

Getting the wiggles out: Movement between tasks predicts future mind wandering during learning activities

Rosy Southwell^[0000-0003-4141-523X], Candace E. Peacock^[0000-0001-5865-0653],
and Sidney K. D'Mello^[0000-0003-0347-2807]

University of Colorado Boulder, Boulder CO 80301, USA
rosy.southwell@colorado.edu

Abstract. Mind wandering (“zoning out”) is a frequent occurrence and is negatively related to learning outcomes, which suggests it would be beneficial to measure and mitigate it. To this end, we investigated whether movement from a wrist-worn accelerometer between tasks could predict mind wandering as 125 learners read long, connected, informative texts. We examined random forest models using both basic statistical and more novel nonlinear dynamics movement features, finding that the former were more predictive of future (i.e., about 5 minutes later) reports of mind wandering. Models generalized across students with AUROCS up to 0.62. Importantly, vertical movement as measured by the Z-axis accelerometer channel, e.g. flexion or extension of the elbow in stretching, was the most predictive signal, whereas horizontal arm movements (measured by X- and Y-axis channels) and rotational movement were not predictive. We discuss implications for theories of mind wandering and applications for intelligent learning interfaces that can prospectively detect mind wandering.

Keywords: mind wandering, fidgeting, reading comprehension, movement

1 Introduction

Mind wandering (MW) is defined as ‘zoning out’, or where attention disengages from the primary task towards task-unrelated thoughts [10, 18]. Students are estimated to mind wander approximately 30% of the time during learning [33], varying from 20% to 40% depending on the learning activity, with greater occurrence in passive forms of learning such as lecture viewing (35%, [33]) than when engaging with interactive learning technology (23% [19]). Although MW has benefits such as enabling future planning and creativity [23], meta-analyses indicate that it is negatively related to performance for activities involving focused attention and concentration [10, 14]. Accordingly, researchers have investigated various interventions to alleviate the effects of MW in education. Proactive interventions are (usually non-individualized) modifications to task parameters made beforehand, such as changing text properties to reduce MW [13, 17, 10] whereas reactive interventions aim to detect and mitigate effects of MW after its onset, for example, by repeating content or asking a probing question to reengage attention and mitigate any learning deficits [8, 22].

This raises the question of whether there are patterns of behavior that can prospectively predict that future MW will occur. If so, detecting such patterns could be used in

learning technologies to enable proactive interventions or to improve prediction of MW when it occurs. These possibilities depend on the ability to identify behaviors that can prospectively predict MW compared to existing research that focuses on detecting it when it occurs [3, 8, 14, 22, 28]. As a step in this direction, the present study examines whether patterns of movement during pre-defined intra-task intervals can prospectively predict self-reports of MW as probed roughly 5 minutes later during learning from complex texts. We model movement because well-established theoretical and empirical research links it to engagement and attention [3, 5, 9, 14, 24, 28] and movement can be measured using low-cost wearable accelerometers scalable for use in classroom settings. In addition to the practical applications, the present work also advances empirical knowledge of the behaviors that underlie MW.

1.1 Background and Related work

Movement and student engagement. Mind wandering is an attentional subcomponent of (dis)engagement [6], and one way that movement has been studied in relation to engagement is through monitoring posture. In a classic study, Mota and Picard (2003) measured how posture (as measured by pressure sensors on a seat and back of a chair) related to a child's level of interest [24]. D'Mello et al. extended this approach to test how a student's posture, also measured via a pressure-sensitive chair, was related to their level of engagement and found that boredom was associated with leaning back and changes in seat pressure and that flow was associated with leaning forward [5, 9]. Fidgeting, which indicates restlessness, has also been associated with MW in classrooms using self-report measures [3]. Seli et al. (2014) expanded beyond self-reports of fidgeting and found that increased MW was associated with more fidgeting as indexed by participants sitting on a Wii balance board [28]. Farley et al. (2013) demonstrated that fidgeting (as coded from video) and self-reports of MW increased as a function of time during lecture viewing and that fidgeting was negatively predictive of the retention of lecture material beyond the role of MW [14].

Whereas the above research suggests a link between body movement and engagement, it is unknown whether there are patterns of movement that are indicative of future MW. Nevertheless, the literature on aerobic breaks (e.g., running on the spot, jumping exercises, stretching) can shed some light. Indeed, when aerobic exercise was added to breaks, participants reported decreases in fatigue and increases in vigor compared to unstructured breaks in which students could do what they wanted as long as they remained in their seats [1]. Furthermore, when exercise breaks (calisthenic exercises), non-exercise breaks (computer game), and no breaks were introduced into computerized lectures, it was found that exercise breaks promoted attention (lower MW) and better retention throughout the lecture relative to the non-exercise breaks and no breaks condition [15]. These studies suggest that moving more during breaks in learning tasks can have some positive benefits in terms of reducing MW, but have yet to yield predictive models of MW from behavior measured during breaks.

Prospectively predicting mind wandering. There is a small body of work that focuses on whether it is possible to prospectively predict MW, although without considering

movement. In a simulation study Kopp et al. (2014) explored whether individual attributes (e.g., working memory capacity) could be used to identify reading situations (defined as text difficulty and utility value) that resulted in the lowest MW rates for an individual [20], suggesting that a learning system that places learners in conditions to reduce MW is feasible (though they did not test this possibility). Another study found that eye movements were more dispersed (spread out) before people later reported MW [34], suggesting that gaze could prospectively predict MW. However, cross-validation was not done so it is unknown whether their models would generalize across learners. In another recent study using electroencephalography, causality analysis of timeseries of oscillatory neural activity indicated that it predicted MW several minutes later [25]. However, taking such neural measurements would be impractical for a learning environment.

1.2 Current Work: Contributions and Novelty

We tested the intriguing possibility as to whether machine-learning models of movement during task breaks can prospectively predict MW. Specifically, we measured movement (via accelerometer data) during short intra-task intervals between reading each of five expository texts and collected probe-caught MW responses as students read each text. Random forest models were trained to predict MW from movement measured prior to reading (i.e., during breaks) in a learner-independent fashion.

Our study is novel in multiple respects. To our knowledge, this is the first attempt to prospectively predict MW in a student-generalizable manner, which has several applications if successful (see Discussion). Second, prior work has employed basic statistical movement features (e.g., mean movement, variance). As an alternative we employed a technique from nonlinear dynamical analyses called recurrent quantification analysis (RQA [32]) to characterize various patterns of movement, including fidgeting, which has been associated with MW [3, 5, 9, 14, 24, 28].

We examined three research questions. First, we evaluated how the prospective MW models performed relative to several baselines including a context model that captured situational factors independent of movement, a “gaze dispersion” model inspired by [34], and a model of movement during the MW episode. Second, we examined whether models consisting of statistical movement features or RQA were better predictors of MW. A final question examined which channels and features of movement were most predictive of mind wandering.

2 Methods

2.1 Data Collection

Data were collected as part of a larger study investigating neurophysiology during learning from complex texts. Only aspects germane to the present study are presented here. The data analyzed here have not been previously published. Participants (N=156, age 23 ± 6 years, 67% female, 32% male, 1% other) were students from a large public University in the Western US (77% White, 9% Asian, 7% Hispanic, 5% Other, 2%

prefer not to say). Participants were paid \$20 per hour plus \$10 for a follow-up survey via Amazon gift cards. All procedures were approved by the Institution's internal review board (IRB00000191 protocol #19-0396) and all participants provided informed consent after being given the opportunity to read the consent form prior to the study and upon arrival, and to ask the researcher any questions. We analyzed data from the 125 participants with complete accelerometer data.

Participants each read five expository texts of around 1000 words each split into 10 pages (screens of text) on the topic of behavioral research methods: Bias, Hypothesis, Casual Claims, Validity, and Variables. The texts had a mean Flesch-Kincaid grade level of 13.2 indicating an advanced reading level [16] suitable for college students. Reading was self-paced in that participants pressed a key to advance to the next page but could not return to a previous page. On average, participants read each text for 5.5 minutes ($SD = 1.8$), for a total of 27.6 minutes ($SD = 9.2$).

As participants read the texts, they were probed for MW at two pseudorandom points roughly corresponding to the first and second half of the text. Specifically, upon advancing to the next page of text, they were presented with the following question: *Were you zoning out when you read the previous page?* Participants responded with “yes” or “no” using mouse clicks. Such self-reports of MW are a standard and validated method for measuring conscious phenomena [10, 29]. Participants were trained on how to respond to MW probes prior to the main reading session. Here, they were given a definition of MW (i.e., abbreviated as thinking about something other than reading or “zoning out”) and were informed that they would be asked about these zone outs as they read. Participants were asked to be honest about these zone outs, because they are perfectly natural and were reassured that responding “yes” would not affect their progress in the study. As a knowledge check, they were asked to identify an example of zoning out in a multiple-choice question.

Participants wore a Shimmer3 (sampling rate = 51.2 Hz) on their non-dominant wrist. The sensors include electrodermal activity, photoplethysmography, two accelerometers, a gyroscope, and a magnetometer, but models focused on the latter three sensors. Participants also underwent gaze tracking using the Eyelink 1000+ desktop mounted eye-tracker. Chin and forehead rests were used to minimize head movement during reading, but participants were encouraged to stretch and reposition prior to reading each text: participants were presented with a prompt (*“Before proceeding, if you need a short break to stretch or close your eyes, you can do so now.”*) and to press a button to proceed when ready. Next the eyetracker was recalibrated by collecting fixation positions from nine known target positions (presented in random order) to map raw eye data to gaze position on-screen. Following calibration, participants went through a validation procedure which measured the difference between the computed fixation position and the fixation position for the target during calibration. The error reflects the accuracy of the calibration. Analyses focused on movement (accelerometer, gyroscope, magnetometer) between texts, i.e. the intra-task interval comprising the break and eye-tracker recalibration. Eye-tracker calibration error was also analyzed as this may increase if the participant moves or fidgets during the calibration. The time stamps corresponding to the initial eyetracker calibration (preceding the first text read) were not recorded so we focus on the recalibration periods prior to the final four texts.

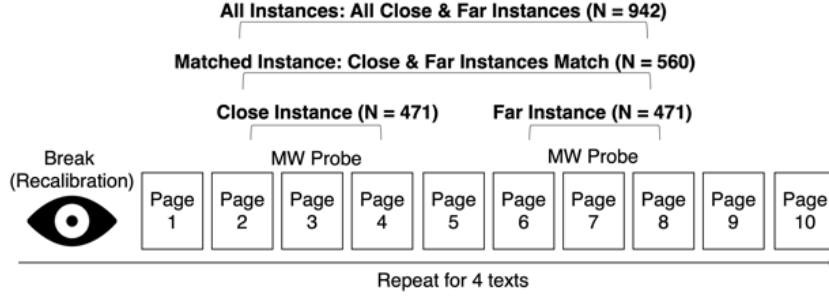


Fig. 1. Example event sequence. First, participants were recalibrated on the eyetracker during a break and then proceeded to read 10 pages of an expository text. They were probed for MW on two pseudo randomly selected pages. N refers to the number of instances.

2.2 Machine Learning Models

We chose Random Forest classifiers since they incorporate nonlinearity and interactivity among features and have good generalization properties. The random forest classifier was implemented in *sklearn*, with 100 estimators, minimum of 15 samples per leaf, and the maximum number of features set to the square root of the total number of features. The class weights of the models were balanced by setting the weights to be inversely proportional to the number of samples in each class. Note that no resampling was done on either the training or testing sets: setting the class weights to ‘balanced’ simply penalizes incorrect predictions made on the minority class.

We tested several different models. One model was trained on only the first MW probe per text (Fig. 1) to test whether the features were more sensitive to future MW that occurred closer in time to calibration (“closest”). A second model was trained on only the second MW probe per text (Fig. 1) to test the endurance of the model (“farthest”). A third model was only tested on cases where MW responses for the nearest and farthest probes were the same (i.e., positive or negative MW) (“matched”). The matched model provided the purest test of whether movement was predictive of future MW since people consistently mind wandered (or not) throughout the duration of the text. A final model was tested on all MW probes administered during each text (“all”).

We used four-fold cross-validation at the participant-level to ensure generalizability to new participants [7]. Here, the dataset was randomly split into four folds, with the data from a given participant only being in a single fold. Predictions were pooled across the four folds prior to computing AUROCs. The process was repeated 101 times with a different random partitioning of the folds for each run. The same fold assignments were used to train all the models (including baseline models) per run. Performance was evaluated using the area under the receiver operator characteristic (AUROC) curve, which ranges from 0 to 1 with an AUROC of 0.5 indicating baseline chance-level performance. Here, we focus on the run with the median-performing model for each model specification as the results were highly consistent across runs (95% confidence intervals

at most differed by 0.01 AUROC across models). We used the `roc.test` function from the `pROC` package in R [27] (bootstrap test with 10,000 iterations) to statistically compare the median run’s AUROC between models. The false discovery rate correction was used to correct for six multiple comparisons.

2.3 Sensors, Data Processing, and Feature Extraction

Sensors & Signals. The Shimmer3 has a low noise accelerometer with a lower range of values (range: ± 2.0 g), a wide range accelerometer that has a wider range of possible values (range: ± 2.0 g; ± 4.0 g; ± 8.0 g; ± 16.0 g), a gyroscope, and a magnetometer. All sensors were sampled at 51.2 Hz. Accelerometers and gyroscopes measure linear and rotational motion in X, Y, and Z coordinates, respectively, whereas magnetometers indicate direction. From both accelerometer’s *acceleration* time series, we additionally derived *velocity* [acceleration * change in time] and *distance* [acceleration * (change in time)²]. This resulted in 6 accelerometer time series (3 measures [acceleration, velocity, distance] x 2 sensors [low noise vs. wide range]) for each axis (X,Y,Z). From the *angular velocity* (i.e. gyroscope), we computed *degrees rotated* [angular velocity * change in time] totaling 2 timeseries for each axis. Combining these accelerometer and gyroscope signals with the magnetometer *direction* time series this totaled 9 movement measures * 3 axes, yielding a timeseries for each of 27 “channels” from which statistical features were derived (see below). We smoothed each time series using the `denoise_wavelet` function in Scikit Image (`wavelet='sym4'`, `mode='soft'`, `wavelet_levels=3`) then removed values that exceeded 3 SD from the mean to account for noise and linearly interpolated over removed data. Fig. 2 depicts example accelerometer time series for intra-task intervals corresponding to positive and negative MW responses.

Statistical Movement Features. The following eight statistical features were computed from each of the 27 time series: mean, median, maximum, minimum, standard deviation, kurtosis, skew, and range [11, 26]. Statistical features were z-scored within-participant. Because the question of interest was whether movement intra-task intervals was prospectively predictive of MW, we computed statistical features from the time series between reading blocks, which included a self-paced break followed by eye-tracker recalibration (total $M = 142.8$ sec; $SD = 90.2$ sec). As the intra-task interval contained both a break and the recalibration, we also extracted features from just the break ($M = 13.4$ sec, $SD = 16.5$ sec), and just the calibration ($M = 101.8$ sec, $SD = 73.6$ sec), and fit models using the features from these intervals alone. For comparison with the pre-reading movement models, we also extracted features during reading the page corresponding to the MW probe (as a baseline check; $M = 34.0$ s; $SD = 10.4$ s).

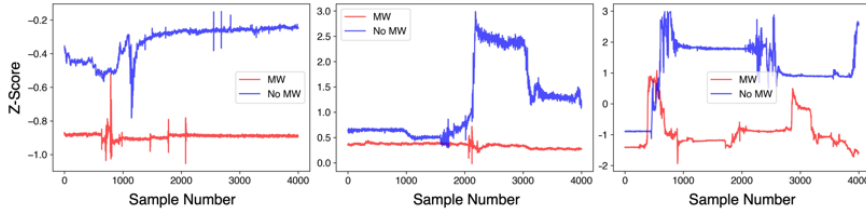


Fig. 2. Example time series from the z component of the low noise accelerometer for breaks corresponding to positive (red lines) and negative (blue lines) responses to MW probes.

Calibration Error Features. During each validation period, the eye-tracker outputs the degrees, horizontal (X pixels), and vertical error (Y pixels) based on how far each fixation was from each of the nine calibration points. We computed the same eight statistical features over the 9 calibration errors, resulting in 24 calibration error features.

Context Features. Context features capture situational factors independent of movement and were used as a baseline measure to gauge the added value of movement features. They comprised calibration duration and text order.

Model Comparisons. To evaluate which channels would provide the most information, we first trained separate models for the 9 movement measures, each on 24 statistical movement features: 8 statistical features * 3 axes, plus the two context features. The angular velocity, degrees rotated, and direction channels were not predictive of MW (all AUROCs at chance). We found that the low noise and wide range accelerometer features yielded very similar (AUROCS within 0.01) results, so we focused on the low-noise accelerometer channels. The most strongly predictive accelerometer-derived channel was distance. For this low noise distance channel, we then trained separate models using the 8 statistical features of the X, Y, and Z components. Overall, the Z component which indexed the direction perpendicular to the wrist surface (i.e., flexion and extension of the elbow/arm, such as may occur during stretching, or lifting the hand from the keyboard and mouse) [30] contained the most information (M AUROC = 0.62, 95% CI = [0.62, 0.62]) relative to X and Y (which indexed horizontal, planar movements such as moving the arm across the desk [30]) which were at chance, so we focused models on the Z component statistical features from the low noise distance channel.

Recurrent Quantification Analysis Features. We conducted recurrent quantification analysis (RQA) to complement the statistical movement features. RQA is an analytic suite used to investigate the dynamics of complex timeseries, and is a powerful tool that avoids several of the assumptions of traditional techniques (e.g., assumptions of stationary, homoscedasticity) [32]. We computed the auto recurrence (the amount that a time series recurs with itself at all time lags) of the Z component of the low noise distance channel during recalibration and during reading of each page with a MW probe. This entailed transforming the time series into a distance matrix where each cell captures the pairwise distance between the elements of the time series with itself (Fig. 3). The diagonal is the line of identity (LOI) or the distance between elements that occur at the same point (0 for auto-recurrence). Diagonal lines parallel to the LOI represent points that occur at a particular time lag with lines further from the LOI representing greater time lags. The distance matrix is then transformed to a recurrence matrix with a radius which captures whether two points recur. If the distance of a given cell is less than the radius, then the cell is defined as a recurrent point and a value of 1 is assigned to the cell (Fig. 3). Diagonal lines represent patterned behaviors (e.g., fidgeting) and vertical lines represent smooth or slow changes to movement.

We used the “crqa” package in R [4], set the delay to 0, the embedding dimension to 1 (i.e., no phase space embedding), standardized each time series, and used Euclidean

distance metric. We tested radius values between 0.03 and 0.15 with a step size of 0.005 on a random 10% of participants and selected the value corresponding to approximately 5% mean recurrence. We repeated this procedure with another random 10% of participants, to confirm values. We proceeded with radii of 0.04 and .095 for time series associated with recalibration and MW pages, respectively. We computed the following RQA features from the recurrence matrices: recurrence rate (the fraction of recurrent time points over the total number of points), and features related to vertical and diagonal lines (determinism, laminarity, averaged diagonal line length, trapping time, longest diagonal line, longest vertical line, ratio, entropy) [4].

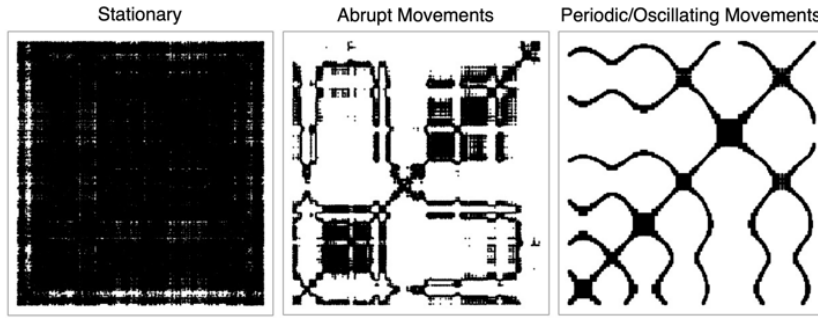


Fig. 3. Example recurrence plots. The first panel represents a stationary system (e.g., no movement), the second panel represents a system with abrupt movements (e.g., stretching), and the third panel represents periodic movements (e.g., fidgeting).

3 Results

3.1 Model comparisons.

All four statistical movement models (all, “matched”, “closest”, “farthest”) computed during intra-task intervals outperformed chance (i.e., AUROCs > 0.5, CIs do not overlap 0.5), whereas the RQA and calibration error models did not (Table 1). The context-only models also performed at chance except for the “farthest” model, but a pairwise ROC test indicated that its accuracy was lower than the “farthest” statistical model ($p = .04$). Pairwise ROC tests between the statistical feature models revealed no significant differences ($ps > 0.05$). Quantitatively, the statistical movement “matched” and “farthest” models yielded the highest AUROCs of 0.62. The best AUROCs for the separate break models (all: 0.53 [0.50, 0.58]) and calibration (matched: 0.59 [0.56, 0.65]) models were significant but lower than for the main model using the entire intra-task interval. Interestingly, the statistical features model during the probed reading page performed at chance, while there was a weak signal for the RQA movement model.

Table 1. AUROCs and 95% bootstrapped CIs from the median performing model. Bolded values indicate models that outperformed chance.

Movement Models	Baseline Models
-----------------	-----------------

Model Type	Statistical	RQA	Context	Calibration Error
Matched	0.62 [0.58, 0.67]	0.46 [0.41, 0.50]	0.48 [0.44, 0.53]	0.49 [0.44, 0.54]
Closest	0.54 [0.50, 0.60]	0.46 [0.41, 0.51]	0.44 [0.39, 0.49]	0.54 [0.49, 0.58]
Farthest	0.62 [0.56, 0.66]	0.50 [0.45, 0.56]	0.55 [0.50, 0.60]	0.48 [0.42, 0.53]
All	0.59 [0.55, 0.62]	0.48 [0.44, 0.51]	0.49 [0.46, 0.53]	0.49 [0.45, 0.52]
Page Of	0.49 [0.49, 0.49]	0.52 [0.51, 0.52]	--	--

3.2 Predictive features.

We used Shapley Additive exPlanations (SHAP) [21] on the median-performing “matched” model as it provides the cleanest mapping of movement during breaks and MW. SHAP gives feature importance and direction of influence. The top three features were kurtosis, median, and minimum of the Z component of distance, capturing flexion and extension of the elbow/arm (e.g., stretching) (Fig. 4). Positive kurtosis (i.e. the peak of the distribution is sharp and the tails are short) was associated with future MW. This might suggest that there were fewer outliers during periods that were followed by MW. Future MW was also associated with a lower median distance. This suggests that when people produced smaller movements during the recalibration break, they mind wandered more later. Finally, the minimum distance was smaller before texts where MW was reported, again suggesting that smaller movements during breaks predict more MW later. Together, these findings suggest that when people produce bigger movements during intra-task intervals via flexion and extension of the arms, later MW is reduced.

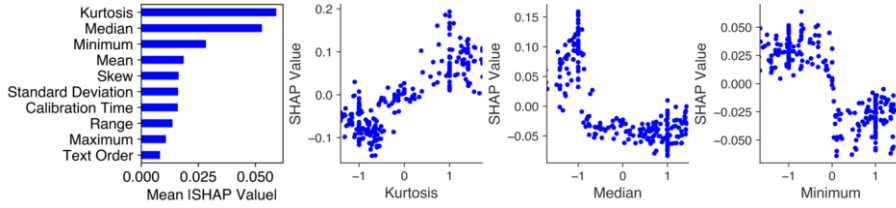


Fig. 4. Feature importances (Panel 1) and directions (Panels 2-4) for the top 3 features. X-axes represent the value of each feature and the Y-axis represents the SHAP value in predicting the probability of MW (values > 0 indicate MW). Each dot represents an instance from the test set.

4 Discussion

The purpose of the present work was to identify whether there are patterns of behavior between task intervals that can prospectively predict MW during later learning. Overall, our results showed a consistent pattern across participants, namely that statistical features computed from an accelerometer time series between task blocks predicted future MW better than the page of MW, a context model, and a model that considered calibration error. Such models were predictive of MW both early during reading (on average 1.7 minutes into each text) and later (after 5.6 min of reading) with a stronger (albeit nonsignificant) effect for the latter. These patterns generalized across

learners. We achieved a highest AUROC of 0.62, while admittedly modest, is consistent with AUROCs of 0.60 obtained by both computer vision algorithms and majority voting by nine human observers in a study detecting MW from visual features during reading [2]. It is lower than AUROCs obtained from eye gaze [12], but this is unsurprising because eye gaze provides a more direct measure of visual attention.

We also found that learners who moved more during intra-task intervals, as indicated by greater flexion and extension of the arms, experienced less MW later, suggesting that how students behave during breaks matters. Indeed, prior work has found that exercise breaks reduce MW during lecture viewing relative to non-exercise breaks [15]. Our findings provide converging evidence and suggest that even in a controlled lab setting, students who moved more tended to MW less during later reading. As participants were seated during the entire task, including the break, these results might be akin to breaks that occur during standardized testing. A surprising finding was that there were no consistent temporal patterns of movement that predicted MW as people read, as prior work has identified a link between movement and concurrent engagement [3, 5, 9, 14, 24, 28]. Although RQA should detect fidgeting behaviors, it might be the case that due to the study constraints and instructions (i.e., participants wore sensors and were asked to minimize movements as they read), participants might have moved less than they would have in more naturalistic reading conditions. Therefore, the present study is likely underestimating the influence of movement during MW.

Applications. The knowledge that there are consistent patterns in movement that can prospectively predict MW has the potential to be incorporated into intelligent learning interfaces. If such a model detected that MW might occur, intelligent technology could suggest to a learner to stretch or walk before completing the next learning activity. If applied to multiple learners, and the prospect of future MW is high, the instructor could be notified that a calisthenic break might be appropriate. Furthermore, a prospective MW detector could be paired with existing proactive and reactive interventions to reduce the negative effects of MW. Specifically, a system might be able to increase the frequency of proactive interventions to decrease MW, such as changing the textual properties [13, 17] or interspersing more test questions [31] if MW is predicted. Similarly, reactive MW detectors could be “seeded” with prospective MW predictions to potentially increase the accuracy of these detectors.

Limitations. As described above, the lab setting might have influenced how learners moved during reading relative to naturalistic reading settings, however participants were encouraged to move at the start of the recalibration, which is the focus of the analyses. Relatedly, because participants were seated the present models might not generalize to all task contexts. For example, the finding that the Z channel, which indexed flexion and extension of the arms, contained the most information (contrary to X and Y which captured horizontal, planar arm movements which might correspond to less naturalistic movements when seated) may only be relevant where learners remain seated during breaks (e.g., standardized testing). However, it is likely the X and Y channels would be more relevant for active breaks, such as walking. It will be critical for future work to test whether the present model generalizes to other break contexts. The outcome measure we predict in the present study is self-reported mind-wandering,

probed after reading on (up to) ten selected pages, and at a coarse, binary level. However, MW is a multifaceted phenomenon [29] and future work should consider using a wider range of MW measures. Other relevant measures of individual differences, e.g. working memory, were not considered here but controlling for these may nuance the findings. In addition, the sample demographics are not representative of the US as a whole [35], with a narrow age range, no Black respondents, and Hispanic participants underrepresented, again highlighting the need to check whether these results generalize.

Conclusion. Although MW has some benefits, multiple studies indicate that it is negatively correlated with learning outcomes (see Introduction), leaving the question open as to how we might mitigate it. Our results suggest that how people move between task blocks is related to how much they will mind wander later. In addition to its theoretical relevance, the ability to prospectively model mind wandering has exciting implications for intelligent learning systems that aim to reduce MW during learning.

5 References

1. Blasche, G. et al.: Comparison of rest-break interventions during a mentally demanding task. *Stress and Health*. 34, 5, 629–638 (2018). <https://doi.org/10.1002/smi.2830>.
2. Bosch, N., D’Mello, S.K.: Can Computers Outperform Humans in Detecting User Zone-Outs? Implications for Intelligent Interfaces. *ACM Trans. Comput.-Hum. Interact.* 29, 2, 1–33 (2022). <https://doi.org/10.1145/3481889>.
3. Carriere, J.S.A. et al.: Wandering in both mind and body: individual differences in mind wandering and inattention predict fidgeting. *Can J Exp Psychol.* 67, 1, 19–31 (2013). <https://doi.org/10.1037/a0031438>.
4. Coco, M.I., Dale, R.: Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Frontiers in Psychology*. 5, (2014).
5. D’Mello, S. et al.: Posture as a Predictor of Learner’s Affective Engagement. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, 29(29). p. 7, Merced, CA (2007).
6. D’Mello, S.K.: Giving Eyesight to the Blind: Towards Attention-Aware AIED. *Int J Artif Intell Educ.* 26, 2, 645–659 (2016). <https://doi.org/10.1007/s40593-016-0104-1>.
7. D’Mello, S.K. et al.: Machine-Learned Computational Models Can Enhance the Study of Text and Discourse: A Case Study Using Eye Tracking to Model Reading Comprehension. *Discourse Processes*. 57, 5–6, 420–440 (2020). <https://doi.org/10.1080/0163853X.2020.1739600>.
8. D’Mello, S.K. et al.: Zone out No More: Mitigating Mind Wandering during Computerized Reading. In: *Proceedings of the 10th International Conference on Educational Data Mining*. International Educational Data Mining Society, Wuhan, China (2017).
9. D’Mello, S.K., Graesser, A.C.: Mining bodily patterns of affective experience during learning. In: *Proceedings of the Third International Conference on Data Mining*. (2010).
10. D’Mello, S.K., Mills, C.S.: Mind wandering during reading: An interdisciplinary and integrative review of psychological, computing, and intervention research and theory. *Language and Linguistics Compass*. 15, 4, e12412 (2021). <https://doi.org/10.1111/lnc3.12412>.
11. Erdaş, Ç.B. et al.: Integrating Features for Accelerometer-based Activity Recognition. *Procedia Computer Science*. 98, 522–527 (2016). <https://doi.org/10.1016/j.procs.2016.09.070>.
12. Faber, M. et al.: An automated behavioral measure of mind wandering during computerized reading. *Behav Res.* 50, 1, 134–150 (2018). <https://doi.org/10.3758/s13428-017-0857-y>.
13. Faber, M. et al.: The effect of disfluency on mind wandering during text comprehension. *Psychon Bull Rev.* 24, 3, 914–919 (2017). <https://doi.org/10.3758/s13423-016-1153-z>.
14. Farley, J. et al.: Everyday attention and lecture retention: the effects of time, fidgeting, and mind wandering. *Front Psychol.* 4, 619 (2013). <https://doi.org/10.3389/fpsyg.2013.00619>.

15. Fenesi, B. et al.: Sweat So You Don't Forget: Exercise Breaks During a University Lecture Increase On-Task Attention and Learning. *Journal of Applied Research in Memory and Cognition*. 7, 2, 261–269 (2018). <https://doi.org/10.1016/j.jarmac.2018.01.012>.
16. Flesch, R.: A new readability yardstick. *Journal of Applied Psychology*. 32, 3, 221–233 (1948).
17. Forrin, N.D. et al.: On the relation between reading difficulty and mind-wandering: a section-length account. *Psychological Research*. 83, 3, 485–497 (2019). <https://doi.org/10.1007/s00426-017-0936-9>.
18. Fox, K.C.R., Christoff, K.: *The Oxford Handbook of Spontaneous Thought: Mind-Wandering, Creativity, and Dreaming*. Oxford University Press (2018).
19. Hutt, S. et al.: Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Model User-Adap Inter*. 29, 4, 821–867 (2019). <https://doi.org/10.1007/s11257-019-09228-5>.
20. Kopp, K.: Identifying Learning Conditions that Minimize Mind Wandering by Modeling Individual Attributes. In: *Intelligent Tutoring Systems*. pp. 94–103 Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-07221-0_12.
21. Lundberg, S.M., Lee, S.-I.: A Unified Approach to Interpreting Model Predictions. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc. (2017).
22. Mills, C. et al.: 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 (2021). <https://doi.org/10.1080/07370024.2020.1716762>.
23. Mooneyham, B.W., Schooler, J.W.: The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology*. 67, 1, 11–18 (2013).
24. Mota, S., Picard, R.W.: Automated Posture Analysis for Detecting Learner's Interest Level. In: *2003 Conference on Computer Vision and Pattern Recognition Workshop*. pp. 49–49 IEEE, Madison, Wisconsin, USA (2003).
25. Nakatani, C. et al.: Context-dependent neural effects predict mind wandering minutes in advance, <http://biorxiv.org/lookup/doi/10.1101/2021.09.04.458977>, (2021). <https://doi.org/10.1101/2021.09.04.458977>.
26. Ravi, N. et al.: Activity Recognition from Accelerometer Data. *American Association for Artificial Intelligence*. 6 (2005).
27. Robin, X. et al.: pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 12, 1, 77 (2011).
28. Seli, P. et al.: Restless mind, restless body. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 40, 3, 660–668 (2014). <https://doi.org/10.1037/a0035260>.
29. Smallwood, J., Schooler, J.W.: The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*. 66, 1, 487–518 (2015). <https://doi.org/10.1146/annurev-psych-010814-015331>.
30. Straczekiewicz, M. et al.: On Placement, Location and Orientation of Wrist-Worn Tri-Axial Accelerometers during Free-Living Measurements. *Sensors (Basel)*. 19, 9, 2095 (2019). <https://doi.org/10.3390/s19092095>.
31. Szpunar, K.K. et al.: Interpolated memory tests reduce mind wandering and improve learning of online lectures. *PNAS*. 110, 16, 6313–6317 (2013).
32. Webber, C., Zbilut, J.: Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*. (2005).
33. Wong, A.Y. et al.: Task-unrelated thought during educational activities: A meta-analysis of its occurrence and relationship with learning. *Contemporary Educational Psychology*. 71, 102098 (2022). <https://doi.org/10.1016/j.cedpsych.2022.102098>.
34. Zhang, H., Jonides, J.: Pre-trial Gaze Stability Predicts Momentary Slips of Attention, <https://psyarxiv.com/bv2uc/>, (2021). <https://doi.org/10.31234/osf.io/bv2uc>.
35. U.S. Census Bureau QuickFacts: United States, <https://www.census.gov/quick-facts/fact/table/US/PST045221>, last accessed 2023/01/14.