The Eye Movement Pattern and Brain Dynamics indicate successful learning during education video viewing

Yizhi Wei* [1] (1155191602@link.cuhk.edu.hk) & Eric Rui-Wen Wang* [1,2] (eric6@uchicago.edu)
1. Chinese University of Hong Kong. 2. University of Chicago

Abstract

Online education has become a prominent feature of modern learning. However, what determines learning success in online education remains unclear. To explore this, the present study applies Hidden Markov Model to eye-tracking and EEG data in an open dataset containing educational video-viewing task. The model revealed two eye movement patterns: a global pattern where fixations were distributed across the video, and a local pattern where fixations focused on task-relevant features. The model also revealed 6 distinct brain states involving attention and memory. Experiments 1 and 2 explore the impact of task engagement on eye movement. We found that compared to intentional learning condition, more participants use a global pattern and suffer from worse memory performance in incidental learning condition. Experiment 3 explores the impact of video style. We found that only in the presenter-and-animation style, the local eye movement patterns and learning brain states are associated with learning success. Overall, this study explores factors that indicate learning success and offers insights for improving learning efficiency in online education.

Keywords: Hidden Markov Model, Eye-tracking, EEG, Online Education, Memory, Learning

Introduction

Online education has become a prominent feature of modern learning given its convenience and accessibility. For example, the rise of MOOC has significantly promoted the global dissemination of knowledge, breaking down traditional barriers to higher education (Alhazzani 2020). For many areas that lack educational resources, it serves as a powerful alternative or supplement to the traditional classroom-based learning, offering students access to world-class resources. Recent progress in AI has catalyzed this trend, allowing anyone to generate and edit content within minutes and upload to video platforms like YouTube. This has greatly democratized access to education. However, while the doors to education are now wider than ever, the cognitive and neural mechanisms behind online learning remain unclear. Specifically, it is not yet clear what kinds of neuro-cognitive patterns enable successful learning, and how test expectancy, like whether learners are explicitly aware of an upcoming assessment, and varying styles of the learning materials, interact with these neuro-cognitive profiles to shape effective learning.

One way to think about this question is through the lens of attention. Indeed, previous research has shown that attention and memory are closely related, as it helps determine what will be remembered and encoded into our memory system, which is critical for successful learning (Cowan et al., 2024). Studies based on Intersubject Correlation (ISC) have shown that co-fluctuation in eye movement can predict test scores in online education video viewing (Gu et al., 2024; Madsen et al., 2021). Apart from capturing attention from eye movements, neuroimaging studies have also shown that the brain networks are stable neuromarkers for sustained attention and can reflect cognitive performance, including recall in reading tasks (Jangraw et al., 2018; Song et al., 2023). Overall, these works pointed out that successful learning depends on optimizing attention strategies, and the neuro-cognitive pattern can be extracted from eye-tracking and neuroimaging data.

However, two critical gaps remain due to the current research methods. Firstly, while the ISC focuses on the co-fluctuation of attention across individuals, it fails to capture what exactly the content in the video is attracting attention. This has undermined the interpretability of attentional dynamics, as instead of identifying specific attention strategies, it only shows whether there is synchronization or not. Secondly, most existing neuroimaging studies employ fMRI to capture attention during video viewing; however, since attention dynamics are fast and happen on an instant timescale (Fiebelkorn & Kastner 2019;), devices with higher temporal resolution, such as EEG, are required to capture more nuanced brain responses. In this study, we address these gaps by applying Hidden Markov Model (HMM) to multimodal data (Hsiao et al., 2021A; Vidaurre et al., 2018). To address the first gap, we use HMM to directly model the latent states underlying the observed fixation in eye-tracking data. This allows us to extract highly interpretable regions of interest and investigate the relationship with the features in the video, thereby answering what is being attended to. To address the second gap, we utilise Canonical HMM-based model to decode source-reconstructed EEG data, capturing the fast transients of brain states at a fine-grained temporal scale (Gohil et al., 2025).

In this study, we analyzed an open dataset and extracted eye movement pattern and brain states using HMM, and studied how these patterns are associated with their learning outcomes measured by memory performance. In experiments 1 and 2, we further examined how eye movements are modulated by test expectancy (intentional vs incidental) [1]. In experiment 3, we further examined how eye movement and

[1] Since there is no EEG data in Experiment 1 in the original dataset, we cannot run analysis to investigate the impact of test expectancy on brain states.

brain states are modulated by the styles of the learning materials.

Design and Method

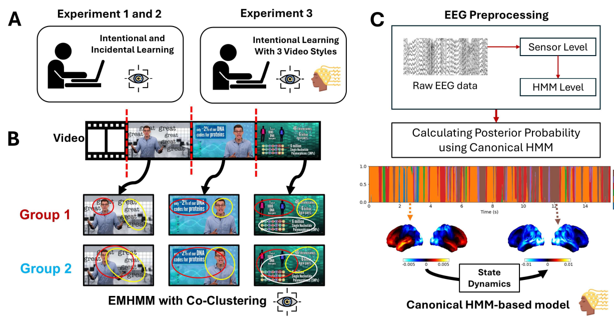


Figure 1A) displaying 3 experiments in the BBBD dataset. B) displaying the procedure to analyze eye-tracking data. C) displaying the procedure to analyze EEG data.

Experiment Design

We analyzed an open dataset that consists of educational video viewing task called the Brain, Body and Behavior Dataset. This dataset consists of multimodal recording including eye-tracking and EEG data. The eye-tracking data were recorded using SR Eyelink 1000 with 500 Hz sampling rate, and the EEG data were recorded using BioSemi ActiveTwo system with a 64-electrode cap (10/10 system) with ground electrodes near Poz at 2048 Hz. For more information, refer to the Madsen et al (2025).

For current analysis pipeline, we used only the session one of the first three experiments. In short, the participants were instructed to watch a series of education videos. Their domain knowledge was assessed before exposed to the video, and their learning outcomes were assessed after watching the video through four-alternative forced-choice questions. As shown in **Fig.1A**, Experiment 1 was conducted under intentional learning condition, where the participants were informed in advance that they would be tested. Experiment 2 was conducted under incidental learning condition, where the participants were not. We analyzed the eye-tracking data in the hope of finding the impact of expectation to be tested on learning outcomes, and how it would reflect as eye movement pattern[2]. Experiment 3 was conducted under intentional learning condition with three video styles: presenter and animation (Style1); animation with writing hand (Style2); presenter with glassboard (Style3). Ratings for enjoyable was recorded. We analyzed both the eye-tracking and EEG data to investigate the impact of styles on eye movement patterns and brain activities, and its relation with learning success.

Eye-Movement Analysis

We used the Eye Movement Analysis with Hidden Markov Model (EMHMM) with co-clustering approach to analyze eye-tracking data (Hsiao et al., 2021A). This method has been successfully applied to a wide range of studies, including reading (Wang et al., 2025), face recognition (Hsiao et al., 2021B), and music reading (Liao et al., 2022). But unlike previous studies where participants look at static images, the feature layout of a video constantly changes as time unfolds. Therefore, as shown in **Fig.1B**, we decided to segment the video following the principle that keeps features within the segments as consistent as possible. Then, we applied EMHMM onto the eye-tracking data in each of these segments. This approach involved two steps. In the first step, a Hidden Markov Model (HMM) was fitted to individual's fixations within each video segments. In each HMM, the hidden states were defined as regions of interest (ROI), formulated as a 2D Gaussian distribution, and transitions between each hidden state (ROIs) follow a Markov process, which means that the probability of being in current hidden state depended on the previous hidden state. Therefore, the individual differences in both spatial and temporal dimensions of eye movements (gaze locations and transitions) were captured. Here, we selected the range of the number of hidden states in each individual model to be from 1 to 3 and used variational hierarchical expectation maximization (VHEM) algorithm to estimate the optimal number of hidden states for each model (Coviello et al., 2014).

In the second step, a co-clustering method was applied to all segments to discover group-level attention patterns. Since the feature layout is different across segments, co-clustering is particularly suitable in discovering participant groups sharing common eye movement patterns across video segments as it considers individual participants' eye movement patterns across all stimuli when performing clustering. The formulation of co-clustering ensures that the participant groups discovered are consistent across all stimuli. Following previous studies, we clustered all HMMs into two representative pattern groups, and within each group, each segment had a representative HMM summarizing eye movement patterns from the participants that belong to the group (Wang et al., 2025; Liao et al., 2024). For each representative passage model, the number of hidden states (ROIs) was set to be the median number of hidden states in the individual segments' HMMs. All models (including the individual HMMs) mentioned above were trained 300 times with different initial Gaussian ROIs, and models with the highest log-likelihood of the data were selected.

Finally, we quantify the similarity between each participant and the representative HMMs using data log-likelihoods. We define a scale such that $(L_1 - L_2) / (|L_1| + |L_2|)$, where L_1 and L_2 are the likelihood that the given fixation data of one participant was generated from either group 1 or 2. So that along the scale, a more positive score indicates higher similarity to group 1, and vice versa. In this study, we discovered that group 1 consisted of participants that had local, focused and non-overlapping ROI in their HMM, while group 2 consisted of participants that had global, dispersed

[2] Due to poor data quality and missing eye fixation data, we removed subject 11,13,22,26,31 in experiment 1 (after removal N=22). And Subject 06, 16,18,23,28,30,35,36 in experiment 2 (after removal N=24).

and overlapping ROI in their HMM. So we also refer to this scale as the Local-Global scale, or L-G scale.

It's worth noting that, the purpose of clustering is to **NOT** to hard parcel subjects, instead, we adapt a soft assignment by comparing their relevant distance to either of pattern (L-G Scale). So even there are some small clusters, it is still meaningful. Moreover, previous HMM study also found similar small clusters (Liao & Hsiao 2025; Zheng et al 2022).

EEG Analysis

Similarly, we used HMM to analyze EEG data (**Fig.1C**). Here, the hidden states are defined as brain states, formulated as N-dimensional Gaussian distribution (where N is the number of parcellation in source-reconstructed space), and the transitions between each hidden state (brain states) follow a Markov process.

The EEG data were preprocessed at sensor level using high pass filter at 0.05 Hz and notch-filtered at 60 Hz to remove line noise. Robust PCA was used to remove artificial signal followed by low-pass filter at 64 Hz. The data were then down-sampled to 128 Hz. Eye movement related artifacts were removed using least-squares noise cancellation via linear regression of EOG signals. Bad channels were replaced via linear interpolation using valid samples from neighboring electrodes within a 40 ms temporal window. Finally, the EEG data were re-referenced by subtracting the average across all sensors excluding those marked as bad.

Based on previous HMM studies on electrophysiology data (Toffoli et al., 2024; Makkinayeri et al., 2025), we co-register structural MRI (using standard MNI template) with EEG data using FSL BET (Smith 2002). Then, we model the electrical conductivity using the inner skull, outer skull and outer head skin in our forward model. We calculated the leadfields for a dipole placed in x, y and z-direction on 8mm isotropic grid inside the volume bound by the inner skull. We used the linearly constrained minimum variance (LCMV) beamforming algorithm to estimate source space activity (Tait et al., 2021). To parcel the source-reconstructed data, we used a 38 ROI parcellation (Gohil et al., 2025). Since source reconstruction suffers from spatial leakage, we correct this using symmetric orthogonalisation proposed by Colclough et al (2015) that removes all zero-lag correlations from the parcel data. Since after beamforming, the sign of the dipoles is arbitrarily assigned, we therefore performed dipole sign flipping described in Vidaurre et al (2016) to correct for the sign ambiguity across the group.

Instead of training a new HMM for the current EEG dataset, we directly applied the canonical HMM solution proposed by Gohil et al. (2025) and calculated the posterior probability on our dataset. This decision was motivated by three key factors: First, training a new HMM on EEG is computationally expensive and can lead to less reliable results, especially with limited sample sizes. Second, the canonical HMM was trained on a high-quality, large-scale MEG dataset (N=621, Cam-CAN) across a wide age range (18–88 years), providing the model with greater stability and generalizability. Third, EEG data typically exhibit higher noise levels, which can reduce the precision of state inference; utilizing MEG-derived brain states helps mitigate this issue. **Importantly,** prior work has demonstrated that network identified by MEG can successfully transfer to EEG data (for further details, see discussion in Gohil et al., 2025).

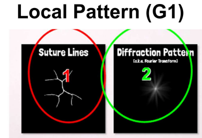
To ensure compatibility with the canonical HMM for inference, we further prepare our data following the exact pipeline used in the canonical HMM. Specifically, we applied time-delay embedding (TDE) with ± 7 lags and performed PCA at the group level to reduce the dimensionality of the TDE data. The data were then temporally standardized (i.e., de-meaned and scaled by the standard deviation). This approach was adopted to mitigate model overfitting and to focus on the low-frequency components of the signal, consistent with the 1/f characteristic of electrophysiological data, where the majority of power and variance is concentrated in lower frequencies. For more discussion on the implementation and rationale of TDE in the canonical HMM, see references (Gohil et al., 2025; Vidaurre et al., 2018; Quinn et al., 2019).

Finally, we obtained 6 HMM brain states that each characterized by a 38-dimensional Gaussian distribution (i.e., the spatial profile across the 38 ROIs). As illustrated in **Fig.1C**, the model yielded the posterior probability of being in each of the six states for every time point. The selection of a hidden state was informed by previous EEG-based HMM studies (Toffoli et al., 2024; Makkinayeri et al., 2025). To quantify the brain dynamics during the viewing of each educational video, we assigned each time point to the state based on the highest posterior probability and calculated the Fractional Occupancy (FO), defined as the fraction of total time spent in a specific state. Therefore, the higher the FO, the longer the participants spend in that brain states[3].

Results

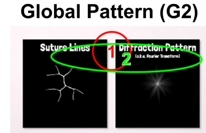
Experiment 1

A Local Pattern (G1)



	To Red	To Green
Prior	0.71	0.29
From Red	0.59	0.41
From Green	0.43	0.57

B Global Pattern (G2)



	To Red	To Green
Prior	0.5	0.5
From Red	1.00	0.00
From Green	0.00	1.00

Figure 2A) Hidden state and transition probability in Local pattern and B) in Global pattern

Eye-tracking data from 22 participants under intentional learning condition were included in the analysis. The

[3] All analysis code for replication can be found on the GitHub https://github.com/Ricwan4/EduVideo_HMM

EMHMM model identified two eye-movement patterns across video segments. Participants assigned to Group 1 showed substantially higher data likelihood under the Group 1 model than under the Group 2 model ($t(139) = 6.95$, $p < .001$, $d = 0.5874$). Conversely, participants assigned to Group 2 exhibited higher data likelihood under the Group 2 model relative to the Group 1 model ($t(139) = 21.12$, $p < .001$, $d = 1.785$), indicating robust differences in eye-movement patterns between the two groups during video viewing. 20 participants were assigned to Group 1, and 2 to Group 2. As shown in **Figure 2A**, Group 1 was characterized by local, focused and non-overlapping ROIs specialized to the features in the video, and the transition rate between the regions is high (41% for Red to Green, 43% from Green to Red). Whereas in **Figure 2B**, Group 2 was characterized by global, dispersed and overlapping ROIs that are not specialized to the video features, and the ROI did not tend to exhibit any transition. This is similar to pattern that has been found in previous EMHMM papers (Wang et al., 2025; Zheng et al., 2022). Therefore, we named Group 1 as Local Pattern, and Group 2 as Global Pattern, and constructed the L-G scale.

A significant relationship was observed between eye-movement pattern, domain knowledge, and memory performance. Shown in **Figure 3**, we found that memory performance was strongly positively correlated with domain knowledge ($r(20) = 0.71$, $p < .001$), indicating that higher prior knowledge tends to make better memory performance. In addition, the L–G scale was significantly correlated with both domain knowledge ($r(20) = 0.50$, $p = .018$), and memory performance ($r(20) = 0.62$, $p = .002$). This suggests that participants who use local, task-focused eye-movement patterns tended to demonstrate both higher prior knowledge and higher memory accuracy.

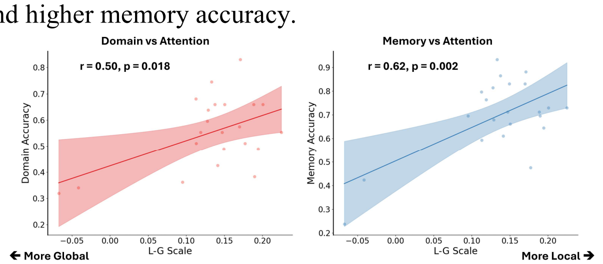


Figure 3: Relationship between L-G Scale and Domain Knowledge and Memory performance in Experiment 1.

Experiment 2

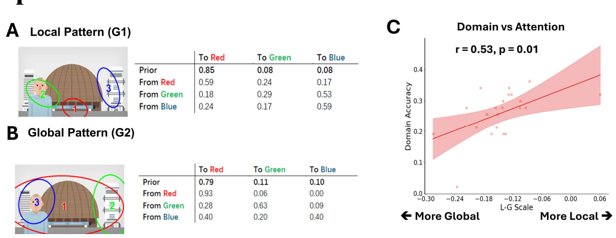


Figure 4A) Hidden state and transition probability in Local pattern and B) in Global pattern. C) relationship between L-G scale and Domain knowledge

Eye-tracking data from 24 participants under incidental learning condition were included in the analysis. Two distinct eye-movement patterns were again identified (Group 1: $t(139) = 6.96$, $p<.001$, $d = 0.59$; Group 2: $t(133) = 4.74$, $p<.001$, $d = 0.41$), confirming reliable differentiation between the two eye-movement patterns. 1 participant was assigned to group1, and 23 to group2. Similar to Experiment 1, we also construct the L-G scale and investigate the relationship between eye-movement patterns, domain knowledge and learning outcomes. However, no relationship between memory performance and domain knowledge ($r(21) = 0.12$, $p = .584$), nor with the L–G scale ($r(21) = −0.05$, $p = .826$). But as shown in **Figure 4C**, the L–G scale was positively correlated with domain knowledge ($r(21)=0.53$, $p = .010$), indicating that participants with higher domain knowledge tend to use more local eye-movement pattern.

To examine the impact of test expectancy (intentional vs. incidental learning), we compare the accuracy and L-G Scale between two experiments. As shown in **Figure 5**, under intentional condition, both Accuracy ($t(43)=13.99$, $p<.001$) and L-G Scale ($t(43)=12.76$, $p<.001$) are higher, indicating better memory and more localized attention, and vice versa.

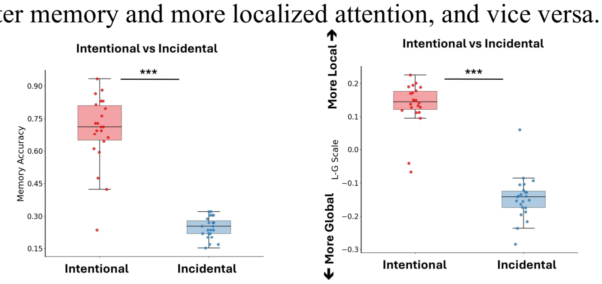


Figure 5: Comparison of Accuracy and L-G Scale between intentional and incidental condition. *** indicates $p<.001$

Experiment 3

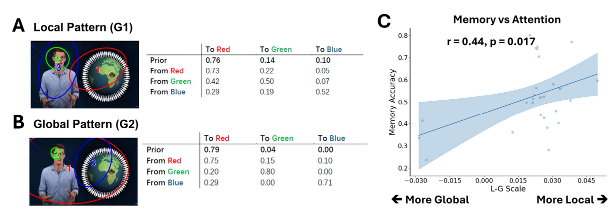


Figure 6A) Hidden state and transition probability in Local and B) in Global pattern. C) relationship with memory

In Experiment 3, we would like to investigate how the style of video influence learning performance. So eye-tracking and EEG data from 29 participants were included in the analysis. We reliably identified two distinct eye-movement patterns (all $ps<.001$), indicating robust differentiation between the two patterns. One example of local and global pattern for style 1 was given (**Figure 6A&B**). 26 participants were assigned to group1, and 3 to group2. The relationship between eye-movement patterns and learning outcomes also differed across styles. In the presenter-and-animation condition (Style 1), eye-movement patterns were significantly associated with memory performance ($r(27) = 0.44$, $p = .017$), indicating that more local, task-specialized eye-movement patterns led to better learning outcomes (**Figure 6C**). In contrast, no significant association between eye-movement and memory performance was observed in either the animation-with-writing-hand condition (Style 2) ($r(27) = 0.26$, $p = .178$), nor the presenter-with-glassboard condition (Style 3) ($r(27) = 0.18$, $p = .350$). Together, these results indicate relationship between eye-movement patterns and learning outcomes was specific to the Style 1.

For HMM of the EEG data, we identified six brain states during video viewing. Among these, States 1, 2, and 6 are shown to be linked with behavior. State 1 was defined by activation of default mode network (DMN), State 2 was defined by the activation of prefrontal and medial temporal lobe (MTL), and State 6 was defined by the deactivation of attention network. See Gohil et al (2025) for all other brain states solutions. For each participant, fractional occupancy (FO) was computed for each state, and correlation analysis was conducted to examine the relationship between FO and memory performance for each instructional video style. To control for multiple comparisons, all p values were corrected using the false discovery rate (FDR) according to Benjamini–Hochberg method (Benjamini & Hochberg, 1995).

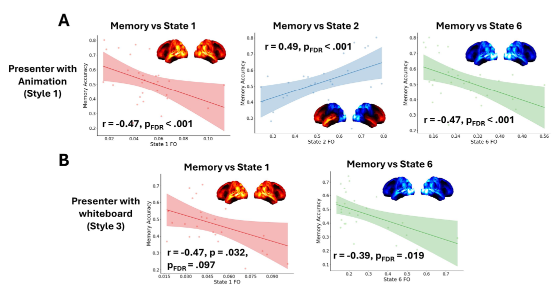


Figure 7A) Relation between brain states identified via HMM and memory in style 1 and B) in style 3.

Figure 7A shows that in the presenter-and-animation condition (Style 1), FO of State 1 ($r(53) = -0.47$, $p_{\mathrm{FDR}}<.001$) and State 6 ($r(53) = -0.47$, $p_{\mathrm{FDR}}<.001$) was negatively correlated with memory performance. In contrast, FO of State 2 showed a significant positive correlation with memory performance ($r(53) = 0.49$, $p_{\mathrm{FDR}}<.001$). No significant relationships were observed between memory performance and the occupancy of States 3, 4, or 5 after FDR correction. In the animation-with-writing-hand condition (Style 2), no significant associations were observed between FO and memory of any of the brain states, indicating an absence of reliable state–behavior relationships under this video style.

Figure 7B shows that the presenter-with-glassboard condition (Style 3), memory performance was negatively correlated with the occupancy of State 6 ($r(53) = -0.39$, $p_{\mathrm{FDR}} = .019$). Before fdr correction, State 1 was also negatively correlate with memory ($r(53) = -0.29$, $p = .032$), but this was not significant after fdr correction ($p_{\mathrm{FDR}} = .097$), and no other brain states showed reliable relationships with learning outcomes. Together, these results demonstrate that while brain states are associated with learning outcomes, it was also modulated by the video style.

Finally, we examined the enjoyment ratings between all three styles and found significant differences ($F(2, 56) = 45.70$, $p < .001$): Style 1 were rated as most enjoyable (M=5.64), followed by Style 3 (M=4.24), and Style 2 (M=3.83). Post hoc test confirmed that all pairwise differences between styles were significant ($ps \leq .012$).

Discussion

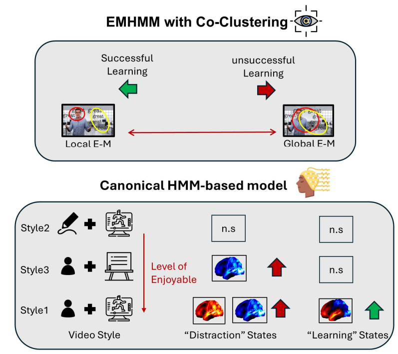


Figure 8: Summary of the findings in this research.

The present study investigated under what conditions attentional and neural processes are efficiently translated into learning outcomes during educational video viewing. Motivated by the rapid expansion of online education, this work addressed a central challenge: although attention and engagement are often assumed to benefit learning, it remains unclear when specific neural and behavioral patterns actually indicate learning success. For Experiment 1 and 2, HMM was applied to analyze the eye movement data. In Experiment 3, we applied the Canonical HMM on our EEG dataset.

Eye-Movement Pattern and Modulatory Effects of Test Expectancy. Experiments 1 and 2 investigate the relationship between eye movement and learning success and how it was modulated by test expectancy. As shown in **Figure 8** the local patterns were associated with better memory performance, whereas the global patterns were associated with poorer performance, this is consistent with previous findings (Wang et al., 2025; Zheng et al., 2022). This pattern can be interpreted in terms of the *expertise effect*. Prior work has shown that compared to novices, experts attend more to relevant aspects of the video stimulus (Jarodzka et al., 2010). Moreover, recent eye-tracking studies of professional esports players have reported similar patterns, suggesting that experts adopt more targeted visual strategies by selectively attending to task-relevant elements (Luo et al., 2025). Consistent with this account, we observed significant correlations between domain knowledge and the L–G Scale in both Experiment 1 and 2, indicating that greater domain-specific expertise is associated with more localized eye movement patterns that focus on specific parts of stimuli.

Crucially, we found that the relationship between eye movement patterns and learning success also depends on learners' test expectancy. Namely, under intentional learning condition, the learning performance is better and more localized attention patterns were used, vice versa. The *test expectancy effect* can largely account for the significantly different performance between Experiment 1 and Experiment 2. According to this account, recall-based testing encourages strategic, value-driven encoding and enhances the prioritization of task-relevant information (Middlebrooks et al., 2017). In Experiment 1, participants were explicitly instructed to memorize the learning materials, whereas no such instruction was given in Experiment 2. This difference in test expectancy likely led to systematic differences in learning strategies and the allocation of cognitive resources during encoding. A value-directed learning framework can also account for this difference, showing that greater perceived importance increases the allocation of encoding resources, resulting in memory representations that better support retrieval (Knowlton & Castel, 2022).

Together, these findings suggest that learning efficiency can be reflected on eye movement patterns, and are also modulated by test expectancy. This shows that learning efficiency depends on whether cognitive resources, like attention, can be deployed in a way that yields sufficient learning-related value (Shenhav et al., 2017).

The impact of Context-Depended Dynamics and the Multimedia instruction on learning success. Experiment 3 investigate the impact of different teaching styles. Generally speaking, the activation of DMN (State1) and deactivation of attentional network (State6) were related to poorer learning outcomes, whereas activation of memory related state (State2) were linked with successful learning. Indeed, DMN is often associated with mind wandering, and deactivation of attention network often suggests failure to invest attention resources (Christoff et al., 2016; Rosenberg et al., 2016), we refer to these as “distraction states”. Prefrontal Cortex and MTL are often associated with memory formation (Simons & Spiers, 2003), we refer to this as the “learning state”.

Interestingly, we found that both brain and eye movements seem to be context specific (**Figure 8**). For example, we only observed a significant association between State 2 and memory performance in the presenter-and-animation condition. This could be explained by the idea that neural states are shaped by task or context. Wang et al. (2025) showed that engagement in unethical decision-making can systematically alter post-task brain states. Others show the configuration of neural dynamics within the attractor landscape varies systematically as a function of situational contexts (Song et al., 2024; Song et al., 2025). Similarly, we only observed a significant association between the local eye-movement patterns and memory performance in the presenter-and-animation condition. Previous eye-tracking study also report similar context depended effect. For example, Zou et al. (2023) demonstrated that attentional allocation during reading comprehension is shaped by different context.

We did not find any significant relationship in style 2. This could also be a result of context-depended dynamics. Some styles are more enjoyable, which influence how attentional and neural resources are deployed, thereby altering the functional relevance of otherwise similar attentional or brain states. Conversely, when engagement is reduced, the same attentional strategies or neural activations may no longer support efficient encoding, resulting in context-dependent variability in learning outcomes.

A perspective article on the principles of multimedia learning provides a close theoretical match to our findings. Specifically, it proposes that learning is most effective when instructional materials are presented in ways that attract attention and guide information from limited working memory into long-term memory through multiple sensory channels (Oakley & Sejnowski, 2019). We suggest that this process may be also realted to increased enjoyability of the instructional videos, as prior work indicates that enjoyability can modulate whether attentional investment translates into meaningful learning outcomes (Hidi & Renninger, 2006). This also explains why we see learning state is associated with better learning in the presenter-and-animation, not the presenter and glassboard style.

But we didn’t find any significant association between the learning state and memory performance in the animation-with-writing-hand condition (style 2). This could be due to the lack of an instructor’s face. Previous research suggests that instructor presence, particularly facial cues, can play an important role in guiding attention and supporting learning-relevant processing (Wang et al., 2020).

Limitations. A key limitation of the present study concerns the limited sample size, which may rise concerns in not having enough power for reliable correlation analysis. Moreover, in experiment 1 & 2 participants differ both in domain accuracy and memory accuracy, as result, this may confound our comparison between conditions (as effect of expertise and effect of instructional condition are mixed). Future work should therefore address these limitations by collecting more samples and revise experiment design.

Conclusion. The present study moved beyond intersubject synchronization approach and low temporal resolution neuroimaging study, and illustrates how interpretable latent cognitive and neural states can be derived from multimodal data to indicate success in education video viewing, and how they might be influenced by test expectancy and video style. From an educational standpoint, our findings shows prior to online learning, informing participants that the knowledge in the video will be tested can effectively enhance learning performance and nudge participants to use more local eye movement pattern. Across video styles, the presenter-and-animation format was associated with relatively more favorable learning-related brain states and eye movement patterns compared to the other conditions.

References

Alhazzani N. (2020). MOOC's impact on higher education. *Social sciences & humanities open*, 2(1), 100030.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289-300.

Bali, C., Tasdelen, B., Bandi, S., & Zsidó, A. (2026). Understanding the cognitive cost of multimedia learning: effects of visual load and language proficiency. *Cognitive research: principles and implications*, *11*(1), 2.

Christoff, K., Irving, Z. C., Fox, K. C., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nature reviews neuroscience*, 17(11), 718-731.

Colclough, G. L., Brookes, M. J., Smith, S. M., & Woolrich, M. W. (2015). A symmetric multivariate leakage correction for MEG connectomes. *Neuroimage*, *117*, 439-448.

Coviello, E., Chan, A. B., & Lanckriet, G. R. (2014). Clustering hidden Markov models with variational HEM. *The Journal of Machine Learning Research*, 15(1), 697-747

Cowan, N., Bao, C., Bishop-Chrzanowski, B. M., Costa, A. N., Greene, N. R., Guitard, D., Li, C., Musich, M. L., & Ünal, Z. E. (2024). The Relation Between Attention and Memory. *Annual review of psychology*, 75, 183–214.

Fiebelkorn, I. C., & Kastner, S. (2019). A Rhythmic Theory of Attention. *Trends in cognitive sciences*, 23(2), 87–101.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23), 8410–8415.

Gohil, C., Huang, R., Higgins, C., van Es, M. W. J., Quinn, A. J., Vidaurre, D., & Woolrich, M. W. (2025). Canonical hidden Markov model networks for studying M/EEG. *bioRxiv*.

Gu, C., Peng, Y., Nastase, S. A., Mayer, R. E., & Li, P. (2024). Onscreen presence of instructors in video lectures affects learners' neural synchrony and visual attention during multimedia learning. *Proceedings of the National Academy of Sciences of the United States of America*, 121(12), e2309054121.

Hahn, M., & Keller, F. (2023). Modeling task effects in human reading with neural network-based attention. *Cognition*, 230, 105289.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist*, 41(2), 111-127.

Hsiao, J.H., Lan, H., Zheng, Y. *et al.* Eye movement analysis with hidden Markov models (EMHMM) with co-clustering. *Behav Res* **53**, 2473–2486 (2021A).

Hsiao, J. H., An, J., Zheng, Y., & Chan, A. B. (2021B). Do portrait artists have enhanced face processing abilities? Evidence from hidden Markov modeling of eye movements. *Cognition*, *211*, 104616.

Jangraw, D. C., Gonzalez-Castillo, J., Handwerker, D. A., Ghane, M., Rosenberg, M. D., Panwar, P., & Bandettini, P. A. (2018). A functional connectivity-based neuromarker of sustained attention generalizes to predict recall in a reading task. *NeuroImage*, 166, 99–109.

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154.

Knowlton, B. J., & Castel, A. D. (2022). Memory and reward-based learning: A value-directed remembering perspective. *Annual review of psychology*, 73(1), 25-52.

Liao, W., & Hsiao, J. H. (2024). Understanding the Role of Eye Movement Pattern and Consistency in Isolated English Word Reading Through Hidden Markov Modeling. *Cognitive science*, 48(9), e13489

Liao, W., Li, S. T. K., & Hsiao, J. H. W. (2022). Music reading experience modulates eye movement pattern in English reading but not in Chinese reading. *Scientific Reports*, *12*(1), 9144.

Liao, W., & Hsiao, J. H. W. (2025). Does word boundary information facilitate Chinese sentence reading in children as beginning readers?. *Learning and Instruction*, *95*, 102034.

Luo, Y., Chen, Y., Cho, J., Yan, C., & Seo, J. (2025). Differences in eye movement characteristics between expert and non-expert eSports players: a systematic review and meta-analysis. *Scientific Reports*, 15(1), 30185

Madsen, J., Júlio, S. U., Gucik, P. J., Steinberg, R., & Parra, L. C. (2021). Synchronized eye movements predict test scores in online video education. *Proceedings of the National Academy of Sciences of the United States of America*, 118(5), e2016980118.

Madsen, J., Kuppa, N., & Parra, L. C. (2025). The Brain, Body, and Behavior Dataset (BBBD): Multimodal recordings during educational videos. *bioRxiv*.

Makkinayeri, S., Guidotti, R., Basti, A., Woolrich, M. W., Gohil, C., Pettorruso, M., Ermolova, M., Ilmoniemi, R. J., Ziemann, U., Romani, G. L., Pizzella, V., & Marzetti, L. (2025). Investigating brain network dynamics in state-dependent stimulation: A concurrent electroencephalography and transcranial magnetic stimulation study using hidden Markov models. *Brain stimulation*, *18*(3), 800–809.

Middlebrooks, C. D., Murayama, K., & Castel, A. D. (2017). Test expectancy and memory for important information. *Journal of experimental psychology: Learning, memory, and cognition*, 43(6), 972–985.

Oakley, B. A., & Sejnowski, T. J. (2019). What we learned from creating one of the world's most popular MOOCs. *NPJ science of learning*, 4(1), 7.

Quinn, A. J., Vidaurre, D., Abeysuriya, R., Becker, R., Nobre, A. C., & Woolrich, M. W. (2018). Task-evoked dynamic network analysis through hidden Markov modeling. *Frontiers in neuroscience*, *12*, 603.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature neuroscience*, 19(1), 165-171.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, 79(2), 217-240.

Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. Nature reviews neuroscience, 4(8), 637-648.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping*, *17*(3), 143-155.

Song, H., Chen, R., Botch, T. L., Braver, T. S., Rosenberg, M. D., Zacks, J. M., & Ching, S. (2025). Geometry of neural dynamics along the cortical attractor landscape reflects changes in attention. *bioRxiv*, 2025-08.

Song, H., Shim, W. M., & Rosenberg, M. D. (2023). Large-scale neural dynamics in a shared low-dimensional state space reflect cognitive and attentional dynamics. *eLife*, 12, e85487. https://doi.org/10.7554/eLife.85487

Tait, L., Özkan, A., Szul, M. J., & Zhang, J. (2021). A systematic evaluation of source reconstruction of resting MEG of the human brain with a new high-resolution atlas: Performance, precision, and parcellation. *Human Brain Mapping*, *42*(14), 4685-4707.

Toffoli, L., Zdorovtsova, N., Epihova, G., Duma, G. M., Del Popolo Cristaldi, F., Pastore, M., Astle, D. E., & Mento, G. (2024). Dynamic transient brain states in preschoolers mirror parental report of behavior and emotion regulation. *Human brain mapping*, *45*(14), e70011.

Vidaurre, D., Hunt, L. T., Quinn, A. J., Hunt, B. A. E., Brookes, M. J., Nobre, A. C., & Woolrich, M. W. (2018). Spontaneous cortical activity transiently organises into frequency specific phase-coupling networks. *Nature communications*, *9*(1), 2987.

Vidaurre, D., Quinn, A. J., Baker, A. P., Dupret, D., Tejero-Cantero, A., and Woolrich, M. W. (2016). Spectrally resolved fast transient brain states in electrophysiological data. *Neuroimage*, *126*:81–95

Wang, E., Zheng, Y., Hsiao, J., & Maurer, U. (2025). Emotion influences behavioral outcomes and attention during goal-directed reading. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 47).

Wang, E., Xu, X., & Wu, H. (2025). The alternated brain states in resting state after immoral decisions. *bioRxiv*, 2025-10.

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

Zheng, Y., Ye, X., & Hsiao, J. H. (2022). Does adding video and subtitles to an audio lesson facilitate its comprehension?. *Learning and Instruction*, 77, 101542.

Zheng, Y., Que, Y., Hu, X., & Hsiao, J. H. (2022). Predicting reading performance based on eye movement analysis with hidden Markov models. In 2022 International Conference on Advanced Learning Technologies (ICALT) (pp. 172-176). IEEE.

Zou, J., Zhang, Y., Li, J., Tian, X., & Ding, N. (2023). Human attention during goal-directed reading comprehension relies on task optimization. *eLife*, *12*