Eye Movement Behavior during Mind Wandering across Different Tasks in Interactive Online Learning

Xiaoru Teng
u3306585@connect.hku.hk
Department of Psychology,
University of Hong Kong

Hui Lan
huilan@bjut.edu.cn
School of Mathematics, Statistics and
Mechanics, Beijing University of
Technology

Gloria H.Y. Wong
hoiyangloria.wong@reading.ac.uk
School of Psychology and Clinical
Language Sciences, University of
Reading

Antoni B. Chan
abchan@cityu.edu.hk
Computer Science, City University
of Hong Kong

Janet H. Hsiao
jhhsiao@ust.hk
Division of Social Science, Hong Kong
University of Science and Technology

Abstract

The recent surge in online learning demands better ways to monitor students' mind wandering (MW) episodes. We examined whether different eye movement measures were associated with MW in tasks with different cognitive demands. We found that a reduced number of fixations was associated with MW in tasks involving searching for information without clearly defined strategies. A larger variance in pupil diameter, as well as reduced eye movement consistency, were associated with MW when imagining a scenario with a central fixation. Reduced eye movement consistency, as well as reduced joint attention with another participant, were both associated with MW in tasks involving a clearly defined strategy. Interestingly, none of these eye movement measures was associated with MW in tasks involving well-learned visual routines such as face and scene identification, suggesting idiosyncrasy in eye movement behavior in these tasks. These findings have important implications for developing effective methods for detecting MW.

Keywords: mind wandering, eye tracking, EMHMM, learning

Introduction

Mind wandering (MW) is defined as a shift of attention from processing task-related information to thoughts unrelated to the task (American Psychological Association, 2018). It can negatively affect learners' learning performance. In traditional classrooms, teachers can gauge students' attention engagement by observing their behavior, which is typically more challenging in online learning. During the Covid-19 pandemic, demands for online learning have significantly surged, and this trend is expected to continue post-pandemic (Statista, 2023). However, due to the limitations of online learning platforms, a high MW rate has been reported (45% in Kane et al., 2017). Thus, ways to effectively monitor MW episodes during online learning has become an important issue in education research.

However, current research on behavioral markers of MW during online learning has been limited. During online learning, typically only videos of a participant's face are available. Thus, in addition to facial expressions, one important source of information regarding MW is from the participant's eye gaze behavior. Indeed, one's eye gaze behavior has been shown to reflect underlying cognitive processes (Rayner, 1998; Reichle et al., 2012). During MW, the mind decoupled from processing external stimuli and the thoughts shifted inward. This shift may result in distinct eye movement behavior, making eye movement measures good potential behavioral markers for MW.

Some previous studies have examined the relationship between eye movement measures and MW during learning. However, the findings have not been consistent. For example, increased fixation duration has been reported to be an indicator of MW during video lectures (Zhang et al., 2020). However, Faber et al. (2020) reported that when participants were learning from an audiobook (while maintaining a central fixation), or when they were asked to identify an infrequent target from a series of centrally presented stimuli, reduced fixation duration, rather than increased duration, was associated with MW episodes.

Faber et al. (2020) speculated that this inconsistency may be because of the differences in task demands on spatial attention allocation. More specifically, for tasks requiring extensive spatial attention allocation, such as reading illustrated texts or scene viewing, a smaller number of fixations, longer fixation duration, or more dispersed fixation distribution may indicate MW. In contrast, for tasks that require a central attention focus, such as the audiobook listening and infrequent target identification tasks, shorter fixation duration or larger saccade amplitude may be indicators of MW. However, although their results were generally consistent with this dichotomy according to task demands on spatial attention allocation, some inconsistencies remained. This suggested that a finer-grained categorization based on task demands may be required.

Therefore, here we aimed to examine whether different aspects of eye movement behavior could be used as behavioral markers of MW for tasks with different demands. More specifically, since eye fixations are typically used for gathering information for performing a task, a smaller number of fixations or longer fixation duration may be good indicators for MW in tasks requiring information search. Indeed, these markers of MW have been observed in scene viewing tasks (Krasich et al., 2018; Zhang et al., 2021).

In contrast, pupil diameters have been reported to be affected by learners' internal states such as affective

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

processes or cognitive load (Kahneman & Beatty, 1966; Partala & Surakka, 2003). Thus, measures such as variance of pupil diameter changes may be effective MW markers for tasks involving imagining scenarios, where one's attention engagement is affected by one's internal states rather than external stimuli. Consistent with our speculation, larger pupil diameter variation was observed during MW in tasks depending more on internal state processing such as letter rearrangement, word/sentence generation, mental arithmetic addition and scenery imagination tasks (Benedek et al., 2017; Annerer-Walcher et al., 2021).

In addition to these summary statistics of eye movement measures, here we considered eye movement measures that aimed to capture temporal dynamics of eye movement behavior such as transitions of eye fixations. In particular, we considered eye movement consistency measures using Eye Movement analysis with Hidden Markov Models (EMHMM; Chuk et al., 2014). In EMHMM, each participant's eye movement pattern was summarized in terms of person-specific regions of interest (ROIs) and transitions among these ROIs using a hidden Markov model (HMM), with the optimal number of ROIs determined using a variational Bayesian approach. Participants' eye movement consistency then can be measured as the entropy of the HMM (Cover & Thomas, 2006). Higher entropy indicates less predictable and thus less consistent eye movement (e.g., Hsiao, Liao et al., 2022; Hsiao, An et al., 2021; Hsiao, An et al., 2022; Hsiao, Chan et al., 2021). Previous research has suggested that high eye movement consistency in a visual task reflects a well-developed visual routine (Hsiao, An et al., 2022). We speculated that MW may interfere with the execution of a visual routine, resulting in reduced eye movement consistency. In other words, lower eye movement consistency may be a good marker of MW for tasks involving well-learned or well-defined visual strategies, such as identifying faces or familiar scenes, or making a decision between two alternative choices. In addition, Hsiao, Chan et al. (2021) showed decreased eye fixation consistency at a central target fixation location when attention was shifted away from the location prior to the saccade (see also Gersch et al., 2008; Zhao et al., 2012). Similarly, Lee et al. (2021) showed that when making judgements on a series of centrally presented stimuli while maintaining a central fixation, individuals with a wider central fixation distribution had higher MW frequency. Thus, when the task involves maintaining a central fixation while engaging in internal thoughts such as imagining a scenario or making judgements on the stimuli, decreased eye fixation consistency may indicate inattention or MW.

Since online learning typically involves multiple participants, whether a participant attends to where other participants are attending (i.e. joint attention (JA), when two or more people attend to the same object/location simultaneously; Langner et al., 2022) may also be a good indicator of MW. Accordingly, we speculated that a participant's decreased JA with others may be good indicators of MW for tasks involving a clearly defined strategy or diagnostic features. Indeed, better JA (measured as time needed to fixate at a common area of interest) has been shown to be associated with better learning outcomes during instructional video viewing (Chisari et al., 2020), suggesting its relevance to reduced MW.

To test these hypotheses, here we recorded participants' eye movements when they performed ten online learning tasks. These tasks could be categorized into four types: (1) tasks involving a clearly defined strategy, such as decision-making between two alternative choices; (2) tasks involving imagining a scenario while maintaining a central fixation; (3) tasks involving well-learned visual routines or diagnostic features, such as identifying persons or scenes; and (4) tasks involving searching for information without a clearly defined strategy such as video viewing, looking for differences, and playing board games. We hypothesized that 1) for eye fixation measures, a smaller number of fixations and longer fixation duration may be good MW indicators for tasks involving searching for information without a clearly defined strategy; 2) For pupil diameter measures, a larger variance in pupil diameter change may be a good MW indicator for tasks involving imagining a scenario while maintaining a central fixation; 3) For eye movement consistency (as measured in entropy), lower consistency may be a good MW indicator for tasks involving a clearly defined strategy or a well-learned visual routine such as viewing faces or familiar scenes. It may also be related to MW in tasks involving imagining a scenario while maintaining a central fixation; 4) For JA measures, it may be a good MW indicator for tasks where participants are expected to look at similar locations, such as tasks involving clearly defined strategies, well-learned routine/diagnostic features (e.g., person or familiar scene recognition).

Methods

Participants

We recruited 84 participants (59 females and 25 males) aged from 18 to 26 (M = 20.9, SD = 2.2), who were all healthy adult Cantonese speakers growing up in Hong Kong. They had similar university education backgrounds with normal or corrected to normal vision with glasses. A power analysis (G*Power, Faul et al., 2007) showed that a sample size of 34 was needed to obtain a medium effect size (d = .50) in a paired sample t test (β = .20; α = .05).

Materials

Participants performed 10 tasks through verbal responses. The design of the tasks was based on the criterion of "recall", "interpret" and "evaluate" according to the National Assessment of Educational Progress (National Assessment governing board, 2022). The 10 tasks were categorized into four task types:

(1) Tasks involving a clearly defined strategy:

(1.1) Sound judgment between two alternative choices This task had 30 trials. In each trial, a sound was played for 5 seconds, together with a sound icon (1.26° x 0.97°) presented at the center of the screen and two pictures (9.58°

x 6.45° each) at the left and right side of the screen respectively. Participants were asked to indicate the picture that better matched the sound. Only common daily-life sounds and pictures were used.

(1.2) Preference of a place between two different times This task had 10 trials. In each trial, participants were presented with two scene images (9.58° x 6.45° each) of Hong Kong side by side, with the one on the left photographed during an earlier time period. Participants had a discussion based on the question: "Do you prefer this place the way it was before or it is now? Why?", which typically lasted for around 35 seconds.

(2) Tasks involving imagining a scenario while maintaining a central fixation

(2.1) Imagining playing with toys This task had 20 trials. Each trial featured a picture of a toy at the screen center (16.11° x 9.69°). Participants judged and explained whether they thought the toy was interesting, which typically lasted for around 10 seconds.

(2.2) Imagining working with celebrities This task had 10 trials, with each featuring a picture of a celebrity from Hong Kong (9.88° x 9.98°). Participants had a discussion based on the question: "If there is an opportunity for you to either shoot a movie or sing with this individual, which option will you choose? Why?", which typically lasted for around 20 seconds.

(3) Tasks involving well-learned visual routines/diagnostic features

(3.1) Person Identification This task had 10 trials, with each trial presenting a photograph of a celebrity from Hong Kong (9.88° x 9.98°). Participants were asked to identify the person with a short discussion, which typically lasted for around 20 seconds.

(3.2) Scene Identification This task had 20 trials. Each trial presented a picture of a scene at the top center of the screen (12.86° x 7.61°). Participants had a discussion based on the question: "What is this scene? Which objects typically occur in this scene?", which typically lasted for around 20 seconds.

(3.3) Preference of a person This task had 10 trials, with each trial presenting a photograph of a celebrity from Hong Kong (9.88° x 9.98°). Participants had a discussion based on the question: "Do you like this person? Why?", which typically lasted for around 20 seconds.

(4) Tasks involving searching for information without a clearly defined strategy

(4.1) Video viewing This task had five trials, with each showing a video of current affairs that lasted for 30 seconds (13.77° x 7.81°). Participants were asked to view videos.

(4.2) Place Comparison This task had 10 trials, with each trial presenting two scene images of Hong Kong (9.58° x 6.45° each), with the one on the left photographed during an older time period, and the one on the right taken more recently. Participants had a discussion based on the question: "How is this place different from before?", which typically lasted for around 35 seconds.

(4.3) Tic-Tac-Toe This task had 10 trials. In each trial, participants played five rounds of the game against the host by drawing on a chessboard (9.74° x 9.74°) shown on the screen, with the other participant observing the process, which typically lasted for around 30 seconds.

Design

Each online learning session consisted of two participants in addition to the host (experimenter). We measured potential eye movement behavioral markers of MW in a task trial including average number of fixations per second, average fixation duration, variance of pupil diameter change, eye movement consistency (including overall entropy and marginal entropy of the first fixation), and percentage of JA (including participants' JA with another participant and with the host; see Eye movement data analysis section for details).

We used both thought probe and self-report methods to identify participants' MW episodes (Risko et al., 2012; Varao-Sousa & Kingstone, 2019). The thought probe method involved asking participants at the end of each trial whether they experienced MW or not during the trial. The self-report method involved asking participants to press a key whenever they noticed that they were MW.

To examine MW as a within-subject effect, following previous studies (e.g., Jang et al., 2020; Reichle et al., 2010; Mills et al., 2016), for each participant in each task, we divided the trials into No-MW and MW trials according to their answer to the thought probe at the end of each trial, and examined whether the eye movement measures obtained from the two trial types differed using paired-sample t-test.

Apparatus

Participants performed the online learning tasks through Zoom on a 10.4-inch tablet with a resolution of 2000 x 1200 pixels, connected to an EyeLink eye tracker (EyeLink 1000, EyeLink 1000 Plus, or EyeLink portable duo) with WebLink software to record their eye movement data while using an external device. A chin rest was used to minimize head movements with a 59 cm viewing distance. In all eye trackers, the tracking mode was pupil and corneal reflection with a sampling rate of 1000 Hz. EyeLink default settings for cognitive research were applied for data acquisition: the threshold for saccade motion was .1° visual angle, the saccade acceleration threshold was 8000° per square second, and the saccade velocity threshold was 30° per second.

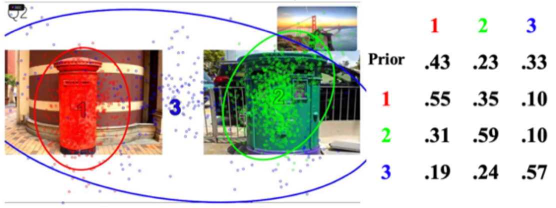
Procedure

Prior to the experiment, participants were instructed on the definition of MW, i.e., engaging in thoughts unrelated to the task. Each participant then performed the ten tasks in a random order together with another participant and the host. At the end of each trial, a thought probe question appeared on the screen, and participants indicated whether they were MW or not during the trial by pressing "Y" and "N" keys respectively. They were also instructed to press the key "M" whenever they noticed that they were MW. A nine-point calibration procedure was performed at the beginning of each task. During the tasks, participants responded to the stimuli and questions raised by the host and engaged in discussions when needed, with a small window at the upper right corner showing either the other participant or the host as in a usual zoom session.

Eye movement data analysis

Entropy

EMHMM (Chuk et al., 2014) was used to quantify participants' eye movement consistency during the tasks. For each participant, we used one HMM to summarize the participant's eye movement pattern in each trial. The hidden states of the HMM corresponded to the participant's ROIs during the task, and the eye gaze transitions were summarized into a transition matrix (Figure 1). The number of ROIs in each individual model was determined using the variational Bayesian approach from a preset range 1 to 8, with each model with a specific number of ROIs trained for 300 times, and the model with the highest log-likelihood selected for the analysis. We then calculated each participant's eye movement consistency using the entropy of the HMMs (Hsiao et al., 2021; Hsiao et al., 2020). Entropy is a measure of predictability, with higher entropy indicating a less predictable and thus less consistent eye movement pattern. In addition to overall entropy of the model, to examine the influence of participants' initial fixation, we measured the marginal entropy of the first fixation of a trial, revealing the variance in the distribution of the first fixation.



	1	2	3
Prior	.43	.23	.33
1	.55	.35	.10
2	.31	.59	.10
3	.19	.24	.57

Figure 1. Example representative model of the "place comparison" task using EMHMM. Ellipses showed ROIs with dots indicating raw fixations. Eye gaze transition probabilities among different ROIs were summarized into a matrix. Priors show probabilities that the first fixation of a trial landed in the ROIs.

Joint Attention (JA)

To identify a JA event, we considered both spatial and temporal thresholds. To determine the spatial threshold, we identified the most prominent object in each recorded video frame and calculated their average size. We used the average length (150 pixels) as the spatial threshold, that is, a JA fixation was identified if the fixation location fell within 150 pixels of another participant's fixation. For the temporal threshold, we considered fixation duration, i.e., how long two people had overlapping fixation durations in the same location. To consider the duration overlap while measuring JA from participant A and B (Figure 2), for a fixation of A with fixation duration a_1, we first determined the fixation duration overlap with B (i.e., b_{11} and b_{12}) and then the percentage of this fixation overlap relative to A's fixation duration (i.e., $(b_{11}+b_{12})/a_1$). Next, we calculated the mean percentage of fixation duration overlap across all A's fixation as the JA measurement (JA_AB_DP in Figure 2).

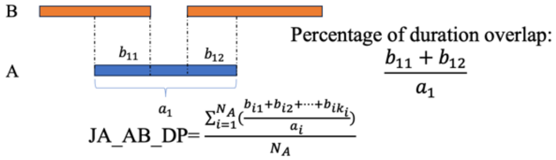


Figure 2. Demonstration of JA calculation. The long strips represent fixation durations.

Results

Fixation behavior

A smaller number of fixations per second was observed in MW episodes than in no-MW episodes in tasks involving searching for information without a clearly defined strategy, including "video viewing", *t*(49) = 3.114, *p* = .003, "place comparison", *t*(44) = 2.790, *p* = .008, and "Tic-Tac-Toe", *t*(62) = 3.007, *p* = .004. This effect was not found in the other tasks.

Similarly, longer average fixation duration was observed in MW episodes than no-MW episodes in "place comparison", *t*(44) = -2.189, *p* = 0.034. This effect marginal in "Tic-Tac-Toe", *t*(62) = -1.997, *p* = 0.050, and was not significant in "video viewing", *t*(62) = -0.903, *p* = 0.371. It was also not significant in the other tasks. Thus, consistent with our hypothesis, a decreased number of fixations and increased average fixation duration may be good MW indicators for tasks involving searching for information without a clearly defined strategy.

Pupil diameter change

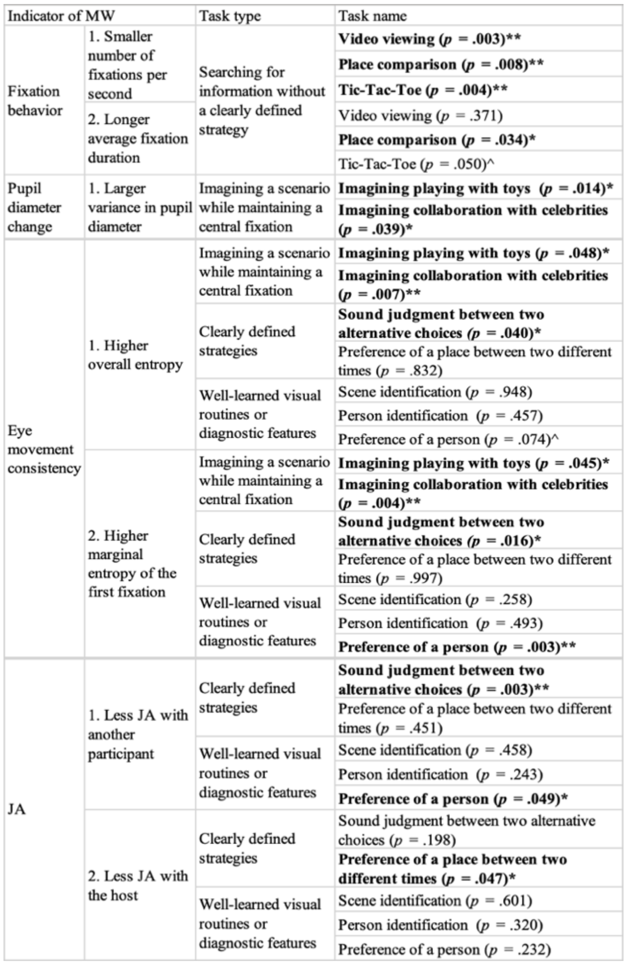
A larger variance in pupil diameter was observed in MW episodes than no-MW episodes in tasks involving imagining a scenario while maintaining a central fixation, including "imagining playing with toys", *t*(51) = -2.538, *p* = .014, and "imagining working with celebrities", *t*(47) = -2.122, *p* = .039. This effect was not observed in the other tasks, consistent with our hypothesis that pupil diameter change may be a good MW indicator of tasks involving attending to internal thoughts.

Entropy

In overall entropy, lower eye movement consistency (i.e. higher entropy) was observed in MW episodes than no-MW episodes in tasks involving imagining a scenario while maintaining a central fixation, including "imagining playing with toys", *t*(51) = -2.027, *p* = .048 and "imagining working with celebrities", *t*(47) = -2.835, *p* = .007. Also, MW episodes showed lower eye movement consistency than no-MW episodes in tasks involving well-defined strategies as in "sound judgment between two alternative choices", *t*(67) = -2.098, *p* = .040, but not for "preference of a place between two different times”. This effect was not observed in the other tasks.

In marginal entropy of the first fixation, lower consistency in the first fixation was observed in MW episodes than in no-MW episodes in tasks involving imagining a scenario while maintaining a central fixation: “imagining playing with toys”, $t(51) = -2.058$, $p = .045$, and “imagining working with celebrities”, $t(47) = -2.993$, $p = .004$. This effect was also observed in tasks involving well-defined strategies or well-learned visual routines/diagnostic features: “sound judgment between two alternative choices”, $t(67) = -2.466$, $p = .016$, and “preference of a person”, $t(43) = -3.199$, $p = .003$. This effect was not found in the other tasks.

Table 1: Summary of differences in eye movement behavior between MW and no-MW episodes observed across different task types. (All significant results are shown in bold type, ** $p < .05$, * $p < .01$, ^ $p < .10$)



Indicator of MW		Task type	Task name
Fixation behavior	1. Smaller number of fixations per second	Searching for information without a clearly defined strategy	**Video viewing (p = .003)****
			Place comparison (p = .008)**
			Tic-Tac-Toe (p = .004)**
	2. Longer average fixation duration		Video viewing (p = .371)
			Place comparison (p = .034)*
			Tic-Tac-Toe (p = .050)^
Pupil diameter change	1. Larger variance in pupil diameter	Imagining a scenario while maintaining a central fixation	**Imagining playing with toys (p = .014)***
			Imagining collaboration with celebrities (p = .039)*
Eye movement consistency	1. Higher overall entropy	Imagining a scenario while maintaining a central fixation	**Imagining playing with toys (p = .048)***
			Imagining collaboration with celebrities (p = .007)**
		Clearly defined strategies	**Sound judgment between two alternative choices (p = .040)***
			Preference of a place between two different times (p = .832)
		Well-learned visual routines or diagnostic features	Scene identification (p = .948)
			Person identification (p = .457)
			Preference of a person (p = .074)^
	2. Higher marginal entropy of the first fixation	Imagining a scenario while maintaining a central fixation	**Imagining playing with toys (p = .045)***
			Imagining collaboration with celebrities (p = .004)**
		Clearly defined strategies	**Sound judgment between two alternative choices (p = .016)***
			Preference of a place between two different times (p = .997)
		Well-learned visual routines or diagnostic features	Scene identification (p = .258)
			Person identification (p = .493)
			Preference of a person (p = .003)**
JA	1. Less JA with another participant	Clearly defined strategies	**Sound judgment between two alternative choices (p = .003)****
			Preference of a place between two different times (p = .451)
		Well-learned visual routines or diagnostic features	Scene identification (p = .458)
			Person identification (p = .243)
			Preference of a person (p = .049)*
	2. Less JA with the host	Clearly defined strategies	Sound judgment between two alternative choices (p = .198)
			Preference of a place between two different times (p = .047)*
		Well-learned visual routines or diagnostic features	Scene identification (p = .601)
			Person identification (p = .320)
			Preference of a person (p = .232)

JA

MW episodes had less JA with another participant than no-MW episodes in a task with well-defined strategies that participants generally would consistently follow, i.e. “sound judgment between two alternative choices”, $t(67) = 3.091$, $p = .003$. MW episodes also showed less JA in another task with well-defined strategies “preference of a place between two different times”, $t(46) = 2.045$, $p = .047$, although it was between JA with the host instead of with another participant, possibly because participants may have different choices. In addition, MW episodes had less JA with another participant in tasks with clear visual routines/diagnostic features, such as “preference of a person”, $t(43) = 2.023$, $p = .049$. These effects were not found in the other tasks. This result was consistent with our hypothesis that JA may be a good MW indicator for tasks involving clearly defined strategies or clear visual routines/diagnostic features.

In summary, our results generally supported our hypothesis that different eye movement behavior was associated with MW episodes across tasks with different task demands. Specifically, a smaller number of fixations and longer fixation duration were good MW indicators for tasks involving searching for information without a clearly defined strategy. A larger variance in pupil diameter was a good indicator of MW for tasks involving imagining a scenario while maintaining a central fixation. Lower eye movement consistency could be a good indicator of MW for tasks involving attending to internal states such as imagining a scenario, well defined strategies such as making a decision between two alternative choices, or well-learned visual routine/diagnostic features such as identifying faces. Similarly, JA measures could be good indicators of MW for tasks involving well defined strategies or clear visual routine/diagnostic features (The results were summarized in Table 1).

Discussion

In this study, we examined whether different eye movement behavioral measures were associated with MW in tasks with different cognitive demands during interactive online learning. Consistent to our hypotheses, we found that eye fixation behavior, including number of fixations and average fixation duration, could be potential MW indicators in tasks involving searching for information without a clearly defined strategy, with a smaller number of fixations and longer average fixation duration observed in MW than no-MW trials. This finding was consistent with previous studies (Krasich et al., 2018; Zhang et al., 2021; Faber et al., 2020). Number of fixations may be a particularly useful indicator in this scenario, as a significant effect was observed across all three tasks involving searching for information without a clearly defined strategy, including “video viewing”, “place comparison”, and “Tic-Tac-Toe”. Since in these tasks, eye fixations were required for gathering information, during MW the processing of external information was suppressed to give priority to the thoughts generated internally (Schooler et al., 2011; Smallwood et al., 2008), resulting in a reduced ability to plan and execute saccades and eye fixations. In contrast, longer fixation duration was not found in MW than no-MW trials during video viewing. Previous research has shown that a more centralized attention strategy to focus on the screen center when learning from video lessons was associated with better comprehension performance (Zheng et al., 2022). Thus, longer fixations during video viewing may sometimes reflect this centralized attention strategy instead of MW, making it an unreliable indicator.

In contrast, pupil diameter change was a potential MW indicator for tasks involving imagining a scenario while maintaining a central fixation, with a larger variance in pupil diameter observed in MW trials than no-MW trials. This finding was consistent with our hypothesis. More specifically, pupil diameters could usually be affected by internal states such as affective processes (Ferrari et al., 2016; Henderson et al., 2014, Snowden et al., 2016), internal attention focus (Annerer-Walcher et al., 2021; Benedek et al., 2017), or cognitive load (Kahneman & Beatty, 1966; Partala & Surakka, 2003). Imagining a scenario while viewing a centrally presented stimulus may not only involve internal attention focus, but also increase cognitive load as compared with viewing the external stimulus alone. The tasks here, i.e., imagining playing with toys or collaborating with celebrities, may also induce affective processes. Thus, pupil diameter change could be a potential indicator for MW in this scenario.

Eye movement consistency, including overall entropy and marginal entropy of the first fixation measured here, was also a potential MW indicator in tasks involving imagining a scenario while maintaining a central fixation, with lower consistency observed in MW than no-MW trials. This result was again consistent with our hypothesis, as a wider eye fixation location distribution (and thus lower consistency) at a centrally presented stimulus during visual tasks have been reported to be associated with inattention, higher MW frequency, and poorer task performance (e.g., Hsiao, Chan et al., 2021; Lee et al., 2021).

In contrast, eye movement consistency did not appear to be a consistent indicator for tasks involving a clearly defined strategy or well-learned visual routines/diagnostic features, although we hypothesized that MW should interfere with strategy or visual routine implementation, resulting in decreased eye movement consistency. Specifically, we found that lower eye movement consistency was a significant MW indicator in “sound judgment between two alternative choices”, but not in “preference of a place between two different times. Perhaps the latter task involved making preference decisions, and thus the attention strategy was less clearly defined as sound judgements and could involve a higher variance across trials (see Chuk et al., 2020; Shimojo et al., 2003). Similarly, eye movement consistency was a significant indicator in “preference of a person” but not in “scene identification” or “person identification”. Perhaps in these well-learned tasks, individuals still need to adaptively adjust attention strategies, which could lead to reduced eye movement consistency. Thus, reduced eye movement consistency was not a reliable indicator in this scenario.

Lastly, reduced JA was found in MW as compared with no-MW trials in tasks involving a clearly defined strategy. Since participants tended to follow a similar strategy and a high level of JA was expected, reduced JA could be a good MW indicator. Interestingly, JA with another participant was a good indicator for “sound judgment between two alternative choices”, whereas JA with the host was a good indicator for “preference of a place between two different times”. Since preference decision making may involve larger individual differences, JA with another participant may not be a good MW indicator. In contrast, the host might attend to participants’ verbal responses in the preference judgment task, and thus JA with the host became a good MW indicator. However, while JA with another participant may be a good MW indicator in this scenario, under what scenarios JA with the host would be a good MW indicator requires further examinations. In contrast, different from our hypothesis, participants did not differ in JA between MW and no-MW trials in tasks involving well-learned visual routines. Indeed, recent research has reported that in well-learned visual tasks such as face and scene perception, individuals develop idiosyncratic eye movement behavior that is consistent over time (Peterson & Eckstein, 2013; Hsiao, Lan et al., 2021; An & Hsiao, 2021). For example, in face recognition individuals could differ in adopting more eyes- or nose-focused face scanning patterns (Chuk et al., 2017; Chan et al., 2018). These individual differences in the visual routines may render JA with another participant/host an ineffective MW indicator.

Thus, different eye movement behavioral measures could be associated with MW in tasks with different cognitive demands, and a slight difference in task demands may change the effect. This finding is generally consistent with the current literature on human attention, where eye movements in visual tasks are reported to be goal-driven and task-dependent (Henderson, 2017; Hsiao & Chan, 2023), as well as being idiosyncratic (Kanan et al., 2015). Future work will examine whether these indicators or a combination of the indicators can be used to detect MW episodes during interactive online learning using machine learning classifiers.

In conclusion, here we showed that different eye movement behavioral measures were associated with MW behavior in different tasks during interactive online learning, depending on the task demands. Specifically, eye fixation behavior, in particular a reduced number of fixations, was associated with MW in tasks involving searching for information without a clearly defined strategy. A larger variance in pupil diameter, as well as reduced eye movement consistency, were associated with MW in tasks related to imagining a scenario while maintaining a central fixation. Reduced eye movement consistency was also associated with MW in tasks involving a clearly defined strategy. Finally, reduced JA with another participant was also associated with MW in tasks involving a clearly defined strategy. However, subtle differences in task demands, such as perceptual vs. preference decision making, could lead to different results. Interestingly, none of the potential indicators that we tested here was associated with MW in tasks involving well-learned visual routines such as face and scene identification, suggesting idiosyncrasy in eye movement behavior in these tasks. These findings have important implications for developing effective methods for detecting MW episodes during interactive online learning.

Acknowledgments

We are grateful to Research Grant Council of Hong Kong (CRF #C7055-21G). We are also grateful to R&D Program of Beijing Municipal Education Commission, China (KM202310005014).

References

American Psychological Association. (2018). Mind Wandering. In *APA Dictionary of Psychology*. Retrieved December 12, 2023, from https://dictionary.apa.org/mind-wandering

An, J., & Hsiao, J. H. (2021). Modulation of mood on eye movement pattern and performance in face recognition. *Emotion, 21*(3), 617-630. https://doi.org/10.1037/emo0000724

Annerer-Walcher, S., Ceh, S. M., Putze, F., Kampen, M., Körner, C., & Benedek, M. (2021). How reliably do eye parameters indicate internal versus external attentional focus?. *Cognitive Science, 45*(4), e12977. https://doi.org/10.1111/cogs.12977

Benedek, M., Stoiser, R., Walcher, S., & Körner, C. (2017). Eye behavior associated with internally versus externally directed cognition. *Frontiers in Psychology, 8*, 1092. https://doi.org/10.3389/fpsyg.2017.01092

Chan, C. Y. H., Chan, A. B., Lee, T. M. C., & Hsiao, J. H. (2018). Eye movement patterns in face recognition are associated with cognitive decline in older adults. *Psychonomic Bulletin & Review, 25*(6), 2200-2207. https://doi.org/10.3758/s13423-017-1419-0

Chisari, L. B., Mockevičiūtė, A., Ruitenburg, S. K., Vemde, L., Kok, E. M., & Gog, T. (2020). Effects of prior knowledge and joint attention on learning from eye movement modelling examples. *Journal of Computer Assisted Learning, 36*(4), 569–579. https://doi.org/10.1111/jcal.12428

Chuk, T., Chan, A. B., & Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden Markov models. *J. Vision, 14*(11), 8. https://doi.org/10.1167/14.11.8

Chuk, T., Chan, A. B., & Hsiao, J. H. (2017). Is having similar eye movement patterns during face learning and recognition beneficial for recognition performance? Evidence from hidden Markov modeling. *Vision Research, 141*, 204-216. https://doi.org/10.1016/j.visres.2017.03.010

Chuk, T., Chan, A. B., Shimojo, S., & Hsiao, J. H. (2020). Eye movement analysis with switching hidden Markov models. *Behavior Research Methods, 52*(3), 1026-1043. https://doi.org/10.3758/s13428-019-01298-y

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory*. John Wiley & Sons.

Faber, M., Krasich, K., Bixler, R. E., Brockmole, J. R., & D'Mello, S. K. (2020). The eye–mind wandering link: Identifying gaze indices of mind wandering across tasks. *Journal of experimental psychology: human perception and performance, 46*(10), 1201. https://doi.org/10.1037/xhp0000743

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. https://doi.org/10.3758/BF03193146

Gersch, T. M., Kowler, E., Schnitzer, B. S., & Dosher, B. A. (2008). Visual memory during pauses between successive saccades. *Journal of Vision, 8*(16):15, 1-18. https://doi.org/10.1167/8.16.15

Henderson, J. M. (2017). Gaze control as prediction. *Trends in cognitive sciences, 21*(1), 15-23. https://doi.org/10.1016/j.tics.2016.11.003

Henderson, R. R., Bradley, M. M., & Lang, P. J. (2014). Modulation of the initial light reflex during affective picture viewing. *Psychophysiology, 51*(9), 815-818. https://doi.org/10.1111/psyp.12236.

Hsiao, J. H, Lan, H., Zheng, Y., & Chan, A. B. (2021). Eye Movement analysis with Hidden Markov Models (EMHMM) with co-clustering. *Behavior Research Methods, 53*, 2473-2486. https://doi.org/10.3758/s13428-021-01541-5

Hsiao, J. H. W., Liao, W., & Tso, R. V. Y. (2022). Impact of mask use on face recognition: an eye-tracking study. *Cognitive Research: Principles and Implications, 7*(1), 1-15. https://doi.org/10.1186/s41235-022-00382-w

Hsiao, J. H., & Chan, A. B. (2023). Visual attention to own- vs. other-race faces: Perspectives from learning mechanisms and task demands. *British Journal of Psychology, 114*, 17-20 https://doi.org/10.1111/bjop.12647

Hsiao, J. H., An, J., & Chan, A. B. (2020). The role of eye movement consistency in learning to recognise faces: Computational and experimental examinations. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 1072-1078). Cognitive Science Society.

Hsiao, J. H., An, J., Hui, V. K. S., Zheng, Y., & Chan, A. B. (2022). Understanding the role of eye movement consistency in face recognition and autism through integrating deep neural networks and hidden Markov models. *npj Science of Learning, 7*(1), 28. https://doi.org/10.1038/s41539-023-00154

Hsiao, J. H., An, J., Zheng, Y., & Chan, A. B. (2021). Do portrait artists have enhanced face processing abilities? Evidence from hidden Markov modeling of eye movements. *Cognition, 211*, 104616. https://doi.org/10.1016/j.cognition.2021.104616

Hsiao, J. H., Chan, A. B., An, J., Yeh, S. L., & Jingling, L. (2021). Understanding the collinear masking effect in visual search through eye tracking. *Psychonomic Bulletin & Review, 28*(6), 1933-1943. https://doi.org/10.3758/s13423-021-01944-7

Jang, D., Yang, I., & Kim, S. (2020). Detecting mind-wandering from eye movement and oculomotor data during learning video lecture. *Education Sciences, 10*(3), 51. https://doi.org/10.3390/educsci10030051

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(3756), 1583-1585. https://doi.org/10.1126/science.154.3756.1583

Kanan, C., Bseiso, D., Ray, N., Hsiao, J. H., & Cottrell, G. (2015) Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research, 108*, 67-76. https://doi.org/10.1016/j.visres.2015.01.013

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

Langner, M., Toreini, P., & Maedche, A. (2022). EyeMeet: A Joint Attention Support System for Remote Meetings. *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3491101.3519792

Lee, H. H., Chen, Z. L., Yeh, S. L., Hsiao, J. H., & Wu, A. Y. (2021). When eyes wander around: mind-wandering as revealed by eye movement analysis with hidden Markov models. *Sensors, 21*(22), 7569. https://doi.org/10.3390/s21227569

Mills, C., Bixler, R., Wang, X., & D'Mello, S. K. (2016). Automatic Gaze-Based Detection of Mind Wandering during Narrative Film Comprehension. *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 1072-1078). International Educational Data Mining Society.

National Assessment governing board. (2022). *Reading assessment framework for the 2022 and 2024 National Assessment of Educational Progress*. National Assessment governing board. https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/reading/2022-nagb-reading-framework-508.pdf

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies, 59*(1-2), 185-198. https://doi.org/10.1016/S1071-5819(03)00017-X

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science, 24*(7), 1216–1225. https://doi.org/10.1177/0956797612471684

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372. https://doi.org/10.1037/0033-2909.85.3.618

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye–mind link. *Psychological Review, 119*, 155–185. http://dx.doi.org/10.1037/a0026473

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological science, 21*(9), 1300-1310. https://doi.org/10.1177/0956797610378686

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

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. Trends in *Cognitive Sciences, 15*, 319 –326. http://dx .doi.org/10.1016/j.tics.2011.05.006

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6*, 1317–1322. https://doi.org/10.1038/nn1150

Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition, 36*(6), 1144-1150. https://doi.org/10.3758/MC.36.6.1144

Snowden, R. J., O'Farrell, K. R., Burley, D., Erichsen, J. T., Newton, N. V., & Gray, N. S. (2016). The pupil's response to affective pictures: Role of image duration, habituation, and viewing mode. *Psychophysiology, 53*(8), 1217-1223. https://doi.org/10.1111/psyp.12668.

Statista (2023). *Online Learning Platforms - Worldwide.* Statista. https://www.statista.com/outlook/dmo/eservices/online-education/online-learning-platforms/worldwide

Varao-Sousa, T. L., & Kingstone, A. (2019). Are mind wandering rates an artifact of the probe-caught method? Using self-caught mind wandering in the classroom to test, and reject, this possibility. *Behavior Research Methods, 51*, 235-242. https://doi.org/10.3758/s13428-018-1073-0

Zhang, H., Anderson, N. C., & Miller, K. F. (2021). Refixation patterns of mind-wandering during real-world scene perception. *Journal of experimental psychology: human perception and performance, 47*(1), 36. https://doi.org/10.1037/xhp0000877

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

Zhao, M., Gersch, T. M., Schnitzer, B. S., Dosher, B. A., & Kowler, E. (2012). Eye movements and attention: The role of pre-saccadic shifts of attention in perception, memory, and the control of saccades. *Vision Research, 74*(1), 40-60. https://doi.org/10.1016/j.visres.2012.06.017

Zheng, Y., Ye, X., & Hsiao, J. H. (2022). Does adding video and subtitles to an audio lesson facilitate its comprehension? *Learning & Instruction, 77*, 101542. https://doi.org/10.1016/j.learninstruc.2021.101542