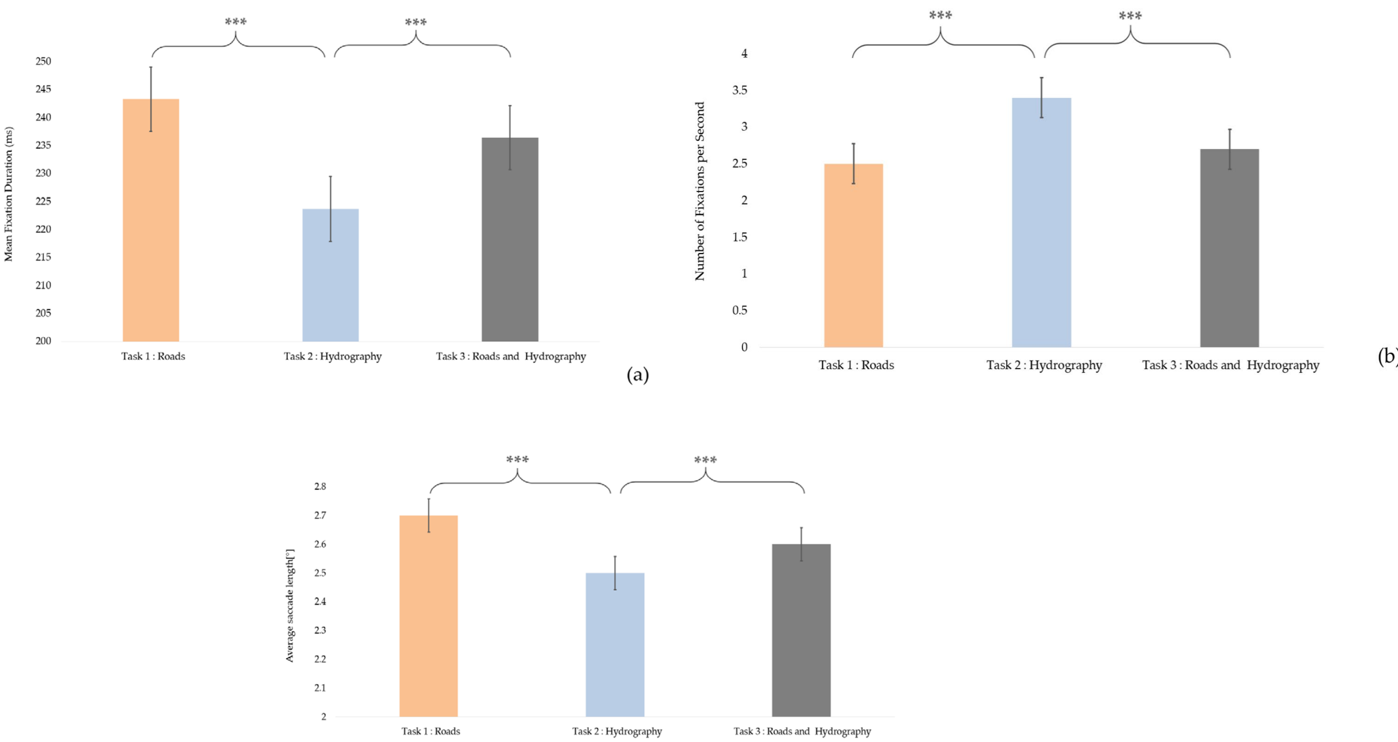
For average fixation duration, Task 1 resulted in longer fixation durations than Task 2, with a large effect size. Similarly, Task 3 exhibited longer fixation durations than Task 2, with a medium effect size.

For the number of fixations per second, Task 1 had a higher fixation rate than Task 2, with a large effect size. Likewise, Task 3 showed a higher fixation rate than Task 2, with a medium effect size. Regarding Average saccade length, Task 1 produced longer saccades than Task 2, with a large effect size. Similarly, Task 3 exhibited longer saccades than Task 2, also with a large effect size.

Overall, the results reveal clear patterns of visual attention and cognitive demand. Task 1 required the greatest attentional engagement, as indicated by longer fixation durations, higher fixation rates, and distinct saccadic behavior. Task 2 showed reduced attentional demand, with shorter fixations and lower fixation rates, suggesting different cognitive processing demands (Table 4.6) .These findings suggest that road related information demands greater visual attention and imposes a high cognitive load, as reflected in prolonged fixation durations and increased fixation rates. The addition of hydrographic features in Task 3 did not significantly reduce this demand, as Task 3 remained statistically similar to Task 1 (Figure 4.6).

Table 4.6. Descriptive statistics in all three tasks (n=225).

Tasks	Eye Tracking Metrics	N	Mean	Median	Standard Deviation (SD)
Task 1: (Roads)	Average fixation duration (ms)	225	246	238	35.3
	Number of fixations per second	225	2.5	2.4	0.8
	Average saccade length [°]	225	2.7	2.5	0.7
Task 2: (Hydrography)	Average fixation duration (ms)	225	226	218	29.3
	Number of fixations per second	225	3.4	3.4	0.6
	Average saccade length [°]	225	2.5	2.4	0.3
Task 3: (Roads and Hydrography)	Average fixation duration (ms)	225	239	239	38.9
	Number of fixations per second	225	2.7	2.7	0.8
	Average saccade length [°]	225	2.6	2.5	0.3



(c)

Figure 4.6.Average of (a) fixation duration (ms) (b) saccade length ([°]) (c) the number of fixation per second (error bars indicate standard deviation).

Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

4.5. Qualitative metrics

Heatmaps and scanpath visualizations. To conduct a preliminary exploratory analysis, gaze behavior was examined by comparing the encoding and decoding stages. Gaze density maps heat maps and scanpath visualizations were generated in Gazealytics to analyze how participants interacted with the map stimulus and its corresponding graphical answer screen.

For Task 1 (Road) depicted in Figure 4.7(a), participants were instructed to study (seven-second long encoding stage) and recall (decoding stage ending with participant's response) the main roads and road junctions. In the decoding stage, they were required to select the correct representation of the road landmarks from a multiple-choice graphical survey. The visualization includes fixations data, gaze based heat maps, illustrating user behavior and accuracy in map interpretation. The scatter plots display individual gaze points (dots) from participants' eye tracking data overlaid on the map. These plots indicate where participants looked during both the encoding and decoding stages, with a clear focus on main roads and junctions, as instructed.

Furthermore, in the encoding stage, the saliency map reveals a wide distribution of attention with multiple focal points across the map, suggesting that participants broadly scanned the map to encode the information.

In contrast, during the decoding stage, the saliency map exhibits more concentrated areas of attention, reflecting participants' efforts to focus on specific regions to complete the graphical survey task. Notably, the saliency contours align more closely with the main roads and junctions, in accordance with the instructions.

Approximately 78% of participants answered correctly on the graphical survey for this map in Task 1, indicating that most effectively encoded and recalled the necessary information about the map's main roads and junctions.

However, discrepancies were observed between gaze patterns and the actual areas of interest, suggesting potential distractions or gaze estimation errors. The calibration accuracy of the eye tracking data was 73%, meaning that 73% of the gaze data was accurately tracked and mapped, while the remaining 27% may reflect noise or system inaccuracies.

For Task 2 (Hydrography)**,** as shown in Figure 4.7(b), participants were instructed to identify major lakes and rivers. During the encoding stage, scatter plots of gaze data reveal that most gaze patterns were strongly centered, despite hydrographic features being more dispersed toward the corners of the map. However, the heatmap for this stage indicates a wide distribution of attention, with multiple focal points spread across the map. In the decoding stage, both scanpath and heatmap show more concentrated gaze patterns, reflecting participants' efforts to focus on specific regions while completing the graphical survey task. In particular, the saliency map of gaze data in the decoding stage highlights the most focused areas of eye tracking data, which were concentrated between graphical response options (C) and (D) on this hydrographic map for Task 2.

In contrast, during the encoding stage, the most concentrated eye tracking data was centered in the middle of the map. Despite these variations in gaze patterns, participants performed well, with 60% answering correctly, suggesting that a majority effectively encoded and recalled the relevant hydrographic features. However, discrepancies were observed between gaze patterns and the actual areas of interest on the map, indicating potential distractions or attentional errors.

For Task 3 (Road & Hydrography), as presented in Figure 4.7(c), participants were required to focus on both roads and hydrographic features. Analysis of raw gaze data, saccade pattern, and heatmap reveals that during the encoding stage, participants broadly scanned the map to encode both feature types. In the decoding stage, gaze patterns indicate that participants concentrated on recalling relevant roads and hydrographic features. The heatmap suggests that participants distributed their attention across both feature types. Specifically, in the decoding stage, heatmap reveals a central fixation bias, although participants made efforts to scan multiple areas. The results indicate that approximately 71% of participants answered correctly on this map for Task 3, despite some discrepancies between gaze patterns and the actual areas of interest. These discrepancies maybe attributed to limitations of webcam-based eye tracking (WebGazer.js), which introduced slight deviations in the gaze data. While the heat map suggests that participants' gaze was not always perfectly aligned with key map features, the high percentage of correct responses indicates that they successfully identified the relevant areas.

Due to deviations in the gaze data, accurate calculation of Areas of Interest (AoIs) analysis was not possible. To assess the extent of these deviations, additional tests were conducted, revealing that WebGazer.js did not accurately capture eye tracking positions relative to the map.

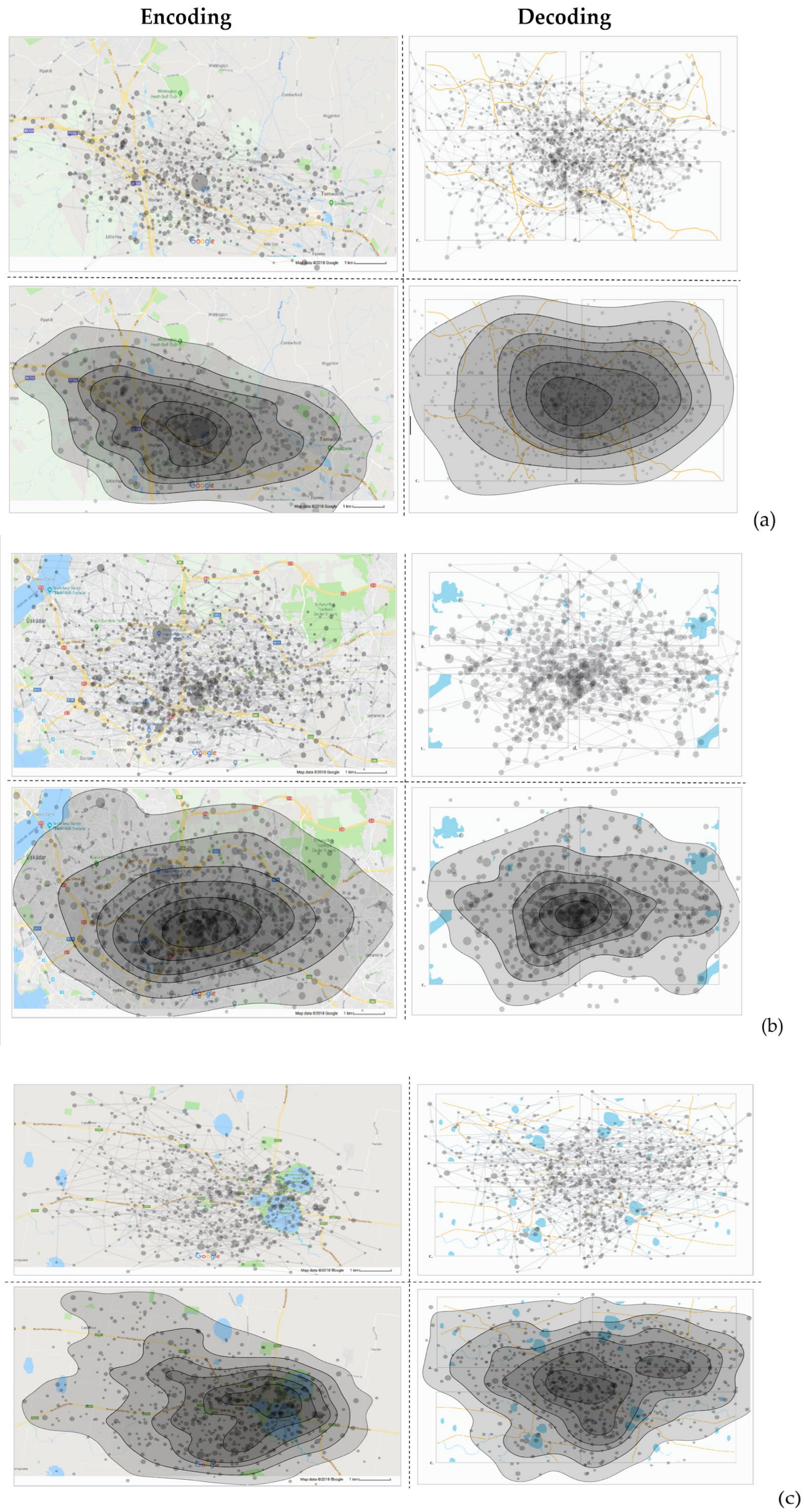


Figure 4.7.Gaze density maps (heatmaps) and (scanpaths) during encoding and decoding stages in Gazealytics. Task 1: Roads (a), Task 2: Hydrography (b), Task 3: Roads and Hydrography(c).

4.6. Post-test questionnaire analysis

The study included 28 participants, with an equal gender distribution of 14 females and 14 males. The average age of the participants was 38 years (SD = 7.5), with ages ranging from 26 to 50 years. Participants had varied educational backgrounds: 9 held a bachelor’s degree (ages 29–46), 10 held a master’s degree (ages 27–45), and 2 held a PhD, with an average age of 47.5 years. The most common fields of study were geography (6 participants), geoinformatics/geoinformatics and cartography (4 participants), and other disciplines such as computer Science, civil engineering, and environmental science. A significant proportion (42.9%) of participants reported having a background in cartography, indicating substantial expertise in map related fields.

Professionally, participants included students, researchers, faculty members, and engineers, with roles such as PhD students, study directors, and urban drainage engineers. Participants with backgrounds in cartography or related fields (e.g., geography, geoinformatics) reported higher confidence in their map interpretation abilities and more frequent map use. Specifically, all 6 participants with geography degrees rated their confidence as “very confident,” and 5 of them used maps daily or weekly. In contrast, participants without cartography backgrounds, such as those in biomedical engineering or educational psychology, reported lower confidence and less frequent use, often-ranking hydrography as particularly challenging.

Participants were also asked whether they wore glasses or contact lenses while completing the survey. More than 17 participants (60.7%) reported not using any visual aids, while 10 participants (35.7%) wore glasses, and one participant (3.6%) used contact lenses. This information may be relevant in understanding potential differences in visual perception during map interpretation.

Furthermore, Participants were asked to evaluate the ease of using Google Maps on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). The majority (50%) strongly agreed (rating of 4), indicating a high level of satisfaction with Google Maps' usability. Additionally, 42.9% of participants reported using Google Maps in their daily lives, while only 3.6% used it less than once or twice a week. Regarding the types of maps used, participants who worked with cartographic products daily primarily used topographical (7.14%) and thematic maps (17.86%), whereas participants without a cartography background predominantly used navigational maps (60.71%) in their daily routines.

Figure 4.8 shows a detailed result of the post-test questionnaire.

Qualitative comments provided further insights into participants' experiences. Many participants praised the clarity of features, effective labeling, and overall design of maps. However, one participant noted that the map scale was inadequate for properly reading map features, suggesting a potential area for improvement. Although not directly relevant to their map learning strategy, participants shared diverse strategies for remembering maps.

The most common approach involved memorizing distinctive features such as rivers, roads, waterways, meanders, intersections, lakes, and streets. For instance, participants concentrated on intersections for roads, meanders for waterways, and major geometries (e.g., highways oriented North to South/East to West) before examining map edges, corners, and counts of lakes and streets. A significant number of participants emphasized visualizing and memorizing the overall structure of the map, including colors, shapes, patterns, and sizes of landmarks. Strategies such as "visualizing the landmark's shape and size," "looking at map shapes," and "memorizing the overall shape of some map parts," indicate a reliance on visual memory and pattern recognition.

These techniques were particularly relevant for navigational and road maps, which were the most frequently used map types (navigational: 60.7%, roads: 10.7%). This reliance on visual strategies may explain the high confidence levels (53.6% confident, 25% very confident) among participants with cartography backgrounds (42.9%).

Participants were asked to rank the experimental tasks from "easy to remember" to "hard to remember" based on their experiences. Their rankings completely aligned with their actual performance in terms of success rate and response time. A majority (50% of females and 32.58% of males) ranked Task 1 (roads) as the hardest task, while all (100%) females and males ranked Task 2 (hydrography) as the easiest. These discrepancies suggest that subjective difficulty assessments influenced by prior experience and familiarity with different map types rather than objective performance measures. These qualitative insights offered valuable feedback for refining the experimental design and understanding participant behavior. A summary chart of the post-test questionnaire results is presented below.

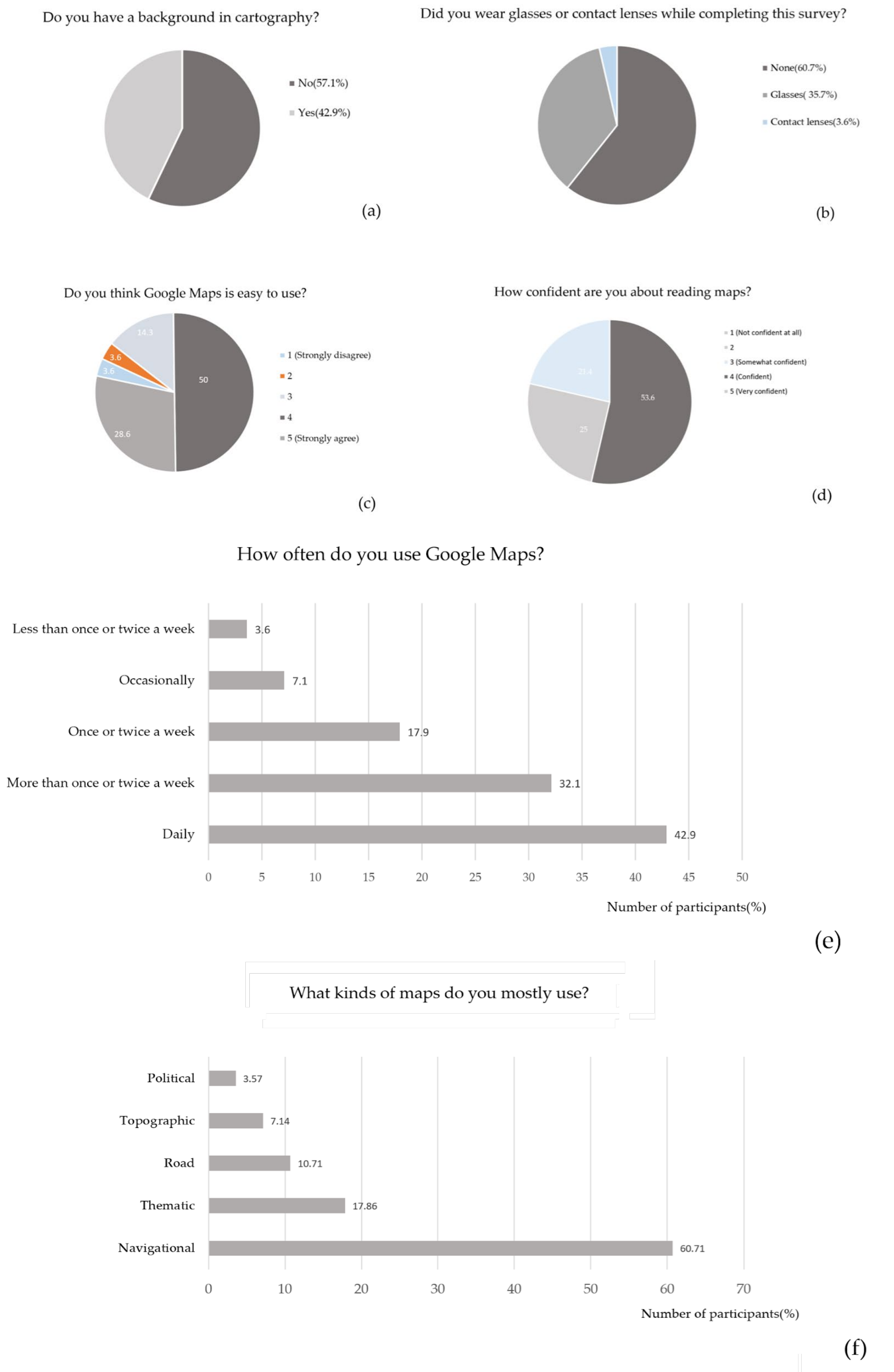


Figure 4.8. Participants' characteristics gathered from the post-test questionnaire.

Chapter 5:

Discussion

5.1. Quantitative metrics interpretation

5.1.1. Behavioral measures

Response time and success rate. The results of this study indicate some interesting differences in behavioral measures between this webcam-based eye tracking study (WebGazer.js) and Keskin's (2020) lab-based study using SMI RED250 software.
Participants in the WebGazer.js study responded more rapidly except for the tasks requiring the retrieval of the road features (Task 1) where we also observed the lowest response accuracy. This might suggest that the controlled lab environment may have fostered greater participant focus or diligence, potentially enhancing performance due to the structured setting. Research by Edler et al. (2015) supports this, showing that the recognition of linear features alone is more difficult than remembering them in combination with other features or within a contextual setting.
On the other hand, the experimental designs in lab and online settings differed markedly. Keskin's (2020) lab-based study used seven task blocks dedicated to the recognition of other map landmarks, too, and each block consisted of the same 50 stimuli in a randomized order to combat the learning effect. In contrast, the WebGazer.js study employed three task types, and each contained 11 different stimuli (in total 33), therefore there was no randomization. This is why we think webcam experiment might introduce cognitive load or unfamiliarity, particularly in the road task, where slower responses did not yield improved accuracy as it was in the lab settings.

5.1.2 Eye tracking metrics

Average fixation duration. Looking at how long participants focused on different tasks in this WebGazer.js study, we saw some interesting patterns. Task 1 likely required significantly longer fixation durations than Task 2, indicating greater visual and cognitive effort required to process road features. The significant difference in fixation durations between Task 1 and Task 2 suggests that the hydrographic features in Task 2 imposed a lower cognitive load when integrated with road features. Similarly, Task 3 showed significantly longer fixation durations compared to Task 2, suggesting that the integration of road and hydrographic features increased attentional demand relative to hydrographic features alone.

However, no significant difference emerged between Task1 and Task 3, suggesting that the presence of both road and hydrographic features in Task3 did not substantially increase task difficulty relative to Task 1, which focused solely on roads. This is likely due to the experimental design, which used unique stimuli in Task 3 with counterbalanced road and hydrographic features. This interpretation is supported by participants' performance, as success rates showed no significant differences across tasks, indicating that the inclusion of hydrographic features alongside roads in Task 3 did not impose a significant additional cognitive load compared to Task 1. Moreover, patterns of visual attention were closely aligned with performance, with fixation durations primarily influenced by task relevant features such as the roads in task 1 rather than by overall complexity.

Keskin's (2020) found that the duration of participant focus remained steady across tasks, with only small differences between simpler and more complex tasks. In contrast, this study showed longer focus for Task 1 than Task 2. This difference may stem from the setup: Keskin separated participants by expertise, while this study grouped all together. However, Keskin et al. (2023) suggest expertise has little effect on memorability performance. Therefore, while combining participants may have overlooked some effects of expertise, it allowed a more focused examination of how the tasks themselves shaped attention patterns. Both studies support the idea that fixation duration is closely tied to task relevant features rather than overall task complexity. Keskin's (2020) lab-based study suggests that increased visual complexity (i.e., the number and arrangement of map elements) does not consistently lead to longer fixations, as attention was influenced more by the type and relevance of map objects than by the overall visual detail.

Similarly, the results of this study showed that fixation durations varied depending on the specific visual demands of each task; for instance, the road networks in Task 1 elicited different attention patterns compared to the hydrographic elements in Task 2. This aligns with Keskin's (2020) observation that added complexity map does not necessarily lead to substantial changes in visual attention. These findings suggest that fixation duration is driven more by task relevant attention where participants focus on elements critical to completing the task than by the sheer quantity of visual stimuli. Consequently, although fixation metrics provide meaningful insights into visual engagement, their utility in evaluating task difficulty may depend significantly on the characteristics of the visual stimuli (repeated measures in each task block vs. unique stimuli in each trial) and the data collection method (remote eye tracker vs. webcam eye tracker) used.

As noted by Slim et al. (2024), methodological differences such as using webcam-based eye tracking versus high precision infrared systems can influence the granularity and reliability of fixation based measures.

Therefore, we should be cautious against over interpreting average fixation durations; as such measures may not fully capture the nuanced cognitive and affective processes involved in complex task performance (Negi and Mitra 2020).

Number of fixations per second. As a complementary metric to average fixation duration, the analysis of the number of fixations per second revealed distinct patterns across tasks. A significant difference was observed between Task 1 and Task 2, which is logical given that Task 2 likely required more active visual engagement due to the complexity of hydrographic features compared to focusing only on roads in Task 1.A statistically significant difference was observed between Task 2 and Task 3, indicating that retrieval of both hydrographic and road features resulted in fewer fixations per second than hydrographic features alone. This suggests that the type and complexity of the task can influence visual scanning behavior and aligns with the eye tracking research in terms of visual complexity of maps (e.g., Edler et al., 2020; Çöltekin et al., 2020)
However, Keskin (2020) reported a steady number of fixations per second across tasks of varying complexity in a lab-based setting, suggesting that increased task complexity does not always lead to substantial changes in eye movement behavior. In conclusion, while the number of fixations per second offers valuable insights into visual engagement during map reading tasks, its effectiveness as a standalone indicator of task difficulty is limited, therefore, is generally used as a complementary metric to fixation duration and is useful for the internal validity of the experiment design.

Average saccade length (Amplitude). To further interpret cognitive load and learning strategy of participants the average distance between the fixation or saccade length (amplitude) was analyzed as it is associated with the ability of focused attention and the complexity of the task and learning strategy of participants. Variations across tasks revealed patterns linked to task demands and participant behavior. Task 1 resulted in the longest saccades whereas; Task 2 resulted in the shortest saccades. This statistically significant difference observed between Task 1 and Task 2 was consistent with the fixation metrics demonstrating that retrieval of road features required broader visual scanning and potentially higher cognitive effort. Task 1 may have encouraged broader eye movements due to the linear and spatially distributed structure of road features, which required participants to scan larger areas of the map compared to the more compact hydrographic elements in Task 2.

The retrieval of hydrography and roads in Task 3 did not substantially alter scanning behavior relative to Task 1 alone. On the other hand, retrieval of both road and hydrographic features in Task 3 resulted with significantly longer saccade lengths compared to remembering only hydrographic features in Task 2, indicating that task type (i.e. recognition of road features) significantly shapes visual exploration behavior beyond simply increasing task complexity. More demanding tasks led to shorter saccades, while easier and moderately difficult tasks resulted in longer saccades (Keskin 2020).

In the current study, the most demanding task (Task 1) exhibited the longest saccades, Gegenfurtner, Lehtinen & Säljö (2011) claimed that longer saccade length reflects the ability for holistic analysis. Interestingly, Task 2, considered less complex, but received the shortest saccades. This shows that hydrographic features required more targeted viewing behavior, possibly encouraged by the specific nature of hydrographic features, to be remembered. Task 3, which received intermediate saccade amplitudes, appears to represent a balance between broader and more focused scanning behaviors, adapting to the combined demands of roads and hydrography. Further supporting this perspective,

Skaramagkas et al. (2023) noted that average saccade amplitude could decrease with increasing cognitive load, emphasizing the complex relationship between task demands and eye movement behavior.

Overall, while saccade amplitude provides valuable insights into cognitive processing during map tasks, it cannot serve as a straightforward indicator of task difficulty. In webcam eye tracking experiments, especially when the trial duration is short due to the low sampling rate of webcam eye tracking not all scan paths can be captured.

5.2. Qualitative metrics interpretation

Heatmaps and Scanpath .Preliminary analysis of heatmaps and scanpath data from this study showed that, despite participants achieving a 60% overall success rate in identifying graphical survey elements such as hydrographic features or roads , the aggregated eye tracking data did not consistently align with expected gaze patterns on the map. This discrepancy raises questions about the reliability of the gaze data as an indicator of participant attention, even when correct responses were provided.

In comparison, Keskin’s (2020) study, which used SMI RED250 a commercial eye tracking software, reported that 94.9% of participants successfully identified target areas on a hydrography map classified as easy.

In that study, heatmaps and scanpaths revealed precise fixations on relevant map regions prior to responses, suggesting a strong correspondence between gaze behavior and task performance. By contrast, our findings indicate a notable gap between successful outcomes and the corresponding eye tracking patterns, hinting at potential inaccuracies in gaze data capture or differences in task demands. Notably, our setup showed a 27% error rate in gaze tracking, which likely contributed to the fixation-offset errors observed in the data. While this level of accuracy is considered acceptable for webcam-based eye tracking systems like WebGazer.js, especially in uncontrolled environments (Papoutsaki et al., 2016), it proved insufficient for this study. The maps used contained a large number of small and closely spaced Areas of Interest (AoIs), increasing the need for high spatial precision.

As a result, even minor tracking inaccuracies made it difficult to determine whether participants were truly fixating on specific AoIs, rendering the dataset unsuitable for reliable AoI-based analysis.

5.3. Limitations of the study

Quality of recorded raw data. While this study aimed to examine the replicability of an in lab system, the quality of data collected did not match the level typically achieved with established commercial software, such as the SMI RED250 system, which samples at 250 Hz compared to the 30 Hz sampling rate of standard webcams.

Furthermore, the online methodology employed introduced several inherent limitations. Although the rate of participant exclusions was comparable to that observed in traditional lab-based studies, WebGazer.js relies on remotely sourced data without control over participants' actions or positioning, contributing to variability in data quality. This was reflected in the 27% gaze tracking error and the lack of consistent fixations observed in the scanpath data, which frequently failed to align with task relevant map regions. These findings highlight the challenges associated with achieving reliable spatial accuracy in remote eye tracking contexts.

Fixation-offset error in gaze data. To assess the accuracy of the eye tracking data collected in this study, recorded gaze points from 28 participants during the calibration phase were compared to the known positions of two predefined targets: a calibration grid consisting of nine calibration points and a single calibration focus point. Gazealytics software was used to evaluate the alignment of the recorded gaze data with the predefined calibration targets. The visualizations presented in Figure 5.1 (calibration points) revealed notable discrepancies. These discrepancies, observed as deviations of gaze points relative to the yellow dots representing the calibration targets, suggest potential errors attributable to hardware limitations, participant movement, or environmental factors. An average calibration accuracy of 73%, as reported by WebGazer.js and verified through raw data inspection, provides a baseline for understanding the distribution and magnitude of these errors.

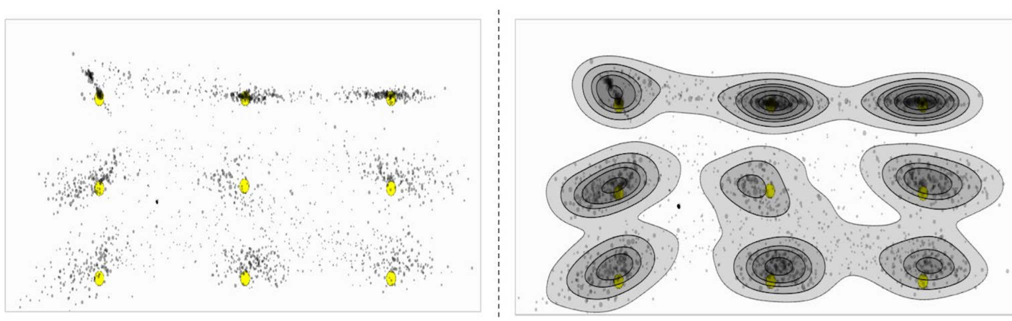


Figure 5.1. A visual representation of gaze tracking data during the calibration step in Gazealytics.

Eye tracking error refers to the difference between position predicted by the eye tracking system (WebGazer.js) as the participant's gaze and the actual point of fixation (ground truth). To calculate eye tracking error in this study, we utilized a single calibration target the center point on which participants were instructed to maintain their focus during the calibration step. A “fixation-offset error” happens when the estimated gaze point does not perfectly align with the fixation point due to systematic inaccuracies in determination of the gaze point. This method for calculating eye tracking errors follows the approach described by Papoutsaki et al. (2016) for WebGazer.js, using focus-point calibration and the Euclidean distance to measure deviations from the ground truth.

To calculate the deviation between the predicted gaze points and the true fixation point, and to ensure consistency across different screen sizes, the following normalized Euclidean distance formula was employed:

$$d_i = \sqrt{\frac{1}{2}\left((\frac{x_{i,pred} - x_{true}}{width/2})^2 + (\frac{y_{i,pred} - y_{true}}{height/2})^2\right)}$$

where $x_{i,\ pred}$ and $y_{i,pred}$ are the ith predicted gaze coordinates, and *width* and *height* denote the display dimensions. Note that d_i is chosen so that its value lies between 0 and 1 when the fixation point coordinates are equal to the central point of the display; that is,

x_{true} = width / 2 , y_{true} = height / 2 .

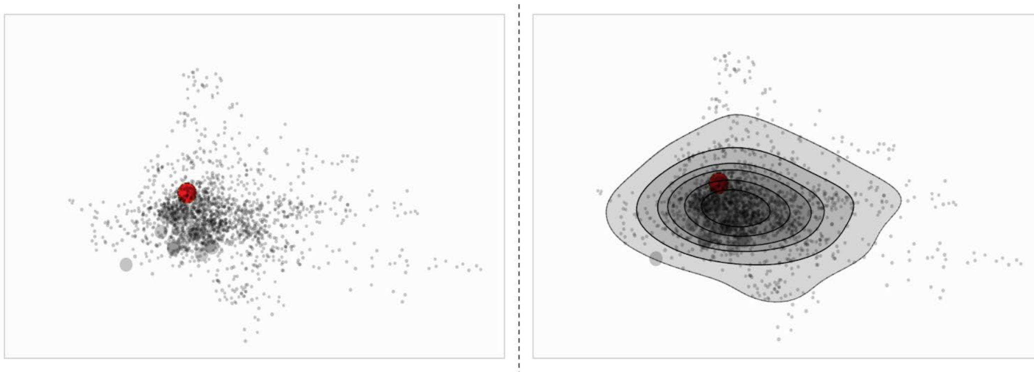
The final value of d is obtained by averaging over 50 predicted points.
This averaged normalized distance is essentially the statistical error of the prediction, obtained empirically.
The precision is then computed accordingly as and reported as a percentage.

$$\text{Precision} = 1 - \overline{d}$$

Using this approach, an average calibration accuracy of 73% was calculated, reflecting the percentage of gaze points reliably aligned with the target fixation point. The corresponding average error was:

$$\text{Avg. Error} = 1 - 0.73 = 0.27(27\%)$$

This eye tracking precision during calibration indicates reliable alignment with focus point for the majority of fixation points. The mean errors confirm good performance, while the 27% error rate implies some room for improvement, such as refining calibration procedures or minimizing external influences or hardware limitations, calibration inaccuracies, participant movement, or environmental factors. These findings, supported by WebGazer.js visualizations, provide a robust foundation for evaluating task performance in subsequent encoding and decoding phases. Figure 5.2 illustrates the distribution of fixation points, providing a visual representation of gaze tracking accuracy and precision during the calibration phase for all 28 participants.



(a)

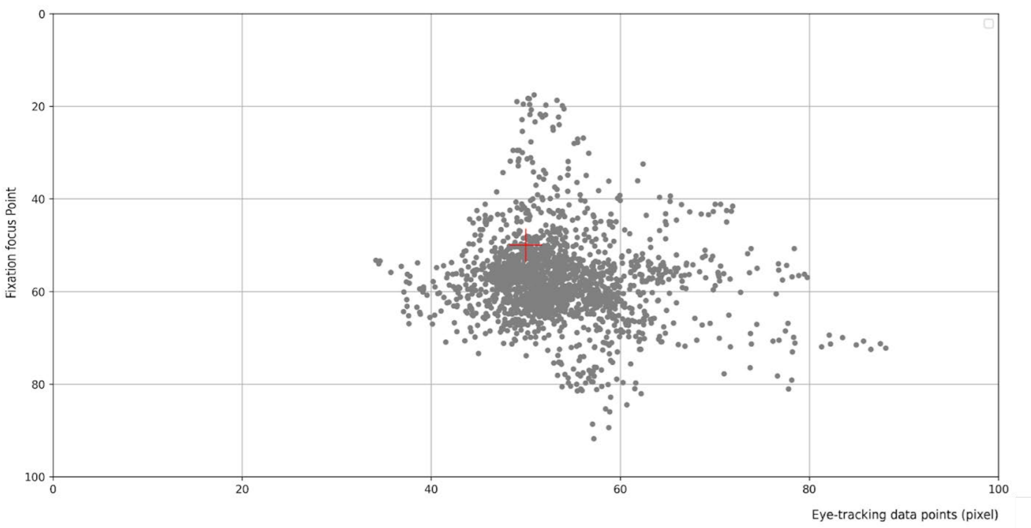


Figure 5.2. Visual representation of gaze tracking accuracy and precision from 28 participants during the calibration phase. (a) Aggregated gaze data displayed in Gazealytics. (b) Scatter plot showing the empirical distribution of gaze points in pixels relative to the fixation target (Red Cross).

Minimize fixation-offset error in gaze data. To study the origins of spatial deviations in eye tracking data and determine whether they arise from device specific factors or individual differences, a dataset from a single participant under tightly controlled conditions was tested. The participant used a high resolution camera equipped device with a display resolution of 1516 pixel in width and 725 pixel in height, positioned at a fixed viewing distance of 60 cm. This setup showed an impressive calibration accuracy of 93%, suggesting that high precision is achievable when variables are carefully controlled.

Furthermore, an analysis of the gaze data and distance calculations revealed a systematic spatial error, characterized by a mean offset of x = +0.38 and y = -1.37 relative to the designated fixation point (50, 50). This translates to a 1.42-pixel deviation, predominantly downward and slightly to the right, as determined by the Euclidean distance metric. Notably, temporal error remained minimal, and participant related variables, such as movement or inconsistent gaze behavior, were effectively controlled in this case.

These findings suggest that the observed error is not a product of individual differences but stems from device specific factors, likely due to inaccuracies in the calibration process or limitations in the mapping algorithm employed by the eye tracking system. In order to minimize fixation-offset error in gaze data, several targeted strategies can be implemented to address the underlying issues.

Fixation-offset error, defined as the systematic inaccuracy where the estimated gaze point fails to align precisely with the actual fixation point, can originate from multiple sources, including hardware limitations, participant movement, environmental influences, and calibration inaccuracies. By systematically addressing these factors, the reliability and precision of eye tracking data can be significantly improved. As mentioned in the fixation-offset error in gaze data section, dynamic recalibration is an effective method to reduce systematic spatial errors. This adaptive process keeps the system responsive to changing conditions throughout the experiment.

Additionally, refining the initial calibration procedure offers another avenue for reducing fixation-offset errors from the outset. This could involve increasing the number of calibration points beyond the nine-point grid used in the study or conducting calibration at multiple intervals to account for temporal variability.

Extending the duration of fixation on a single focus point or incorporating additional focus points could further establish a more robust baseline, thereby mitigating systemic inaccuracies in gaze estimation. These enhancements would strengthen the calibration framework and reduce initial offset discrepancies. Participant movement represents a significant contributor to fixation-offset error.

To address this, physical stabilization techniques, such as headrests, chin rests, or other restraining devices, could be employed to limit head and body shifts. By maintaining the eyes in a consistent position relative to the eye tracking hardware, these measures would minimize deviations between predicted and actual gaze coordinates, particularly during extended experimental sessions. In addition to these techniques, hardware limitations, such as low camera resolution or inconsistent frame rates in the WebGazer.js system, may also introduce offset errors.

Upgrading to higher-quality cameras or improving the sampling rate of the eye tracking system could enhance the precision of gaze point detection. Moreover, ensuring proper alignment and calibration of the hardware such as optimizing camera angle and focus prior to each session could reduce systematic biases, further improving data accuracy.
Environmental variables, including fluctuating lighting conditions or screen reflections, can compromise eye tracking performance. By maintaining consistent, diffuse lighting and minimizing glare on the display, noise in the gaze data can be reduced. Shielding the experimental setup from external light sources or utilizing anti-reflective screens could further stabilize system performance, decreasing fixation-offset errors.
Furthermore, advanced post processing techniques within the gaze prediction model provide another means of correcting fixation-offset errors. For instance, machine learning algorithms could be trained on calibration phase data to detect and adjust for systematic offset patterns, refining the predicted gaze coordinates. Alternatively, filtering outliers from the 50 predicted points used to compute the averaged normalized distance (d) could help diminish the influence of erratic data points, leading to a more accurate error estimate. Similarly, participant behavior can be optimized to reduce offset errors. Providing real-time feedback during calibration, such as a visual indicator of gaze position, could encourage more precise fixation on target points. Training participants to maintain focus and minimize blinks or extraneous eye movements during critical phases might further reduce variability in the gaze data, contributing to lower offset errors. Collectively, these strategies will help determine whether the observed discrepancies arise from methodological limitations or behavioral factors. Such insights are crucial for refining our approach and improving the validity of gaze data in future studies, ensuring that fixation-offset errors are minimized, and the eye tracking system provides a robust foundation for scientific inquiry.

Technical limitation. Despite many advantages of online eye tracking several limitations of WebGazer.js compared to in lab eye tracking methods should be highlighted. In this study, 37 participants were recruited, of whom 35 provided complete datasets. However, one participant's data was only partially received, with 20-recorded choice trials, while another had only 10-recorded trials. This data loss underscores the challenges associated with online implementations, particularly in ensuring data completeness.
A significant technical challenge stemmed from the use of Google Apps Script, a cloud-based platform that integrates with Google services for efficient data storage.

At times, errors arose during data transmission, limiting the ability to receive complete datasets. The WebGazer.js source code offers limited troubleshooting and debugging resources, making it difficult to independently resolve integration issues.

One major constraint is WebGazer's low sampling rate, typically around 20–30 Hz, which is significantly lower than the 250 Hz offered by conventional infrared eye trackers. This lower temporal resolution limits the precision of gaze data and reduces the reliability of fixation-based metrics, particularly in tasks requiring fine-grained temporal analysis.

Tracking accuracy was also affected by hardware variability among participants. Differences in webcam quality, screen size, and ambient lighting conditions introduced inconsistencies in gaze estimation across users.

Compounding this issue, WebGazer.js does not dynamically correct for head pose shifts. Once calibration is completed, any subsequent head movement can result in misalignment and increased noise in gaze data. While WebGazer.js is designed to display warning messages when participants move excessively, this feature did not function consistently across browsers particularly in Firefox further diminishing its effectiveness.

During the final testing phase of the code, it was observed that WebGazer.js did not function correctly in certain browsers, such as Firefox. Specifically, a warning message intended to alert participants if they moved their heads excessively did not appear in Firefox. This lack of real-time head pose compensation presents a significant challenge for maintaining gaze accuracy in uncontrolled, naturalistic settings.

Moreover, the WebGazer.js source code lacks thorough testing across different devices and browsers, leading to potential performance inconsistencies. Additionally, the browser exhibited significantly slower reload times, further compromising usability and responsiveness. Another limitation involved the retrieval of video data from the user's camera.

Due to privacy constraints, WebGazer.js accesses a processed video stream without providing the actual camera resolution, making it impossible to assess the impact of webcam quality on tracking performance.

Chapter 6:
Conclusions and Future Work

6.1. Revisiting the research questions

Advances in webcam-based, eye tracking, such as WebGazer.js, enable online studies as an alternative to traditional lab setups eye tracking has become a widely used tool for exploring how researchers interpret and understand eye-tracking data. Traditionally, most eye tracking studies rely on controlled lab settings with high-end equipment, but conducting these experiments remotely using everyday devices could make it easier to involve more participants and expand research efforts.

To explore this possibility, a webcam eye tracking experiment with 28 participants was conducted who engaged in map reading tasks (i.e., memorability of map landmarks) through a custom web application that runs in the Google Chrome browser. WebGazer.js, an open-source camera-based eye tracking software, was used in this study, paired with standard consumer webcams. This shift from lab settings to online settings sets the foundation and the motivation of this thesis:

Can webcam-based eye tracking effectively replicate and scale the results of lab-based eye tracking studies in the context of map reading? How far can webcam-based eye tracking mirror lab-based behavioral insights in map reading?

To explore this, a subset of tasks and stimuli of Keskin's (2020) lab-based study performed using SMI RED250 eye tracking software (the dataset is openly available under CartoGAZE (Keskin, 2020)) was replicated in a webcam eye tracking experiment using the WebGazer.js open-source tool. Due to the nature of the experiment design considerations of webcam eye tracking studies (i.e., shorter experiment durations, lack of participant control and potential calibration or connection issues), our study focuses only on three tasks recognition of roads, hydrography, and the combination of roads and hydrography as these tasks include the most attention-grabbing features which acted as memory anchors (Keskin et al., 2023).

This allowed partial overlap with Keskin's (2020) Blocks 7 (Roads), 6 (Hydrography), and 2 (Roads and Hydrography) and also helped shorten the experiment time to avoid fatigue and drop-outs. Despite WebGazer.js providing accessibility and scalability, the results of this study highlight its limitations in precision and accuracy compared to lab-based systems. These insights directly inform our research objectives and questions:

Research Objective 1: Evaluating webcam eye tracking software solutions

RQ 1: *What are the best free webcam eye tracking tools for map reading?*

A comparative evaluation of three free webcam-based eye-tracking tools WebGazer.js, GazeRecorder, and RealEye highlights WebGazer.js as the most suitable open-source solution for map-reading studies. Its primary strengths lie in its scalability and accessibility, making it particularly advantageous for large-scale, remote research. To support this assessment, a comparison table (see Chapter 3, Table 3.1) was developed, drawing on a preliminary analysis adapted from Vaníček et al. (2025). This comparison highlighted the features and limitations of WebGazer.js, GazeRecorder, and RealEye. WebGazer.js offers a web browser based, real-time gaze tracking system that does not require participants to download any software, a key advantage over many other tools. This made it particularly useful when conducting an online, large-scale experiment with 35 participants, of which 28 were retained after filtering for data quality (Chapter 3, Section 3.5).
Additionally, WebGazer.js's open-source nature and lack of subscription costs make it an ideal choice for researchers with limited resources, supporting the goal of democratizing cartographic usability studies.
In contrast, GazeRecorder offers a desktop solution with slightly higher accuracy but at the cost of scalability since it has to be installed, thus being less suitable for remote, large-scale research. Meanwhile, RealEye offers built in calibration and cloud-based data processing but operates on a subscription model, which may restrict access for budget-constrained researchers (Chapter 3, Table 3.1). Despite its limitations, WebGazer.js strikes a balance between ease of use and flexibility, making it the preferred choice for our study.
Since WebGazer.js tracks gaze points by predicting the participant's gaze position on the screen based on sampled head and eye positions recorded via webcam (relative to stimulus displayed), it offers an effective solution to the task of tracking gaze in online settings (Steffan et al., 2023).It provides consistent gaze path capturing and coarse eye tracking across the entire map, not linked to specific landmarks or Area of Interest (AoI).This flexibility permitted us to calculate significant eye tracking metrics, such as average fixation duration, fixation count, and average saccade length (Chapter 4), by writing custom code to process gaze data and calculate these measures based on gaze coordinates and time.
Despite its strengths, WebGazer.js has notable drawbacks. Its performance is sensitive to external variables like lighting, participant movement, and camera quality, which can

undermine data reliability. Compared to GazeRecorder and RealEye, which offer superior spatial precision, WebGazer.js provides moderate accuracy, which is not suitable for such a large-scale Area of Interest (AoI) analysis.
To compensate for its lack, very large AoIs must be defined which undermines the precision and specificity of the analysis (Thilderkvist & Dobslaw, 2024). Despite this limitation, its usability and flexibility make it a valuable tool for our purposes particularly for exploratory map reading research where broad gaze patterns, rather than pinpoint accuracy, are the focus.

RQ 2: *How accurate and practical are the tools for map reading tasks?*

Despite the accessibility and cost effectiveness of webcam-based eye tracking systems like WebGazer.js, there are notable limitations when compared to laboratory based systems such as the SMI RED250 their accuracy and practicality for map reading tasks where participants focus on specific features like place names, roads, or symbols are limited by factors such as sample rate, gaze estimation accuracy, and suitability for Area of Interest (AoI) analysis. According to the results of this study, most participants using WebGazer achieved sampling rates between 15 and 20Hz, underscoring its limitations in real-world conditions. Tasks such as map reading which involve rapid saccades between landmarks and smooth pursuits along routes require higher sampling rates to accurately capture the timing and dynamics of eye movements. WebGazer's sampling rate is limited by typical webcam frame rates (30–60 Hz, or 16–33 ms per sample) (Papoutsaki et al., 2016), falling well below that of lab-based systems like the SMI RED250, which operates at 250 Hz (4 ms per sample). The SMI RED250's higher temporal resolution enables more precise tracking of dynamic eye movements.
In contrast, WebGazer's lower effective rate, especially in the observed 15–20 Hz range, introduces noticeable lag. This temporal lag may lead to the omission of brief fixations or the smoothing of transitions between map elements, thereby diminishing the system's effectiveness for tasks requiring real-time monitoring or fine-grained temporal analysis. WebGazer's gaze estimation accuracy is reported at approximately 3 degrees of visual angle (around 50 mm or 100-150 pixels on a standard screen), which is significantly lower than the SMI RED250's 0.4-0.5 degrees (6-8 mm or 15-20 pixels). Studies, such as Papoutsaki et al. (2016), indicate that WebGazer's error ranges from 100 to 200 pixels under optimal conditions, whereas lab-based systems like the SMI RED250 consistently achieve sub degree precision Holmqvist et al.(2011).

WebGazer.js employs a self-calibrating Ridge Regression model that leverages user interactions, such as mouse clicks, to improve accuracy over time. However, its performance remains sensitive to factors like lighting, head movement, and webcam quality. In contrast, the SMI RED250 uses infrared illumination and high fidelity sensors to maintain accuracy even with moderate head motion. For map reading tasks, WebGazer.js can reliably distinguish gaze between broad regions (e.g., map quadrants); however, its larger error margin limits precision when it comes to smaller features, such as closely spaced labels or icons areas where the SMI RED250 performs significantly better.

Area of Interest (AoI) analysis, which is essential for examining attention to specific map elements, highlights WebGazer's limitations. With a reported gaze accuracy of approximately 3 degrees, this translates to a positional uncertainty of 100–150 pixels (Papoutsaki et al., 2016), making it unreliable for identifying gaze on small or densely packed AoIs such as adjacent street names or symbols spaced less than 50 mm apart.

In this study, an average eye tracking accuracy of 73% was achieved using WebGazer.js, a reasonable outcome for a webcam-based system operating in uncontrolled environments (Papoutsaki et al., 2016).However, this level of accuracy was still insufficient for precise AoI identification. In contrast, the SMI RED250, with its 0.4 – 0.5-degree precision, supports fine-grained AoI analysis, accurately mapping gaze to targets as small as 15–20 pixels (Holmqvist et al., 2011).

While WebGazer's coarser spatial resolution may suffice for broader analyses such as tracking attention to the northern versus southern half of a map its performance is further limited by environmental factors, including ambient light variation and the use of lower resolution webcams (typically 720p or 1080p), as noted by Xu et al. (2015).

Lab-based systems like the SMI RED250 mitigate these issues through controlled conditions and high precision hardware, resulting in superior data quality. WebGazer.js is practical and sufficiently accurate for map reading tasks involving broad gaze patterns, such as identifying regions of interest, but its lower sample rate, reduced gaze estimation accuracy, and limited Area of Interest (AoI) precision make it less effective compared to the SMI RED250. For detailed map analysis tracking exact fixations or small Area of Interest (AoI) lab, based systems remain the gold standard, while WebGazer.js serves as a viable, budget friendly option when precision is secondary to accessibility.

Research objective 2: Comparing webcam-based eye tracking for online studies and lab-based eye trackers.

RQ 3: *How do user performance and data quality differ between the two methods?*

By comparing two studies between webcam-based eye tracking (WebGazer.js) and lab-based eye tracking Keskin's (2020) study, we can see significant differences in how participants perform and how reliable the data is when they interact with 2D static maps across tasks like roads, hydrography, and a combination of both. For performance, in this study participants were observed to be slower and less accurate completing the tasks compared to participants in Keskin's (2020) study who used, eye tracking in the lab setting.

The participants in Keskin's (2020) study likely had better resolution eye tracking; this can help track small details that participants look for during fixation on the map. For instance, in the road task while WebGazer.js participants were slower and less accurate during the road task by comparing lab study, they completed the hydrography task and the combined task faster. This indicates that perhaps novice participants with web based, eye tracking are in a more naturalistic environment and they may value speed over precision during these tasks. Success rates in this comparison illustrate the performance gap; lab-based participants had better success rates than their webcam counterparts on all tasks. This performance difference can be explained by an ability to track fixation and saccade location and duration more accurately in a lab setting. However, several factors, such as the design of the tasks, randomization of task order, and greater expertise or familiarity with map reading tasks, may also influence the results of, eye tracking. These factors likely contribute to the better performance observed in the lab-based setting. In terms of data quality, the lab-based eye tracking approach from Keskin's (2020) study captured clearer and more reliable saccade rates and patterns of fixation. Their approach even captured saccades occurring frequently in more tasks that are complex.

The WebGazer.js webcam-based system yielded noisier, less granular results and were particularly poor at capturing rapid saccades. This could be explained due to limited sampling and the environment lighting conditions or accuracy of calibration step when the webcam and the eye tracker software were used or participant's posture during the session. Although WebGazer.js offered briefer fixations per user, the brief fixations may not be reliable for tracking task focus compared to the accuracy and reliability of lab-based eye tracking suggesting the possibility for fast but unreliable attention tracking in lab-based eye tracking methods.

Overall, it is evident that lab-based eye tracking methods provide better accuracy and reliability when tracking novices or beginning learners about of eye fixations in more detail. However, with a controlled participant environment is it possible to obtain eye-tracking data comparable to lab-based eye tracking data

RQ 4: *Can webcam-based, eye tracking replicate lab-based results, especially for online studies, particularly in evaluating novice users' interactions with 2D static maps?*

In addressing the research question of whether webcam-based, eye tracking can replicate lab-based results, particularly in evaluating novice users' interactions with 2D static maps, it is crucial to consider findings from previous studies. Keskin's (2020) study Experiment 2 involved a complex map learning task with multiple stimuli and fixed study times, closely mirroring our own design.

The comparison between this study and Keskin's (2020) lab-based results demonstrates that webcam-based, eye tracking can effectively replicate key attentional metrics, such as fixation duration and saccade length, while capturing user interactions with 2D static maps. Although there are some differences particularly in fixation rates and saccade lengths the general consistency in trends and patterns supports the validity of using webcam-based methods in online environments. The minor discrepancies observed could be attributed to differences in experimental context, task design, and user familiarity with the map stimuli. Nevertheless, the ability of webcam-based tracking to reproduce comparable visual behavior patterns suggests that it is a feasible and practical alternative for studying attentional behavior, especially in large-scale or remote experiments. This study showed slightly shorter fixation durations compared to Keskin's study (2020) across all tasks, indicating faster information processing, especially in the Hydrography task. While Keskin's study (2020) recorded a higher number of fixations per second in the Roads task, reflecting more frequent visual transitions, additionally, this study showed a higher fixation rate during the Hydrography task, indicating a more dynamic visual scanning behavior. In the combined Roads and Hydrography task, fixation frequencies were similar, with Keskin's study (2020) slightly higher. Regarding saccade length, our study recorded longer saccades in the Roads task, indicating broader visual jumps, while Keskin's study (2020) showed slightly longer saccades in the Hydrography and combined tasks, suggesting a more detailed scanning pattern in the lab-based environment. (Chapter 5). Keskin et al.'s (2023) findings suggest that harder tasks constrain eye movements, and Keskin's (2020) lab-based study emphasized the role of user expertise in shaping visual behavior, whereas the present results demonstrate that the nature of map landmarks (map features) may instead encourage broader scanning patterns. Despite some differences, the overall trends between the two studies were comparable, further supporting the feasibility of webcam-based tracking for evaluating attentional behavior in online environments.

Research objective 3: Identifying limitations, challenges, and developing best practices

RQ 5: *What methodological and technical challenges arise when using webcam-based eye tracking how do factors like environmental variables and device differences affect data quality?*

One major methodological challenge in this study was adapting an experimental task from Keskin's study (2020) Experiment 2, which originally lasted approximately 2.5 hours. Running such a long experiment online was impractical due to participant fatigue and a high risk of dropouts. To address this, the task was refined into a 20-minute version that retained the core elements of the original study. This modification balanced participant engagement while maintaining data quality and validity.

A significant methodological concern was the variability in raw data quality from webcam-based, eye tracking. Data precision is heavily influenced by factors such as camera resolution and processing power. Low-resolution cameras (e.g., 720p or lower) struggle with accurately detecting pupil and eye reflections, particularly in fluctuating lighting conditions. Similarly, weak processing power can cause frame rate inconsistencies, resulting in gaze data inaccuracies. These issues complicate the interpretation of Area of Interest (AoI) and contribute to offset fixation errors (Chapter 5, Section 5.3). Environmental variables further introduced challenges. Unlike controlled lab settings, online environments vary significantly in lighting, camera angles, and background distractions. Such inconsistencies negatively affect the reliability of gaze tracking and create difficulties in standardizing data collection conditions. Another technical obstacle was managing device variability among participants. As participants used their own hardware, disparities in webcam resolution, frame rates, and processing power directly affected tracking accuracy. Privacy constraints prevented direct access to camera specifications, so screen sizes were recorded as an indirect measure. Additional complications arose from individual differences, including participant positioning, eyewear, headwear, and varying device specifications, all of which affected calibration and reduced data precision.

Controlling participant head movement was particularly challenging. To mitigate inconsistent gaze estimates, a real-time visual feedback system was implemented that alerted participants when their eyes moved outside a designated validation box. Despite this, tracking accuracy remained inconsistent due to device limitations and uncontrolled participant movement. Additionally, browser compatibility issues, particularly with Firefox, introduced further inconsistencies due to system lag and privacy restrictions. While webcam-based eye tracking provides accessibility advantages, it has notable limitations that must be addressed in future research.

RQ 6: *What best practices and recommendations can improve to best design a map reading experiment with webcam eye tracking?*

The results indicate that eye tracking using a webcam lacks sufficient accuracy for studies that examine fine detail in gaze such as fixating on individual words or specific lines of a map. It is appropriate for lower accuracy gaze studies only if the area of interest is at least 105 pixels wide (Thilderkvist & Dobslaw, 2024). In addition, for data quality to be the best possible, it is recommended that Participants should remain as still as possible, because even slight head movement can affect how accurately their gaze is tracked. Lastly, even under the best-case scenario, webcam-based eye tracking approaches do not have established procedures for systematic validation of accurate and reliable data.
If researchers are looking to enhance the quality of their map reading experiments using webcam-based, eye tracking, it is essential to consider the following best practices:

Do's:

- Provide participants with clear and structured instructions before the experiment to ensure they understand setup requirements and procedures.
- Incorporate a short pre experiment training to help participants familiarize themselves with the eye tracking software, reducing errors during data collection.
- Design maps using strong with a clear visual structure and use strong contrast size, and placement of critical elements to guide participant focus.
- Keep maps simple to avoid distractions and make it easier for the webcam to track gaze accurately.
- Optimize map layouts by ensuring focal areas are well defined and appropriately sized to accommodate webcam-tracking limitations.

- Use a timer to keep tasks short and focused, helping participants stay engaged and reducing fatigue.
- Test experiments across multiple browsers and devices to identify and address potential technical problems before deployment.
- Use real-time feedback systems to help participants maintain proper positioning during the experiment.
- Calibrate the eye tracker for each participant before every task to improve accuracy, even if they already had a basic calibration.
- Ask participants to use a plain background behind them to reduce visual noise that might confuse the eye tracking software.
- Collect feedback from participants after the experiment to learn about any setup or usability issues.
- Run a small pilot study first to test the map design and eye tracking setup before the full experiment.

Don'ts:

- Avoid using webcam-based eye tracking for high precision studies, as it lacks the resolution and stability required for detailed gaze tracking because the accuracy of webcam eye tracking is good enough to track larger areas of map not details of map
- Avoid long experimental sessions; participant fatigue leads to higher dropout rates and unreliable data.
- Do not rely on automated tracking adjustments alone to compensate for head movements; encourage participants to maintain a stable position instead.
- Avoid overloading maps with excessive visual details, as too much complexity can introduce noise and reduce gaze-tracking accuracy.
- Do not skip checking the webcam quality Low-resolution cameras can worsen data reliability.

As of now, the quality of webcam-based, eye tracking differs widely. Based on current scientific evidence, running remote eye tracking studies without strict controls is not recommended, even with predefined rules for participants' equipment and surroundings. However, by paying close attention to certain details, the data quality and reliability of webcam-based eye tracking experiments can be improved. Although this method presents challenges, careful planning and smart design can enhance the usefulness of online eye tracking studies.

6.2. Future work

This study highlighted several ways to improve webcam-based eye tracking, particularly with tools like WebGazer.js, to make it more useful for research in areas like map design and user interaction studies in cartography.

The current study adapted Keskin's (2020) original 2.5-hour lab-based experiment into a 20-minute online session to maintain participant engagement and data quality in an unsupervised setting. This shortening posed a limitation, and future studies could aim to replicate the full 2.5-hour experiment in a controlled lab environment using webcam-based eye tracking technology. The study used data from the CartoGAZE dataset (https://doi.org/10.7910/DVN/ONIAZI), which includes 2D static screenshots of Google road/navigational maps (2018), as part of replicating Keskin's (2020) lab-based study. However, webcam-based eye tracking presented challenges, as it struggled to detect fine details like pupil and corneal reflections, which are essential for accurate tracking. Compared to traditional lab-based eye tracking systems, webcam-based tracking has lower spatial accuracy, making it unreliable for analyzing small Areas of Interest (AoIs). Consequently, researchers should use larger AoIs to obtain meaningful gaze data, at the cost of reduced analytical granularity. Larger AoIs also exacerbate gaze point inaccuracies, making tools like WebGazer.js better suited for studying general browsing patterns and attentional focus rather than tasks requiring high precision. However, webcam eye tracking can be particularly effective in scenarios involving dashboard style maps or interfaces, where elements are relatively large and spatially distinct. In such contexts, the method allows for meaningful insights into users' attentional distribution and interaction behaviors without requiring pinpoint gaze precision.

A significant challenge encountered in this study was the effect of head movement on data quality. Even small shifts in a participant's position caused inconsistent gaze estimates, showing the need for better ways to ensure correct eye positioning. We added JavaScript code to WebGazer.js to monitor eye alignment in real-time, providing visual feedback and alerts when participants' eyes moved outside a validation box. However, the system struggled to handle head movements consistently, and the feedback did not fully stabilize tracking over time. Future work could address this by using advanced head pose estimation algorithms or adaptive recalibration methods that adjust to movement automatically. Adding clearer feedback, such as brighter visual cues or sound alerts, might also help participants stay in position and improve data quality. Another issue we noticed was a consistent misalignment in WebGazer's gaze data.

The recorded gaze coordinates often did not match the intended targets. This could be due to problems with the initial calibration, limitations in WebGazer's prediction model, or factors like screen resolution and webcam placement. Future research could focus on identifying the exact cause of this misalignment. For example, testing calibration under different lighting conditions, distances, or hardware setups could reveal what affects accuracy. Improving WebGazer's algorithms perhaps by training them on more varied data, adding real-time corrections for head tilt and distance, or repeating calibration during the experiment could reduce this problem and make the tool more reliable for tasks needing precision. WebGazer.js is a helpful tool for studying cognitive processes in cartographic tasks, but it has limitations that affect data quality. Factors like changing lighting (e.g., glare or dimness) can hide the pupil and lower accuracy, especially with poorly placed webcams, such as those below laptop screens. Low sampling rates, differences in hardware, and inconsistent calibration also create challenges. Participants often find it hard to follow calibration steps correctly, leading to errors in gaze detection.

These issues reduce the tool's reliability for detailed studies needing precise fixation times and eye movement measurements. To address these problems, future work could improve calibration with adaptive algorithms that adjust for head position and lighting in real-time. Clearer instructions for participants such as tips for stable lighting, fewer distractions, and proper camera placement could also help. Better processing power and camera quality might narrow the gap between webcam-based and infrared-based systems. While most eye tracking research uses expensive equipment in controlled labs, studies in cognitive cartography suggest that simple gaze data collected remotely with webcams can be similar to lab results, even from unsupervised participants. Future work should prioritize improving tracking accuracy by enhancing head movement correction, diagnosing and resolving deviation of gaze data in WebGazer.js, and optimizing the quality of visual stimuli. These improvements could help bridge the gap between webcam-based and lab-based systems, enabling more effective and scalable eye tracking research in areas such as map design and user experience studies.

Data availability

The experiment and the code of this study are available in the link below:

https://github.com/rahgh/WebcamET_CartoGAZE-data-set-.git

-References

Chen, K. T., Prouzeau, A., Langmead, J., Whitelock-Jones, R. T., Lawrence, L., Dwyer, T., & Goodwin, S. (2023). Gazealytics: A unified and flexible visual toolkit for exploratory and comparative gaze analysis. *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, 1–7. https://arxiv.org/abs/2303.17202

Çöltekin, A., Lokka, I., & Fabrikant, S. I. (2020). Visual attention and recognition differences based on expertise in a map reading and memorability study. *ISPRS International Journal of Geo-Information, 9*(2), 112. https://doi.org/10.3390/ijgi9020112

Edler, D., Bestgen, A. K., Kuchinke, L., & Dickmann, F. (2015). The processing and integration of map elements during a recognition memory task is mirrored in eye movement patterns. *Cartography and Geographic Information Science, 42*(4), 339–350. https://doi.org/10.1080/15230406.2015.1035428

Edler, D., Keil, J., & Dickmann, F. (2020). Effects of visual map complexity on the attentional processing of landmarks. *Journal of Eye Movement Research, 13*(1). https://doi.org/10.16910/jemr.13.1.4

Hassoumi, A., Peysakhovich, V., & Hurter, C. (2019). Improving eye tracking calibration accuracy using symbolic regression. *PLoS ONE, 14*(3), e0213675. https://doi.org/10.1371/journal.pone.0213675

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Keskin, M. (2020). Exploring the cognitive processes of map users employing eye tracking and EEG [Doctoral dissertation, Ghent University]. https://biblio.ugent.be/publication/8684280

Keskin, M., Krassanakis, V., & Çöltekin, A. (2023). Visual attention and recognition differences based on expertise in a map reading and memorability study. *ISPRS International Journal of Geo-Information, 12*(1), 21. https://doi.org/10.3390/ijgi12010021

Kiefer, P., Giannopoulos, I., Raubal, M., & Duchowski, A. T. (2017). Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation, 17*(1–2), 1–19. https://doi.org/10.1080/13875868.2016.1254634

Krassanakis, V., & Cybulski, P. (2021). Eye tracking in cartography: A review of recent advances and applications. *Cartography and Geographic Information Science, 48*(3), 123–135. https://doi.org/10.1080/15230406.2021.1893759

Krassanakis, V., & Cybulski, P. (2021). Eye tracking research in cartography: Looking into the future. *ISPRS International Journal of Geo-Information, 10*(6), 411. https://doi.org/10.3390/ijgi10060411

Kvålseth, T. O. (2017). Coefficient of variation: The best single measure of dispersion? *The Journal of Applied Statistics, 44*(3), 502–515. https://doi.org/10.1080/02664763.2016.1174195

Negi, A., & Mitra, S. (2020). Learning analytics and eye tracking: An exploratory study of fixation duration distribution in online learning environments. *Smart Learning Environments, 7*(1), 1–15. https://doi.org/10.1186/s40561-020-00128-5

Niehorster, D. C., Zemblys, R., Beelders, T., & Holmqvist, K. (2020). Characterizing webcam-based eye tracking for remote usability studies. *Behavior Research Methods, 52*(4), 1476–1492. https://doi.org/10.3758/s13428-019-01336-7

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, 3839–3845.

Robal, T., Zhao, Y., Lofi, C., & Hauff, C. (2018). Webcam-based attention tracking in online learning: A feasibility study. *Proceedings of the 23rd International Conference on Intelligent User Interfaces (IUI '18)*. https://doi.org/10.1145/3172944.3172987

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

Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2023). Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering, 16*, 260–277.

Slim, H., Courtemanche, F., & Dufresne, A. (2024). Webcam-based versus infrared eye tracking in educational interfaces: A comparative study of usability and fixation detection accuracy. *Behavior Research Methods*. https://pubmed.ncbi.nlm.nih.gov/39654819/

Slim, M., et al. (2024). Webcams as windows to the mind? A direct comparison between in-lab and web-based eye tracking methods. *Open Mind, 8*, 1369–1424. https://doi.org/10.1162/opmi_a_00171

Steffan, A., Bryck, R. L., Baker, R., & Mills, C. (2023). Validation of an open-source, remote web-based eye tracking method (WebGazer) for research in early childhood. *Infancy, 29*(1), 31–55. https://doi.org/10.1111/infa.12564

Thilderkvist, E., & Dobslaw, F. (2024). On current limitations of online eye tracking to study the visual processing of source code. *Information and Software Technology, 174*, 107502. https://doi.org/10.1016/j.infsof.2024.107502

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited: An overview of some recommended measures of effect size. *Trends in Sport Sciences, 21*(1), 19–25.

Vaníček, T., Gharibpour, R., Keskin, M., & Popelka, S. (2025). Exploring the possibilities and limitations of webcam-based eye tracking for interactive and static map studies: A comparative perspective on WebGazer and RealEye. *Manuscript submitted for publication*.

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the Visual World Paradigm. *Glossa Psycholinguistics, 6*(1), 131. https://doi.org/10.5070/G6011131

Wisiecka, K., Krejtz, K., Krejtz, I., Sromek, D., Cellary, A., Lewandowska, B., & Duchowski, A. T. (2022). Comparison of webcam and remote eye tracking. *Proceedings of the 2022 Symposium on Eye Tracking Research and Applications (ETRA '22)*. https://doi.org/10.1145/3517031.3529615

Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015). TurkerGaze: Crowdsourcing eye tracking. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI)*, 1893–1902. https://doi.org/10.1145/2702123.2702367

Yang, B., & Li, H. (2021). A visual attention model based on eye tracking in 3D scene maps. *ISPRS International Journal of Geo-Information, 10*(10), 664. https://doi.org/10.3390/ijgi10100664

Yang, X., & Krajbich, I. (2023). Webcam-based online eye tracking for behavioral research. *Behavior and Brain Sciences, 46*, e152. https://doi.org/10.1017/S1930297500008512