Proceedings of the Design Society, Volume 5: ICED25
https://doi.org/10.1017/pds.2025.10099

Can I catch up later? Design of personalized intervention for online learning using eye-tracking-based video reconstruction and replay

Chunzhi Li and Ting Liao✉

Department of Systems and Enterprises, Stevens Institute of Technology, New Jersey, US

✉ tliao@stevens.edu

ABSTRACT: While online learning allows learners to access materials flexibly and at their own pace, many struggle to self-regulate without supervision. Real-time interventions like pop-out quizzes, screen flashes, and text warnings aim to improve attention focus but risk distracting learners and segmenting the learning process. Despite eye-tracking technology being widely used for real-time intervention design, its potential for delayed and personalized interventions remains underexplored. To address this gap, we proposed and tested an eye-tracking-based video reconstruction and replay (EVRR) method, offering targeted review at the end of online classes without disrupting the learning process. EVRR shows significant positive effects on improving learning outcomes compared to self-paced reviews, especially for learners who are unfamiliar with the concepts.

KEYWORDS: education, user-centred design, design methods, eye-tracking, mind-wandering

1. Introduction

Online learning has long been an essential part of education since the popularization of computers. Especially over the past decade, with the high demands of learning, pre-recorded classes that can be taken at learners' own pace bloomed due to their flexibility and convenience (Castro & Tumibay, 2021). Despite the advantages of online learning, there remain specific challenges in maintaining effective learning outcomes (Schacter & Szpunar, 2015). Online learning is typically regarded as self-regulated learning because it requires learners to regulate their attention independently. Without the supervision of teachers, learners usually find it hard to stay focused and allocate their attention appropriately (Jansen et al., 2020). While watching pre-recorded classes, learners might pause the video for a long time or just let the video play, leading to missing or misunderstanding key concepts or disproportionately concentrating on certain parts of the material, resulting in knowledge gaps.

Particularly, shifting attention from the task at hand to self-generated, irrelevant thoughts refers to mind-wandering (MW), a prevalent phenomenon during online learning (Zhang et al., 2020). MW can occupy up to 45% of online learning, raising the significance of understanding and reducing MW (Risko et al., 2012; Kane et al., 2017; Szpunar et al., 2013).

To address the challenges of MW and its impact on learning outcomes in online education, it is essential to enhance MW detection and intervention methods. Recent advances in eye-tracking technology have provided researchers with an invaluable tool to investigate MW through physiological metrics such as pupil dilation, fixation duration, and gaze dispersion. These metrics offer critical insights into learners' attention allocation and cognitive engagement, forming the foundation for designing effective interventions.

Many researchers have used eye trackers to understand how learners collect visual information via eye movements and provide an advanced view of humans' cognitive processes, such as where they looked at, how long, and in what order (Carter et al., 2020). The research is built based on the principle of the eye-mind link, where people have a high motivation to move their eyes to focus on the stimulus they are currently thinking about or processing (Rayner, 2009). The potential eye movement and pupillometry metrics include fixation duration, saccade amplitude, pupil diameter, etc. (Mahanama et al., 2022). These metrics can be used to reveal cognitive levels, for example, pupil dilation increases when people are concentrated on demanding tasks (Skaramagkas et al., 2021). On top of these fundamental metrics, eye-tracking data has been further analyzed for the Area of Interest (AOI) designation, heatmap analysis, and scan path analysis.

Using eye-tracking data, existing studies have explored strategies for redirecting learner attention, including real-time intervention, such as pop-out quizzes, text recall, oral reading, and text or sound alerts, (McMaster et al., 2015; Han et al., 2022), and AI-based systems like Avatar learning companion and Nao robot (Lee et al., 2022; Blancas et al., 2018). However, there are limitations in their ability to personalize interventions and minimize disruptions to the learning process. To address this gap, we propose a novel framework with delayed intervention by utilizing insights from MW detection metrics and existing intervention strategies. This framework involves an eye-tracking-based video reconstruction and replay (EVRR) system. It uses eye data to reconstruct learning material by highlighting knowledge that learners missed or misunderstood or MW and replay to them. The detailed system design and development are described in Section 4.1.

To examine if the system could effectively guide learners' attention and improve learning outcomes, we conducted a human-subject experiment regarding engineering concepts.

2. Related Works

2.1. Mind-wandering Metrics

MW is typically assessed through eye movement metrics, such as pupil dilation, fixation duration, and gaze patterns. Previous research has shown that pupil dilation is associated with the brain's locus coeruleus-norepinephrine (LC-NE) system, which is vital for attention and cognitive arousal (Eckstein et al., 2017). Furthermore, mean pupil dilation positively correlates with the cognitive workload, indicating comprehension difficulty (Skaramagkas et al., 2021).

Zhang et al. (2020) examined the correlation between MW and eye movement patterns during video lectures. They found that MW is typically associated with longer fixations on the slides, because participants may process the information on the slides more slowly or not actively. Additionally, MW could reduce fixation dispersion on the slides, indicating that participants focused their attention on a smaller area of the slides, possibly reflecting a less active engagement with the content (Jang et al., 2020; Zhang et al., 2020; Krasich et al., 2018).

2.2. MW Intervention Methods

Numerous intervention methods have been developed to address MW and redirect learner attention. The most common types of intervention are pop-out quizzes, screen flashes, and text or sound alerts. Based on traditional interventions, some researchers implemented message reminders based on their developed AI learning systems to help learners better learn and stay focused (Hutt et al., 2021; Lee et al., 2022). Moreover, considering the future of virtual learning environments, virtual teachers' eye contact with learners has also become a novel real-time intervention method (Han et al., 2022). Despite the growing number of real-time interventions, false MW detections can still disrupt learners' (Lee et al., 2022). Distracted learners might also ignore interventions, especially if they were not motivated to obey (Arakawa et al., 2021).

Many tools, for example, Eye-Mind Reader, have been designed to help learners comprehend and address the occasionally inaccurate MW detection Mills et al. (2021). It implemented an intelligent reading interface, which uses two primary intervention techniques, re-reading prompts, and self-explanation exercises, to mitigate MW in comprehension. A nonlinear probabilistic approach is used to decide when to intervene. When the system detects reduced engagement, such as prolonged fixation durations or inconsistent reading pace, the system encourages learners to revisit specific content sections, reinforcing critical information. The self-explanation technique prompts the reader to express their understanding about the text in their own words, encouraging active engagement.

However, interventions such as self-explanation exercises can significantly increase their learning time, decreasing learning efficiency. Additionally, individual differences in eye movement patterns and the lack of personalization further limit the effectiveness of these interventions.

3. Hypotheses

In this study, we investigated if the EVRR effectively enhances learning outcomes, which were assessed by two quizzes on the learning material. Participants in the experiment group were exposed to an EVRR-based intervention by watching a reconstructed video. In contrast, those in the control group were prompted to watch the learning material at their own pace as a comparison. First, we compared the learning outcomes between the experiment and control groups to test the EVRR's effectiveness. We further investigated whether the improvement was based on the reconstruction and replay of the material enabled by the EVRR method. Therefore, we proposed the following two sequential hypotheses.

- H1: EVRR leads to a higher positive score difference than the self-review.
- H2: Replaying more AOIs that are related to incorrect quiz questions leads to greater score improvement.

4. Methods

4.1. Experiment Design

We designed a human-subject experiment to investigate the effectiveness of the EVRR system with the approval of the Institutional Review Board. Participants were randomly divided into experiment and control groups, where the experiment group utilized the EVRR.

4.1.1. Procedure

This experiment consists of four steps. Step 1: Participants were asked to complete a questionnaire to collect their demographic information and self-report their familiarity with seven computer networking topics in the learning material. Step 2: Participants watched a 12-minute video while their eye data were collected and took a quiz after the video. Step 3: Following a 5-minute break, participants entered the review phase, after which they retook the quiz *without* knowing their previous scores or which answers were correct. Step 4: Participants completed an experience survey.

During the review phase in Step 3, participants in the experiment group watched a reconstructed video generated based on their eye data. The process of video reconstruction is explained in detail in Section 4.2.2. Participants in the control group reviewed the original video. They were allowed to control the playback using the progress bar, allowing them to navigate to any part of the learning material during the review phase.

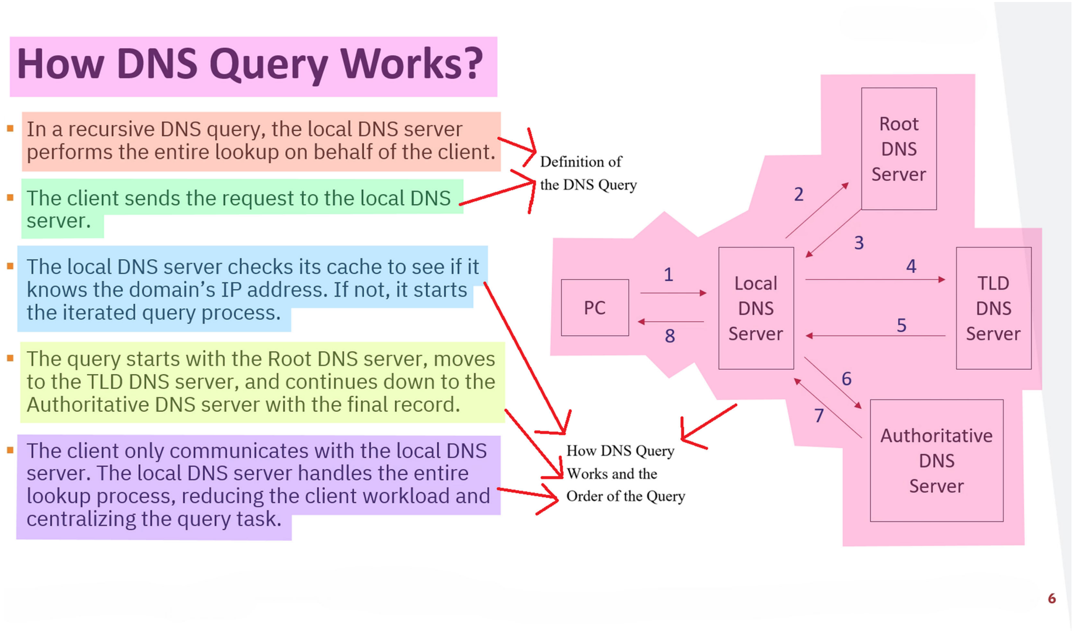
4.1.2. Experiment Setup

We used Tobii Pro Fusion to collect eye data, using its maximum sampling frequency of 250 Hz (Tobii, n.d.). To ensure the accuracy and reliability of the data collection, we asked participants to face the screen with a resolution of 1920 × 1080 pixels and maintain a distance of about 65 cm from the screen when watching the learning video. Further, we used the Tobii Pro Lab to apply nine-point calibration and pre-process participants' eye movements, such as noise-cleaning and clean data exporting.

4.1.3. Learning Materials and Quiz

The learning video was obtained by slide recording. The slides cover the fundamental concepts of computer networking, including IP Address, Domain Name System (DNS), HTTP request, Transmission Control Protocol (TCP), Three-way Handshake (TWH), Forward Proxy, Reverse Proxy, Cache, and Cookies. These concepts about the computer network were chosen from Kurose et al. (2021) to ensure that the task was not too difficult for participants with no prior knowledge and was not common sense for college students. Different paragraphs in each slide are divided into different AOIs based on the specific concept points described.

Figure 1 shows an example of an AOI design in which all AOIs are highlighted in different colors. Certain AOIs are grouped as associated AOIs to explain technical concepts or processes. For instance, the first two paragraphs in Figure 1 are grouped because they introduce the DNS queries. Similarly, the last three paragraphs and the figure are associated AOIs, explaining how DNS queries work and the order of the queries.



How DNS Query Works?

- In a recursive DNS query, the local DNS server performs the entire lookup on behalf of the client.
- The client sends the request to the local DNS server.
- The local DNS server checks its cache to see if it knows the domain's IP address. If not, it starts the iterated query process.
- The query starts with the Root DNS server, moves to the TLD DNS server, and continues down to the Authoritative DNS server with the final record.
- The client only communicates with the local DNS server. The local DNS server handles the entire lookup process, reducing the client workload and centralizing the query task.

Definition of the DNS Query

How DNS Query Works and the Order of the Query

Figure 1. AOI Design Example

To determine when to proceed to the next slide, we calculated the Estimated Reading Time (ERT) for each AOI across all pages. The calculation is derived from the formula outlined by Brysbaert & Marc (2019), which calculated the predicted reading speed with different average word lengths in different texts during silent reading. The average word length of non-fiction texts is 4.6, and the average reading rate is 238 words per minute (wpm). For the learning material and the material-based quiz, a 130 wpm recall reading rate is adopted from the literature Carver et al. (2019). The ERT calculation is shown in Equation (1).

As shown in Figure 1, the AOI "03 DNSQ D1" contains 20 words, totaling 87 characters (excluding spaces). Consequently, the average word length is 4.4. Using Equation (1), an estimated reading time of 8.8 seconds is calculated for this AOI. The ERT for an entire page is calculated from the sum of all AOIs reading times.

$$ERT = \frac{W \cdot L \cdot f}{R \cdot 4.6} \tag{1}$$

where W indicates the *total word count* in the AOI, while L represents the *average word length.* The reading rate R is set at 130 wpm based on recall reading speed and f is a *conversion factor of 60*, used to transform the rate from minutes to seconds.

The learning outcome quiz used to quantify participants' learning outcomes consisted of 15 questions. Each question was worth 5 points for a total score of 75 points. The answers to all the questions can be found in the corresponding slides. We acknowledged that participants might have various levels of knowledge regarding the topics. To control for its influence, we asked participants to report their familiarity level with each topic on a five-point scale, from "Not at all familiar," "Slightly familiar," "Somewhat familiar," "Moderately familiar," to "Extremely familiar."

4.2. EVRR Design

The EVRR system design consists of an Eye (E) movement data threshold design and a Video Reconstruction and Replay (VRR) design.

We collected participants' eye movement data while participants engaged with the learning material in Step 2, focusing on average fixation duration (AFD) and average pupil diameter (APD) during the entire video, average pupil diameter during the AOI ($APD_{average}$), MW time (MWT), total fixation time (TFT), and valid reading time (VRT).

4.2.1. E-Eye Movement Data Threshold Design

To identify MW during learning, it is essential to determine appropriate fixation duration thresholds based on prior reading and visual attention research.

Rayner (2009) demonstrated that fixation durations can range from 50-75 ms to 500-600 ms, depending on task complexity. Particularly, the average fixation duration ranged from 225-250 ms for silent reading. Following this, Trabulsi et al. (2021) adopted a minimum fixation duration of 60 ms in their research on reading optimization. In contrast, Hooge et al. (2022) suggested that fixations shorter than 100 ms are unlikely to reflect meaningful visual processing, underscoring the variability in fixation thresholds across studies. In addition, 2000 ms is a commonly used upper boundary for attention-related fixations during video lectures Zhang et al. (2020); Cornelissen & Võ (2017). However, we adopted a more strict threshold of 600 ms based on the prior study from Rayner (2009), because fixation longer than 600 ms may indicate disengagement or unrelated cognitive processes, particularly in video-based learning contexts.

In summary, we adopted 250 ms as the benchmark AFD for reading. To address the significant individual differences, we set two adaptive thresholds, 100 * (AFD / 250) as the minimum threshold and 600 * (AFD / 250) as the maximum threshold. To further validate the thresholds, we conducted a pilot test of eight participants and found that the average fixation time of participants watching the video ranged from 187 ms to 320 ms (250 ± 70). To refine the minimum threshold, we tested 0.4 * AFD, 0.5 * AFD, 0.6* AFD, comparing MW detection results with self-reported MW. 0.5 * AFD yielded the best results. Thus, we used 0.5 * AFD as the minimum threshold and 600 * (AFD / 250) as the maximum threshold for MW detection. Fixations shorter than the minimum or longer than the maximum threshold will be considered MW and added to MWT. Subsequently, the VRT for each AOI can be calculated in Equation (2), depending on the presence or absence of MWT.

$$VRT = \begin{cases} \text{TFT} - \text{MWT}, & \text{if MWoccurs}, \\ \text{TFT}, & \text{otherwise}. \end{cases} \tag{2}$$

We integrated pupil sizes to enhance MW detection validity because the mean pupil dilation positively correlates with the cognitive workload, indicating comprehension difficulty (Skaramagkas et al., 2021). The $\text{APD}_{\text{average}}$ of all fixations on each AOI is calculated to determine whether MW occurred on certain AOI. Given our strict MW detection thresholds, we set an initial threshold 35% of TFT instead of 45%, which was suggested by Zhang et al. (2020). Also, we compared 25%, 30%, 40%, and 45% in the pilot test to determine when the MW should be considered harmful to the learner's learning. We concluded that an AOI required intervention if MWT exceeded 30% of the TFT and its APDaverage was larger than its APD.

By analyzing the relationship between the proportion of VRT in ERT and missing AOIs, MW AOIs, and misunderstanding AOIs in the pilot test, we set 30% * (AFD / 250) of ERT as the missed AOI threshold (T_{missed}), as most participants had little to no recall of these AOIs. 80% * (AFD / 250) of ERT was used as the MW detection threshold (T_{MW}). Additionally, 120% * (AFD / 250) of ERT was identified as the misunderstood AOI threshold (Tmisunderstood), as exceeding this threshold indicated comprehension difficulty, often leading participants to misunderstand the AOI and reduced the reading time of other AOIs on the same page. In this case, we re-evaluated the AOI associated with the misunderstood AOI and adjusted the threshold to 50% * (AFD / 250) of ERT as the misunderstood associated threshold ($\text{T}_{\text{associated}}$). Using the aforementioned threshold values, Figure 2 shows the flow of detecting MW and determining if an AOI will be highlighted and replayed.

4.2.2. VRR-Video Reconstruction and Replay Design

Guided attention was used in our study to ensure learners focused on replayed AOIs. Figure 3 shows two examples of reconstructed pages. AOIs that learners have fully learned are mosaiced in Figure 3(a). AOIs that learners need to relearn are highlighted in the red line, and related AOIs are generally displayed without intervention. In Figure 3(b), the diagram is associated with the definition, and bullet points are associated with each other. So, no AOI is mosaiced on this page.

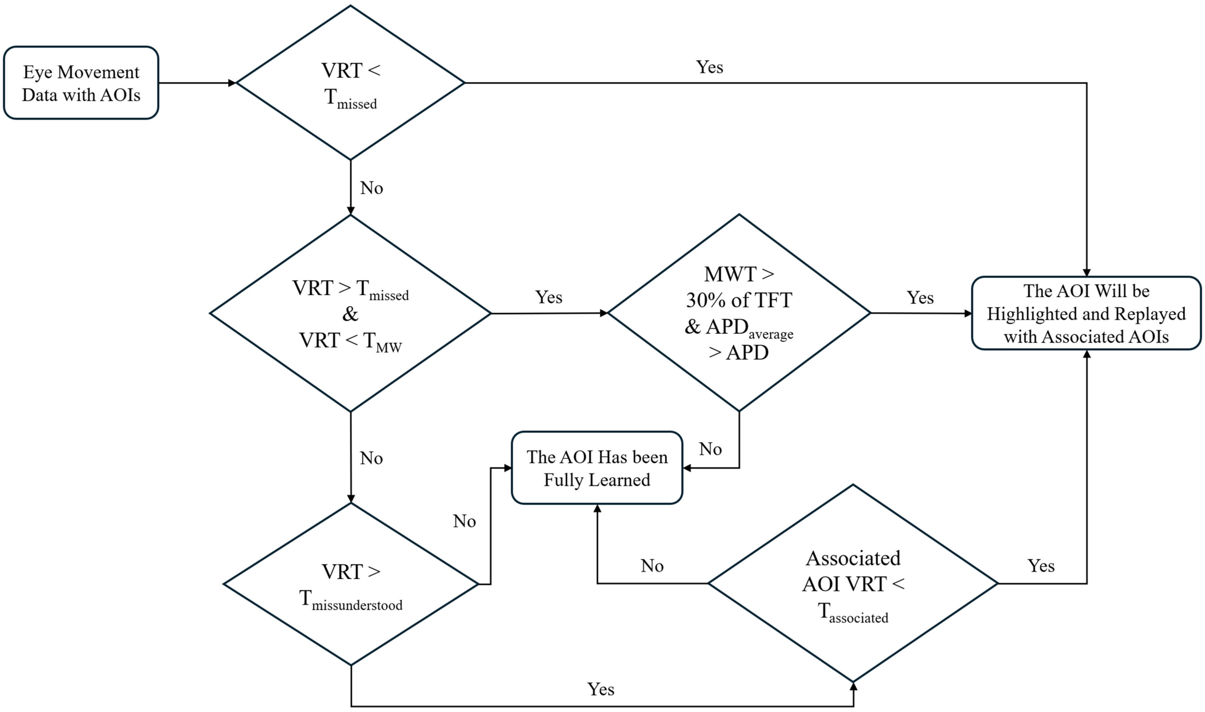
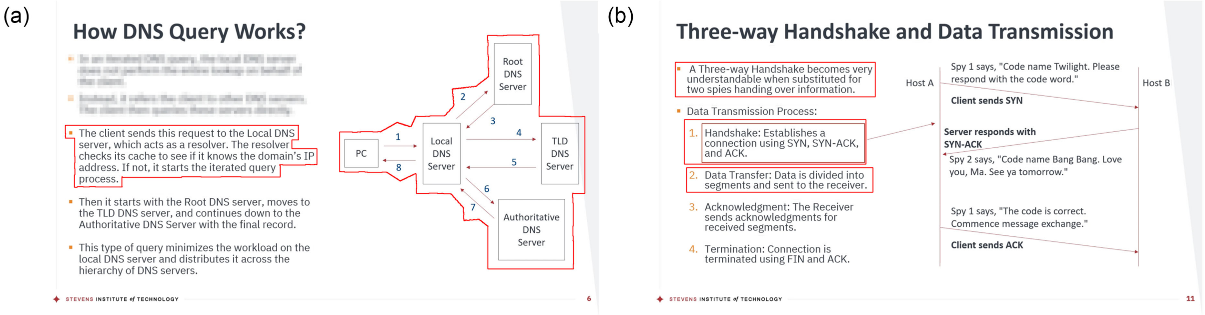

Eye Movement Data with AOIs
VRT < T_{missed}
Yes
No
VRT > T_{missed} & VRT < T_{MW}
Yes
MWT > 30% of TFT & $APD_{average}$ > APD
Yes
The AOI Will be Highlighted and Replayed with Associated AOIs
No
The AOI Has been Fully Learned
No
No
VRT > $T_{missunderstood}$
No
Associated AOI VRT < $T_{associated}$
Yes
Yes

Figure 2. MW and Knowledge Gap Detection Flow



How DNS Query Works?

- [illegible]
- [illegible]
- The client sends this request to the Local DNS server, which acts as a resolver. The resolver checks its cache to see if it knows the domain's IP address. If not, it starts the iterated query process.
- Then it starts with the Root DNS server, moves to the TLD DNS server, and continues down to the Authoritative DNS Server with the final record.
- This type of query minimizes the workload on the local DNS server and distributes it across the hierarchy of DNS servers.

Three-way Handshake and Data Transmission

- A Three-way Handshake becomes very understandable when substituted for two spies handing over information.
- Data Transmission Process:
  1. Handshake: Establishes a connection using SYN, SYN-ACK, and ACK.
  2. Data Transfer: Data is divided into segments and sent to the receiver.
  3. Acknowledgment: The Receiver sends acknowledgments for received segments.
  4. Termination: Connection is terminated using FIN and ACK.

Spy 1 says, "Code name Twilight. Please respond with the code word."

Client sends SYN

Server responds with SYN-ACK

Spy 2 says, "Code name Bang Bang. Love you, Ma. See ya tomorrow."

Spy 1 says, "The code is correct. Commence message exchange."

Client sends ACK

Figure 3. VRR Examples. (a) Replayed AOIs with associated AOIs, (b) All AOIs are associated with Replayed AOIs

5. Analysis and Results

5.1. Participant Demographics and Knowledge Familiarity Distribution

Forty-one participants were recruited for our study. Participants' eye data were collected, and no video recording was stored. The demographics and English proficiency data are shown in Table 1. Considering the small sample size, Near Native and Native Speakers were combined into the high Proficiency group, and Proficient and Fluent were combined into the medium Proficiency group. Overall, the high Proficiency group has 19 participants, and the medium Proficiency group has 22 participants.

The weighted familiarity accounts for the varying occurrences of concepts across quiz questions. Each concept's familiarity score is multiplied by its number of occurrences in the quiz. The weighted scores are then summed and divided by the total number of quiz questions to standardize the result. The average familiarity of all participants with all concepts and their weighted familiarity in the experiment and

Table 1. Demographic Distribution

Group	Age Distribution				Gender Distribution				English Proficiency			
	1825	2634	3554	5564	Male	Female	Non Binary	No say	Nat. Speaker	Near Nat.	Fluent	Prof.
Experimental	14	4	2	0	13	6	0	1	6	1	12	1
Control	13	6	1	1	15	5	1	0	9	3	8	1

control groups are shown in Figure 4(a). Figure 4(b) shows the number of samples in the experimental and the control groups at different levels of familiarity.

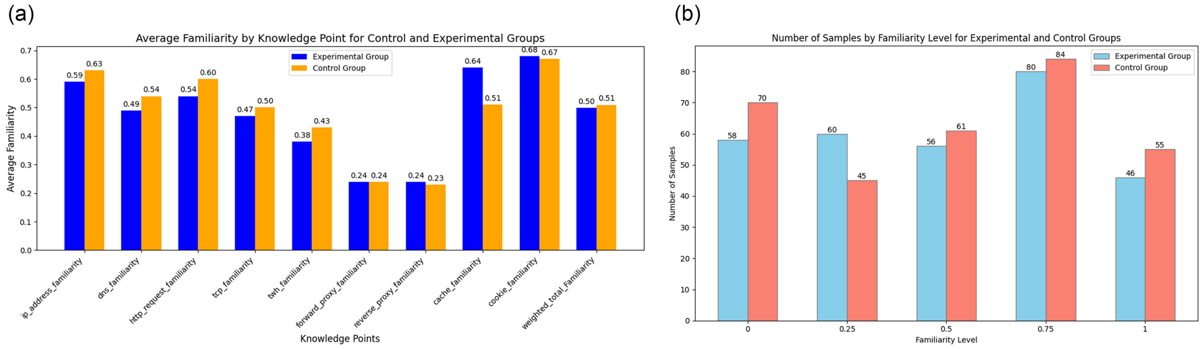


Figure 4. Participants Average Familiarity. (a) Average Familiarity by Knowledge Point for Control and Experimental Groups, (b) Number of Samples by Familiarity Level for Experimental and Control Groups

5.2. Hypotheses Test

The mixed-effect model was utilized in this part to examine the relationships between the following variables while controlling for individual differences. The replay represented the hit rate for the correct option for the AOIs highlighted in the reconstructed video. The higher the value is, the more likely the participant was to get more questions right in the second quiz. Since only the experience group watched the reconstructed video, it is used to examine the effects relationships among the experiment group. In the model, DS_{ij} represents the **score difference** for the jth participant in the ith group, calculated as the difference between the second and first quiz scores to measure performance improvement after the revision. G_i is a categorical variable indicating group membership. GP_i denotes **grouped English** proficiency, as detailed in Section 5.1. Participants' **nervousness level** while using the eye tracker is denoted by N_{ij}, while L_{ij} represents their **perceived usefulness** of the eye tracker in improving learning performance. The variable R_{ij} refers to the **hit rate** of the AOI corresponding to the correct answer for a previously incorrect question, which is highlighted for review.

The model includes **random effects** *(u_j)* to account for participant-level variations not captured by the fixed effects, assuming a normal distribution ($u_j \backslash N(0, \sigma_u^2)$). The **residual error term** (ε_{ij}) represents unexplained variability in the score difference, also assumed to follow a normal distribution ($\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$). The terms ($\beta$, θ, and γ represent the fixed-effect **coefficients** in the models, corresponding to the intercept and various predictors, including group membership, proficiency, nervousness, likelihood, replay, weighted familiarity, and their interaction effects.

To test H1, we controlled for differences in weighted familiarity by dividing participants into low, medium, and high familiarity groups using the quantile method. A mixed-effect model was built to examine factors influencing score differences, as shown in Equation (3).

The results indicated that high English proficiency significantly improves scores for low and medium-familiarity learners, while medium-English proficiency has a more significant positive effect for high-familiarity learners. The results suggested that when familiarity is low, high proficiency aids understanding, whereas when familiarity is high, those with medium proficiency may compensate by revising more carefully. Since eye tracker nervousness and likelihood were not significant, they were removed in Equation (4). The revised model confirms that the experimental group significantly outperforms the control group among low-familiarity learners, demonstrating the effectiveness of the EVRR method. High English proficiency also continues to benefit low-familiarity learners. The p-values are provided in Table 2. Thus, *H1 is partially supported.*

$$DS_{ij} = \beta_0 + \beta_1 G_i + \beta_2 GP_i + \beta_3 N_{ij} + \beta_4 L_{ij} + u_j + \varepsilon_{ij} \tag{3}$$

$$DS_{ij} = \beta_0 + \beta_1 G_i + \beta_2 GP_i + u_j + \varepsilon_{ij} \tag{4}$$

Table 2. P-values for Various Mixed-Effect Models

Effect	Equ(3)-L	Equ(3)-M	Equ(3)-H	Equ(4)-L	Equ(4)-M	Equ(4)-H
Group (Exp vs. Control)	n.s.	n.s.	n.s.	0.040	n.s.	n.s.
Grouped English Proficiency	0.023	0.016	0.039	0.005	n.s.	n.s.
Eye Tracker Likelihood	n.s.	n.s.	n.s.	—	—	—
Eye Tracker Nervousness	n.s.	n.s.	n.s.	—	—	—

Note: n.s. indicates a non-significant result ($p > 0.05$); — indicates that the effect was not included in the model.

Regarding H2, we analyzed the experimental group to test if *replaying more AOIs that are related to incorrect quiz questions leads to a greater score improvement.* The results of the full model shown in Equation (5) confirmed that a higher replay hit rate significantly enhances scores, while eye tracker nervousness negatively impacts scores. Additionally, when replay and weighted familiarity increase, their interaction significantly reduces scores. The p-values are summarized in Table 3. Since replay represents guided attention, a higher replay rate increases focused learning through EVRR. Thus, *H2 is supported.*

$$DS_{ij} = \theta_0 + \theta_1 R_{ij} + \theta_2 WF_{ij} + \theta_3 (R_{ij} \cdot WF_{ij}) + \theta_4 GP_i + \theta_5 N_{ij} + \theta_6 L_{ij} + u_j + \varepsilon_{ij} \tag{5}$$

Since eye-tracking data was only collected in the first session, we further explored why eye-tracker nervousness negatively impacts scores by analyzing participants' first quiz scores in the model described by Equation (6). The results reveal that higher weighted familiarity improves scores, whereas higher eye tracker likelihood reduces them. Then, we speculated that there was an interaction between English proficiency, eye tracker likelihood, and eye tracker nervousness. We used the model in Equation (7). Compared to high proficiency, medium proficiency had a significant negative effect on the first test score, i.e., lower English proficiency led to lower learning outcomes, which is consistent with common sense. Furthermore, higher weighted familiarity improves scores, while eye tracker likelihood and nervousness negatively affect first quiz performance. However, when English proficiency and nervousness increase, first scores significantly improve, suggesting that moderate anxiety, combined with higher proficiency, enhances focus and understanding, as shown in Table 3.

$$FS_{ij} = \gamma_0 + \gamma_1 GP_i + \gamma_2 WF_{ij} + \gamma_3 N_{ij} + \gamma_4 L_{ij} + u_j + \varepsilon_{ij} \tag{6}$$

$$FS_{ij} = \gamma_0 + \gamma_1 GP_i + \gamma_2 WF_{ij} + \gamma_3 N_{ij} + \gamma_4 L_{ij} + \gamma_5 (GP_i \cdot N_{ij}) + \gamma_6 (GP_i \cdot L_{ij}) + u_j + \varepsilon_{ij} \tag{7}$$

6. Conclusion

We proposed a new personalized and delayed intervention, EVRR, that adapts to individual differences but does not interfere with the learner's learning process. EVRR shows significant positive effects on improving learning outcomes compared to self-review for learners who are unfamiliar with the concepts.

Table 3. P-values for Various Mixed-Effect Models

Effect	Equ(5)	Equ(6)	Equ(7)
Grouped English Proficiency	n.s.	n.s.	n.s.
Replay	0.012	—	—
Weighted Familiarity	n.s.	0.007	0.002
Replay * Weighted Familiarity	0.018	—	—
Eye Tracker Likelihood	n.s.	0.023	0.020
Eye Tracker Nervousness	0.050	n.s.	0.000
Proficiency * Likelihood	—	—	n.s.
Proficiency * Nervousness	—	—	0.002

***n.s.** indicates a non-significant result (p > 0.05); *— indicates that the effect was not included in the model.

While they have difficulty effectively acquiring and absorbing knowledge during the initial learning process, EVRR's personalized intervention compensates for this by prompting them to review content they have not mastered. Additionally, increased replay of AOIs related to participants' incorrect quiz questions resulted in a higher positive score difference, further verifying guided attention's effectiveness. Our analysis also revealed that for learners with medium English proficiency, giving them moderate nervousness during learning can help them concentrate better and improve their learning performance. For learners with high English proiciency, the nervousness in the learning process should be minimized as much as possible to help them concentrate. These findings highlight the importance of tailoring interventions based on individual characteristics. The EVRR intervention is based on each individual's eye movement data to carry out learning interventions and, therefore, has robust scalability in personalized education. It can also be applied to live classes, especially online classes that require the camera to be turned on to interact with the teacher.

While the likelihood of using an eye tracker negatively impacted initial scores, its interaction with other factors was not signficant, making it challenging to speculate this indirectly as the trust of the eye tracker or the reconstructed video warrants further exploration. EVRR's use of eye movement data offers robust scalability in personalized education and can be applied to live or online classes requiring teacher interaction. Future research should explore the role of eye tracker likelihood and refine EVRR for broader applications.

References

Arakawa, R., & Yakura, H. (2021). Mindless Attractor: A False-Positive Resistant Intervention for Drawing Attention Using Auditory Perturbation. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3411764.3445339

Blancas-Muñoz, M., Vouloutsi, V., Zucca, R., Mura, A., & Verschure, P. F. M. J. (2018). Hints vs Distractions in Intelligent Tutoring Systems: Looking for the proper type of help. *Interaction Design and Children (IDC-CRI2018) Workshop* https://doi.org/10.48550/arXiv.1806.07806

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, 109, 104047. https://doi.org/10.1016/j.jml.2019.104047

Carver, R. P. (1992). Reading Rate: Theory, Research, and Practical Implications. *Journal of Reading*, 2(2), 84–95.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, 155, 49–62.

Castro, M. D. B., & Tumibay, G. M. (2021). A literature review: Efficacy of online learning courses for higher education institution using meta-analysis. *Education and Information Technologies*, 26(2), 1367–1385. https://doi.org/10.1007/s10639-019-10027-z

Cornelissen, T. H. W., & Võ, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics*, 79(1), 154–168. https://doi.org/10.3758/s13414-016-1203-7

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eye-tracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, 25, 69–91. https://doi.org/10.1016Zi.dcn.2016.11.001

Han, Y., Miao, Y., Lu, J., Guo, M., & Xiao, Y. (2022). Exploring intervention strategies for distracted students in VR classrooms. *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–7. https://doi.org/10.1145/3491101.3519627

Hooge, I. T. C., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2022). Fixation classification: How to merge and select fixation candidates. *Behavior Research Methods*, 54(6), 2765–2776. https://doi.org/10.3758/s13428-021-01723-1

Hutt, S., Krasich, K., Brockmole, J. R., & D"Mello, S. K. (2021). Breaking out of the lab: Mitigating mind-wandering with gaze-based attention-aware technology in classrooms. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3411764.3445269

Jang, D., Yang, I., & Kim, S. (2020). Detecting mind-wandering from eye movement and oculomotor data during learning video lecture. *Education Sciences*, 10(3), Article 3. https://doi.org/10.3390/educsci10030051

Jansen, R. S., van Leeuwen, A., Janssen, J., Conijn, R., & Kester, L. (2020). Supporting learners" self-regulated learning in massive open online courses. *Computers & Education*, 146, 103771. https://doi.org/10.1016/j.compedu.2019.103771

Kane, M. J., Smeekens, B. A., von Bastian, C. C., Lurquin, J. H., Carruth, N. P., & Miyake, A. (2017). A combined experimental and individual-differences investigation into mind-wandering during a video lecture. *Journal of Experimental Psychology: General*, 146(11), 1649–1674. https://doi.org/10.1037/xge0000362

Krasich, K., McManus, R., Hutt, S., Faber, M., D”Mello, S. K., & Brockmole, J. R. (2018). Gaze-based signatures of mind wandering during real-world scene processing. *Journal of Experimental Psychology: General*, 147(8), 1111–1124. https://doi.org/10.1037/xge0000411

Kurose, J. F., & Ross, K. W. (2021). *Computer networking: a top-down approach* (Eighth edition).

Pearson. Lee, T., Kim, D., Park, S., Kim, D., & Lee, S.-J. (2022). Predicting mind-wandering with facial videos in online lectures. *CVPRW2022*, 2104–2113. https://doi.org/10.1109/CVPR52729.2023.02322

Lee, W., Allessio, D., Rebelsky, W., *et al.* (2022). Measurements and interventions to improve student engagement through facial expression recognition. In R. A. Sottilare & J. Schwarz (Eds.), *Adaptive Instructional Systems*, 286–301. Springer International Publishing. https://doi.org/10.1007/978-3-031-05887-5_20

Mahanama, B., Jayawardana, Y., Rengarajan, S., Jayawardena, G., Chukoskie, L., Snider, J., & Jayarathna, S. (2022). Eye movement and pupil measures: A review. *Frontiers in Computer Science*, 3. https://doi.org/10.3389/fcomp.2021.733531

McMaster, K. L., van den Broek, P., Espin, C. A., Pinto, V., Janda, B., Lam, E., Hsu, H.-C., Jung, P.-G., Leinen, A. B., & van Boekel, M. (2015). Developing a reading comprehension intervention: Translating cognitive theory to educational practice. *Contemporary Educational Psychology*, 40, 28–40. https://doi.org/10.1016Zi.cedpsych.2014.04.001

Mills, C., Gregg, J., Bixler, R., & D”Mello, S. K. (2021). Eye-Mind reader: An intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering. *Human-Computer Interaction*, 36(4), 306–332. https://doi.org/10.1080/07370024.2020.1716762

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday attention: Variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology*, 26(2), 234–242. https://doi.org/10.1002/acp.1814

Schacter, D. L., & Szpunar, K. K. (2015). Enhancing attention and memory during video-recorded lectures. *Scholarship of Teaching and Learning in Psychology*, 1(1), 60–71. https://doi.org/10.1037/stl0000011

Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2021). Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering*, 16, 260–277.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, 110(16), 6313–6317. https://doi.org/10.1073/pnas.1221764110

Tobii. (n.d.). (2024). *Reach further with your research | Tobii Pro Fusion - Tobii. Tobii Pro Fusion: Specifications.* https://www.tobii.com/products/eye-trackers/screen-based/tobii-pro-fusion#specifications

Trabulsi, J., Norouzi, K., Suurmets, S., Storm, M., & Rams0y, T. Z. (2021). Optimizing fixation filters for eye-tracking on small screens. *Frontiers in Neuroscience*, 15. https://doi.org/10.3389/fnins.2021.578439

Zhang, H., Miller, K. F., Sun, X., & Cortina, K. S. (2020). Wandering eyes: Eye movements during mind wandering in video lectures. *Applied Cognitive Psychology*, 34(2), 449–464. https://doi.org/10.1002/acp.3632