Zone out no more: Mitigating mind wandering during computerized reading

Sidney K. D'Mello, Caitlin Mills, Robert Bixler, & Nigel Bosch
University of Notre Dame
118 Haggar Hall
Notre Dame, IN 46556, USA
sdmello@nd.edu

ABSTRACT
Mind wandering, defined as shifts in attention from task-related processing to task-unrelated thoughts, is a ubiquitous phenomenon that has a negative influence on performance and productivity in many contexts, including learning. We propose that next-generation learning technologies should have some mechanism to detect and respond to mind wandering in real-time. Towards this end, we developed a technology that automatically detects mind wandering from eye-gaze during learning from instructional texts. When mind wandering is detected, the technology intervenes by posing just-in-time questions and encouraging re-reading as needed. After multiple rounds of iterative refinement, we summatively compared the technology to a yoked-control in an experiment with 104 participants. The key dependent variable was performance on a post-reading comprehension assessment. Our results suggest that the technology was successful in correcting comprehension deficits attributed to mind wandering (d = .47 sigma) under specific conditions, thereby highlighting the potential to improve learning by "attending to attention."

Keywords
Mind wandering; gaze tracking; student modeling; attention-aware.

1. INTRODUCTION
Despite our best efforts to write a clear and engaging paper, chances are high that within the next 10 pages you might fall prey to what is referred to as zoning out, daydreaming, or mind wandering [45]. Despite your best intention to concentrate on our paper, at some point your attention might drift away to unrelated thoughts of lunch, childcare, or an upcoming trip. This prediction is not based on some negative or cynical opinion of the reader/reviewer (we read and review papers too), but on what is known about attentional control, vigilance, and concentration while individuals are engaged in complex comprehension activities, such as reading for understanding.

One recent study tracked mind wandering of 5,000 individuals from 83 countries with a smartphone app that prompted people with thought-probes at random intervals throughout the day [24]. People reported mind wandering for 46.9% of the prompts, which confirmed lab studies on the pervasiveness of mind wandering (see [45] for a review). Mind wandering is more than merely incidental; a recent meta-analysis of 88 samples indicated a negative correlation between mind wandering and performance across a variety of tasks [34], a correlation which increases with task complexity. When compounded with its high frequency, mind wandering can have serious consequences on the performance and productivity of society at large.

Mind wandering is also unfortunately an under-addressed problem in education and is yet to be deeply studied in the context of learning with technology. Traditional learning technologies rely on the assumption that students are attending to the learning session, although this is not always the case. For example, it has been estimated that students mind wander approximately 40% of the time when engaging with online lectures [38], which are an important component of MOOCs. Some advanced technologies do aim to detect and respond to affective states like boredom, but evidence for their effectiveness is still equivocal (see [9] for a review). Further, boredom is related to but not the same as attention [12]. There are technologies that aim to prevent mind wandering by engendering a highly immersive learning experience and have achieved some success in this regard [40, 41]. But what is to be done when attentional focus inevitably wanes as the session progresses and the novelty of the system and content fades?

Our central thesis is that next-generation learning technologies should include mechanisms to model and respond to learners' attention in real-time [8]. Such attention-aware technologies can model various aspects of learner attention (e.g., divided attention, alternating attention). Here, we focus on detecting and mitigating mind wandering, a quintessential signal of waning engagement. We situate our work in the context of reading because reading is a common activity shared across multiple learning technologies, thereby increasing the generalizability of our results. Further, students mind wander approximately 30% of the time during computerized reading [44]. And although mind wandering can facilitate certain cognitive processes like future planning and divergent thinking [2, 28], it negatively correlates with comprehension and learning (reviewed in [31, 45]), suggesting that it is important to address mind wandering during learning.

Towards this end, we developed and validated a closed-loop attention-aware learning technology that combines a machine-learned mind wandering detector with a real-time interpolated testing and re-study intervention. Our attention-aware technology works as follows. Learners read a text on a computer screen using a self-paced screen-by-screen (also called page-by-page) reading paradigm. We track eye-gaze during reading using a remote eye tracker that does not restrict head movements. We focus on eye-gaze for mind wandering detection due to decades of research suggesting a tight coupling between attentional focus and eye movements during reading [36]. When mind wandering is detected, the system intervenes in an attempt to redirect attentional focus and correct any comprehension deficits that might arise due to mind wandering. The interventions consist of asking comprehension question on pages where mind wandering was detected and providing opportunities to re-read based on learners' responses. In this paper, we discuss the mind wandering detector, intervention approach, and results of a summative evaluation study[1].

1.1 Related Work

The idea of attention-aware user interfaces is not new, but was proposed almost a decade ago by Roda and Thomas [39]. There was even an article on futuristic applications of attention-aware systems in educational contexts [35]. Prior to this, Gluck, et al. [15] discussed the use of eye tracking to increase the bandwidth of information available to an intelligent tutoring system (ITS). Similarly, Anderson [1] followed up on some of these ideas by demonstrating how particular beneficial instructional strategies could only be launched via a real-time analysis of eye gaze.

Most of the recent work has been on leveraging eye gaze to increase the bandwidth of learner models [22, 23, 29]. Conati, et al. [5] provide an excellent review of much of the existing work in this area. We can group the research into three categories: (1) offline-analyses of eye gaze to study attentional processes, (2) computational modeling of attentional states, and (3) closed-loop systems that respond to attention in real-time. Offline-analysis of eye movements has received considerable attention in cognitive and educational psychology for several decades [e.g., 16, 19], so this area of research is relatively healthy. Online computational models of learner attention are just beginning to emerge [e.g., 6, 11], while closed-loop attention-aware systems are few and far between (see [7, 15, 42, 48] for a more or less exhaustive list). Two known examples, GazeTutor and AttentiveReview, are discussed below.

GazeTutor [7] is a learning technology for biology. It has an animated conversational agent that provides spoken explanations on biology topics which are synchronized with images. The system uses a Tobii T60 eye tracker to detect inattention, which is assumed to occur when learners' gaze is not on the tutor agent or image for at least five consecutive seconds. When this occurs, the system interrupts its speech mid utterance, directs learners to reorient their attention (e.g., "I'm over here you know"), and repeats speaking from the start of the current utterance. In an evaluation study, 48 learners (undergraduate students) completed a learning session on four biology topics with the attention-aware components enabled (experimental group) or disabled (control group). The results indicated that GazeTutor was successful in dynamically reorienting learners' attentional patterns towards the interface. Importantly, learning gains for deep reasoning questions were significantly higher for the experimental vs. control group, but only for high aptitude learners. The results suggest that even the most basic attention-aware technology can be effective in improving learning, at least for a subset of learners. However, a key limitation is that the researchers simply assumed that off-screen gaze corresponded to inattention, but did not test this assumption (e.g., students could have been concentrating with their eyes closed and this would have been perceived as being inattentive).

AttentiveReview [32] is a closed-loop system for MOOC learning on mobile phones. The system uses video-based photoplethysmography (PPG) to detect a learners' heart rate from the back camera of a smartphone while they view MOOC-like lectures on the phone. AttentiveReview ranks the lectures based on its estimates of learners' "perceived difficulty," selecting the most difficult lecture for subsequent review (called adaptive review). In a 32-participant between-subjects evaluation study, the authors found that learning gains obtained from the adaptive review condition were statistically on par with a full review condition, but were achieved in 66.7% less review time. Although this result suggests that AttentiveReview increased learning efficiency, there is the question as to whether the system should even be considered to be an "attention-aware" technology. This is because it is arguable if the system has anything to do with attention (except for "attention" appearing in its name) as it selects items for review based on a model of "perceived difficulty" and not on learners' "attentional state." The two might be related, but are clearly not the same.

1.2 Novelty

Our paper focuses on closing the loop between research on educational data and learning outcomes by developing and validating the first (in our view) real-time learning technology that detects and mitigates mind wandering during computerized reading. Although automated detection of complex mental states with the goal of developing intelligent learning technologies that respond to the sensed states is an active research area (see reviews by [9, 18]), mind wandering has rarely been explored as an aspect of a learner's mental state that warrants detection and corrective action. And while there has been some work on modeling the locus of learner attention (see review by [5]), mind wandering is inherently different than more commonly studied forms of attention (e.g., selective attention, distraction), because it involves more covert forms of involuntary attentional lapses spawned by self-generated internal thought [45]. Simply put, mind wandering is a form of "looking without seeing" because the eyes might be fixated on the appropriate external stimulus, but very little is being processed as the mind is consumed by stimulus-independent internal thoughts. *Offline* automated approaches to detect mind wandering have been developed (e.g., [3, 11, 27, 33]), but these detectors have not yet been used to trigger *online* interventions. Here, we adapt an offline gaze-based automated mind wandering detector [13] to trigger real-time interventions to address mind wandering during reading. We conduct a randomized control trial to evaluate the efficacy of our attention-aware learning technology in improving learning.

2. MIND WANDERING DETECTION

We adopted a supervised learning approach for mind wandering detection. Below we provide a high-level overview of the approach; readers are directed to [3, 13] for a detailed discussion of the general approach used to build gaze-based detectors of mind wandering.

2.1 Training Data

We obtained training data from a previous study [26] that involved 98 undergraduate students reading a 57-page text on the surface tension of liquids [4] on a computer screen for an average of 28 minutes. The text contained around 6500 words, with an average of 115 words per page, and was displayed on a computer screen with Courier New typeface. We recorded eye-gaze with a Tobii TX300 eye tracker set to a sampling frequency of 120 Hz.

[1] This paper reports updated results of an earlier version [10] presented as a "Late-Breaking Work" (LBW) poster at the *2016 ACM CHI conference*. LBW "Extended Abstracts" are not included in the main conference proceedings and copyright is retained by the authors.

Participants could read normally and were free to move or gesture as they pleased.

Participants were instructed to report mind wandering (during reading) by pressing a predetermined key when they found themselves "thinking about the task itself but *not the actual content of the text*" or when they were "thinking about *anything else besides the task*." This is consistent with contemporary approaches (see [45]) that rely on self-reporting because mind wandering is an internal conscious phenomena. Further, self-reports of mind wandering have been linked to predictable patterns in physiology [43], pupillometry [14], eye-gaze [37], and task performance [34], providing validity for this approach.

On average, we received mind wandering reports for 32% of the pages *(SD* = 20%), although there was considerable variability among participants (ranging from 0% to 82%). Self-reported mind wandering negatively correlated (r = -.23, p < .05) with scores on a subsequent comprehension assessment [26], which provides evidence for the predictive validity of the self-reports.

2.2 Model Building

The stream of eye-gaze data was filtered to produce a series of fixations, saccades, and blinks, from which *global eye gaze* features were extracted (see Figure 1). Global features are independent of the words being read and are therefore more generalizable than so-called local features. A full list of 62 global features along with detailed descriptions is provided in [13], but briefly the features can be grouped into the following four categories: (1) Eye movement descriptive features (n = 48) were statistical functionals (e.g., min, median) for fixation duration, saccade duration, saccade amplitude, saccade velocity, and relative and absolute saccade angle distributions; (2) Pupil diameter descriptive features were statistical functionals (n = 8) computed from participant-level z-score standardized estimates of pupil diameter; (3) Blink features (n = 2) consisted of the number of blinks and the mean blink duration; (4) Miscellaneous gaze features (n = 4) consisted of the number of saccades, horizontal saccade proportion, fixation dispersion, and the fixation duration/saccade duration ratio. We proceeded with a subset of 32 features after eliminating features exhibiting multicollinearity.

Features were calculated from only a certain amount of gaze data from each page, called the *window*. The end of the window was positioned 3 seconds before a self-report so as to not overlap with the key-press. The average amount of time between self-reports and the beginning of the page was 16 seconds. We used this time point as the end of the window for pages with no self-report. Pages that were shorter than the target window size were discarded, as were pages with windows that contained fewer than five gaze fixations as there was insufficient data to compute some of the features. There were a total of 4,225 windows with sufficient data for supervised classification.

We experimented with a number of supervised classifiers on window sizes of 4, 8, and 12 seconds to discriminate positive (pages with a self-report = 32%) from negative (pages without a self-report) instances of mind wandering. The training data were downsampled to achieve a 50% base rate; testing data were unaltered. A leave-one-participant-out validation approach was adopted where models were built on data from *n-1* participants and evaluated on the held-out participant. The process was repeated for all participants. Model validation was conducted in a way to simulate a real-time system by analyzing data from every page. When classification was not possible due to a lack of valid gaze data and/or because participants did not spend enough time on the page, we classified the page as a positive instance of mind wandering. This was done because analyses indicated that participants were more likely to be mind wandering in those cases (but see [13] for alternate strategies to handle missing instances).

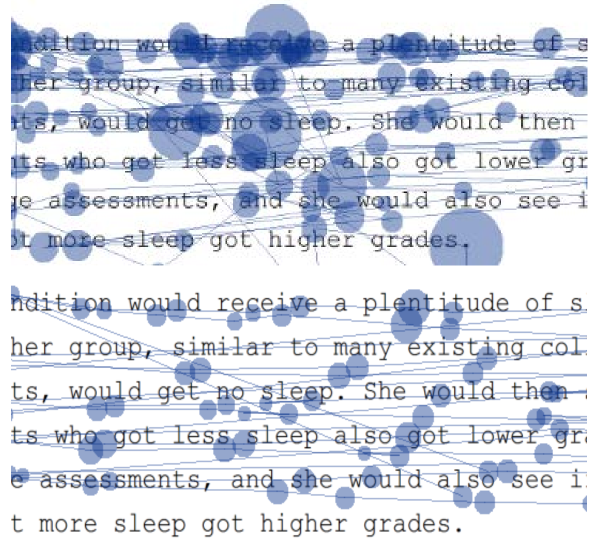

Figure 1: Gaze fixations during mind wandering (top) and normal reading (bottom)

2.3 Detector Accuracy

The best model was a support vector machine that used global features and operated on a window size of 8-seconds. The area under the ROC curve (AUC or AUROC or A') was .66, which exceeds the 0.5 chance threshold [17].

We assigned each instance as mind wandering or not mind wandering based on whether the detector's predicted likelihood of mind wandering (ranges from 0 to 1) was below or above 0.5 We adopted the default 0.5 threshold as it led to a higher rate of true positives while maintaining a moderate rate of true negatives. This resulted in the following confusion matrix shown in Table 1. The model had a weighted precision of 72.2% and a weighted recall of 67.4%, which we deemed to be sufficiently accurate for intervention.

Table 1: Proportionalized confusion matrix for mind wandering detection

	Predicted mind wandering (MW)	
Actual MW	*yes*	*no*
yes	0.715 (hit)	0.285 (miss)
no	0.346 (false positive)	0.654 (correct rejection)

3. Intervention to Address Mind Wandering

Our intervention approach is grounded in the basic idea that learning of conceptual information involves creating and maintaining an internal model (*mental model*) by integrating information from the text with prior knowledge from memory [25]. This integration process relies on attentional focus and breaks down during mind wandering because information from the external environment is no longer being integrated into the internal mental model. This results in an impaired model which leads to less effective suppression of off-task thoughts. This increase in mind wandering further impairs the mental model, resulting in a vicious cycle. Our intervention targets this vicious cycle by redirecting attention to the primary task and attempting to correct for comprehension deficits attributed to mind wandering. Based on research demonstrating the effectiveness of interpolated testing [47], we propose that asking questions on pages where mind wandering is detected and encouraging re-reading in response to incorrect responses will aid in re-directing attention to the text and correct knowledge deficits.

3.1 Intervention Implementation

Our initial intervention was implemented for the same text used to create the mind wandering detector (although it could be applied to any text). The text was integrated into the computer reading interface. Mind wandering detection occurred when the learner navigated to the next page using the right arrow key. In order to address ambiguity in mind wandering detection, we used the detector's mind wandering likelihood to probabilistically determine when to intervene. For example, if the mind wandering likelihood was 70%, then there was a 70% chance of intervention on any given page (all else being equal). We did not intervene for the first three pages in order to allow the learner to become familiar with the text and interface. To reduce disruption, there was a 50% reduced probability of intervening on adjacent pages, and the maximum number of interventions was capped at 1/3 × the number of pages (19 for the present 57-page text). Table 2 presents pseudo code for when to launch an intervention.

Table 2: Pseudo code for intervention strategy

```
launch_intervention:
    if current_page >= WAITPAGES
    and
        total_interventions < MAXINTRV)
    and
        gaze_likelihood > random(0,1)
    and
        (!has_intervened(previous_page)
        or 0.5 < random (0,1)):
            do_intervention()
    else:
        show_next_page()

do_intervention:
    answer1 = show_question1()
    if answer1 is correct:
        show_positive_feedback()
        show_next_page()
    else:
        show_neg_feedback()
        suggest_rereading()
        if page advance detected:
            answer2 = show_question2();
            show_next_page()
```

Figure 2 presents an outline of the intervention strategy. The intervention itself relied on two multiple choice questions for each page (screen) of the text. When the system decided to intervene, one of the questions (randomly selected) was presented to the learner. If the learner answered this *online question* correctly, positive feedback was provided, and the learner could advance to the next page. If the learner answered incorrectly, negative feedback was provided, and the system encouraged the learner to re-read the page. The learner was then provided with a second (randomly selected) online question, which could either be the same or the alternate question for that page. Feedback was not provided and the learner was allowed to advance to the next page regardless of whether the second question was answered correctly, so as not to be overly burdensome.

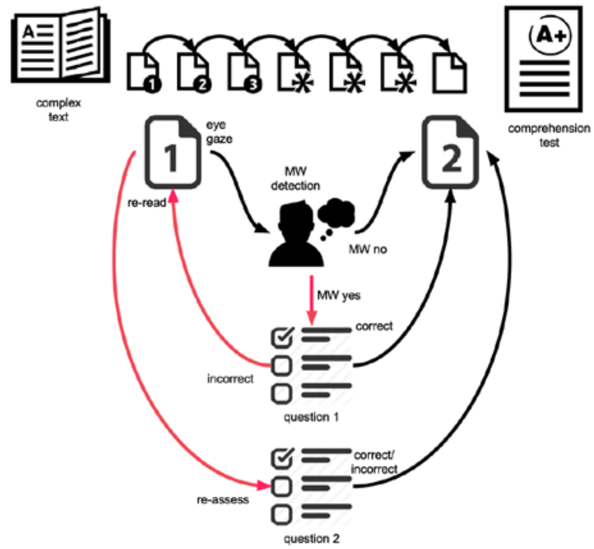

Figure 2: Outline of intervention strategy

3.2 Iterative Refinement

The technology was refined through multiple rounds of formative testing with 67 participants, recruited from the same institution used to build the detector. Participants were observed while interacting with the technology, their responses were analyzed, and they were interviewed about their experience. We used the feedback gleaned from these tests to refine the intervention parameters (i.e., when to launch, how many interventions to launch, whether to launch interventions on subsequent pages), intervention questions themselves, and instructions on how to attend to the intervention. For example, earlier versions of the intervention used a fixed threshold (instead of the aforementioned probabilistic approach) to trigger an intervention. Despite many attempts to set this threshold, the end result was that some participants received many interventions while others received almost no interventions. This issue was corrected by probabilistically rather than deterministically launching the intervention. Additional testing/refinement of the comprehension questions used in the intervention was done using crowdsourcing platforms, specifically Amazon's Mechanical Turk (MTurk).

4. Evaluation Study

We conducted a randomized controlled trial to evaluate the technology. The experiment had two conditions: an intervention condition and a yoked control condition (as described below). The yoked control was needed to verify that any learning benefits are attributed to the technology being sensitive to mind wandering and not merely to the added opportunities to answer online questions and re-read. This is because we know that interpolated testing itself has beneficial comprehension effects [47].

4.1 Method

Participants (N = 104) were a new set of undergraduate students who participated to fulfill research credit requirements. They were recruited from the same university used to build the MW detector and for the iterative testing and refinement cycles.

We did not use a pretest because we expected participants to be unfamiliar with the topic. Participants were not informed that the interface would be tracking their mind wandering (until the debriefing at the end), Instead, they were instructed as follows: “While reading the text, you will occasionally be asked some questions about the page you just read. Depending on your answer, you will re-read the same page and you will be asked another question that may or may not be the same question.”

Participants in the intervention condition received the intervention as described above (i.e., based on detected mind wandering likelihoods). Each participant in the yoked control condition was *paired* with a participant in the intervention condition. He or she received an intervention question on the same pages as their paired intervention participant regardless of mind wandering likelihood. For example, if participant A (i.e., intervention condition) received questions on pages 5, 7, 10, and 25, participant B (i.e., yoked control condition) would receive intervention questions on the same pages. However, if the yoked participant answered incorrectly, then (s)he had the opportunity to re-read and answer another question regardless of the outcome of their intervention-condition partner.

After reading, participants completed a 38-item multiple choice comprehension assessment to measure learning. The questions were randomly selected from the 57 pages (one per page) with the exception that a higher selection priority was given to pages that were re-read on account of the intervention. Participants in the yoked control condition received the same posttest questions as their intervention condition counterparts.

4.2 Results

Participants received an average of 16 (min of 7 and max of 19) interventions. They spent an average of 27.5 seconds on each screen prior to receiving an intervention. There was no significant difference across conditions (p = .998), suggesting that reading time was not a confound. In what follows, we compared each intervention participant to his/her yoked control with a two-tailed paired-samples t-test and a 0.05 criteria for statistical significance.

Mind wandering detection. The detector’s likelihood of mind wandering was slightly higher for participants in the yoked-control condition (M = .431; SD = .170) compared to the intervention condition (M = .404; SD = .112), but the difference was not statistically significant (p = .348). This was unsurprising as participants in both groups received the same interventions, which itself was expected to reduce mind wandering. Importantly, mind wandering likelihoods were negatively correlated with performance on the online questions (r = -.296, p = .033) as well as on posttest questions (r = -.319, p = .021). This provides evidence for the validity of the mind wandering detector when applied to a new set of learners and under different conditions (i.e., reading interspersed with online questions compared to uninterrupted reading).

Comprehension assessment. There was some overlap between the online questions and the posttest questions. To obtain an unbiased estimate of learning, we only analyzed performance on previously unseen posttest questions. That is, questions that were used as part of the intervention were first removed before computing posttest scores.

There were no significant condition differences on overall posttest scores (p = .846). The intervention condition answered 57.6% (SD = .157) of the questions correctly while the yoked control condition answered 58.1% (SD = .129) correctly. This finding was not surprising as both conditions received the exact same treatment except that the interventions were triggered based on detected mind wandering in the intervention condition but not the control condition.

Next, we examined posttest performance as a function of mind wandering during reading. Each page was designated as a low or high mind wandering page based on a median split of mind wandering likelihoods (medians = .35 and .36 on a 0 to 1 scale for intervention and control conditions, respectively). We then analyzed performance on posttest questions corresponding to pages with low vs. high likelihoods of mind wandering (during reading). The results are shown in Table 3.

We found no significant posttest differences on pages where both the intervention and control participants had low (p = .759) or high (p = .922) mind wandering likelihoods (first and last rows in Table 3, respectively). There was also no significant posttest difference (p = .630) for pages where the intervention condition had high mind wandering likelihoods but the control condition had low mind wandering likelihoods (row 3). However, the intervention condition significantly (p = .003, d = .47 sigma) outperformed the control condition for pages where the intervention participants had low likelihoods of mind wandering but control participants had high mind wandering likelihoods (row 2). These last two finding suggests that the intervention had the intended effect of reducing comprehension deficits attributable to mind wandering because it led to equitable performance when mind wandering was high and improved performance when it was low.

Table 3: Posttest performance (proportion of correct responses) as a function of mind wandering during reading. Standard deviations in parenthesis.

	Mind wandering		**Posttest scores**	
N	*Int.*	*Cntrl.*	*Int.*	*Cntrl.*
43	Low	Low	.604 (.288)	.623 (.287)
40	**Low**	**High**	**.643 (.263)**	**.489 (.298)**
43	High	Low	.535 (.295)	.566 (.305)
45	High	High	.522 (.312)	.515 (.291)

Note. Int. = intervention. Cntrl. = control. Bolded cells represent a statistically significant difference. N = number of pairs (out of 52) in each analysis. It differs slightly across analyses as not all participants were assigned to each mind wandering group.

After-task interview. We interviewed a subset of the participants in order to gauge their subjective experience with the intervention. A few key themes emerged. Participants reported paying closer attention to the text after realizing they would be periodically answering multiple-choice questions. This was good. However, participants also reported that they adapted their reading strategies in one of two ways in response to the questions. Since the questions targeted factual information (sometimes verbatim) from the text, some participants paid more attention to details and precise wordings instead of the broader concepts being discussed in the text. More discouragingly, some participants reported adopting a preemptive skimming strategy in that they would only look for keywords that they expected to appear in a subsequent question.

Participants were encouraged to re-read text when they answered incorrectly before receiving another question (or the same question in some cases). Many participants reported simply scanning the text (when re-reading) to locate keywords from the question before moving on. Since the scanning strategy was often successful to answer the subsequent question, participants reported that the questions were too easy and it took relatively little effort to locate the correct answer compared to re-reading. They suggested that it may have been better if the questions had targeted key concepts rather than facts.

Finally, participants reported difficulties with re-engaging with the text after answering an online question because the text was cleared when an intervention question was displayed; an item that can be easily corrected in subsequent versions.

5. Discussion

We developed the first educational technology capable of real-time mind wandering detection and dynamic intervention during computerized reading. In the remainder of this section, we discuss the significance of our main findings, limitations, and avenues for future work.

5.1 Significance of Main Findings

We have three main findings. First, we demonstrated that a machine-learned mind wandering detector built in one context can be applied to a different (albeit related) interaction context. Specifically, the detector was trained on a data set involving participants silently reading and self-reporting mind wandering, but was applied to an interactive context involving interpolated assessments, which engendered different reading strategies. Further, self-reports of mind wandering were *not* collected in this interactive context, which might have influenced mind wandering rates in and of itself. Despite these differences, we were able to demonstrate the predictive validity of the detector by showing that it negatively correlated with both online and offline comprehension scores when evaluated on new participants.

Second, we showed promising effects for our intervention approach despite a very conservative experimental design, which ensured that the intervention and control groups were equated along all respects, except that the intervention was triggered based on the mind wandering detector (key manipulation). Further, we used a probabilistic approach to trigger an intervention, because the detector is inherently imperfect. As a result, participants could have received an intervention when they were not mind wandering and/or could have failed to receive one when they were mind wandering. Therefore, it was essential to compare the two groups under conditions when the mind wandering levels differed. This more nuanced analysis revealed that although the intervention itself did not lead to a boost in overall comprehension (because it is remedial), it equated comprehension scores when mind wandering was high (i.e., scores for the intervention group were comparable when the control group was low on mind wandering). It also demonstrated the cost of not intervening during mind wandering (i.e., scores for the intervention group were greater when the control group was high on mind wandering). In other words, the intervention was successful in mitigating the negative effects of mind wandering.

Third, despite the advantages articulated above, the intervention itself was reactive and engendered several unintended (and presumably suboptimal) behaviors. In particular, students altered their reading strategies in response to the interpolated questions, which were a critical part of the intervention. In a sense, they attempted to "game the intervention" by attempting to proactively predict the types of questions they might receive and then adopting a complementary reading strategy consisting of skimming and/or focusing on factual information. This reliance on surface- rather than deeper-levels of processing was incongruent with our goal of promoting deep comprehension.

5.2 Limitations

There are a number of methodological limitations with this work that go beyond limitations with the intervention (as discussed above). First, we focused on a single text that is perceived as being quite dull and consequently triggers rather high levels of mind wandering [26]. This raises the question of whether the detector will generalize to different texts. We expect some level of generalizability in terms of features used because the detector only used content- and position- (on the screen) free global gaze features. However, given that several supervised classifiers are very sensitive to differences in base rates, the detector might over- or under- predict mind wandering when applied to texts that engender different rates of mind wandering. Therefore, retraining the detector with a more diverse set of texts is warranted.

Another limitation is the scalability of our learning technology. The eye tracker we used was a cost-prohibitive Tobii TX300 that will not scale beyond the laboratory. Fortunately, commercial-off-the-shelf (COTS) eye trackers, such as Eye Tribe and Tobii EyeX, can be used to surpass this limitation. It is an open question as to whether the mind wandering detector can operate with similar fidelity with these COTS eye trackers. Our use of global gaze features which do not require high-precision eye tracking holds considerable promise in this regard. Nevertheless, replication with scalable eye trackers and/or scalable alternatives to eye tracking (e.g., facial-feature tracking [46] or monitoring reading patterns [27]) is an important next step (see Section 5.3).

Our use of surface-level questions for both the intervention and the subsequent comprehension assessment is also a limitation as is the lack of a delayed comprehension assessment. It might be the case that the intervention effects manifest as richer encodings in long-term memory, a possibility that cannot be addressed in the current experiment that only assessed immediate learning.

Other limitations include a limited student sample (i.e. undergraduates from a private Midwestern college) and a laboratory setup. It is possible that the results would not generalize to a more diverse student population or in more ecological environments (but see below for evidence of generalizability of the detector in classroom environments). Replication with data from more diverse populations and environments would be a necessary next step to increase the ecological validity of this work.

5.3 Future Work

Our future work is progressing along two main fronts. One is to address limitations in the intervention and design of the experimental evaluation as discussed above. Accordingly, we are exploring alternative intervention strategies, such as: (a) tagging items for future re-study rather than interrupting participants during reading; (b) highlighting specific portions of the text as an overt cue to facilitate comprehension of critical information; (c) asking fewer intervention questions, but selecting inference questions that target deeper levels of comprehension and that span multiple pages of the text; and (d) asking learners to engage in reflection by providing written self-explanations of the textual content. We are currently evaluating one such redesigned intervention – open-ended questions targeting deeper levels of comprehension (item c). Our revised experimental design taps both surface- and inference-level comprehension and assesses comprehension immediately after reading (to measure learning) and after a one-week delay (to measure retention).

We are also developing attention-aware versions of more interactive interfaces, such as learning with an intelligent tutoring system called GuruTutor [30]. This project also addresses some of the scalability concerns by replacing expensive research-grade eye tracking with cost-effective COTS eye tracking (e.g., the Eye Tribe or Tobii EyeX) and provides evidence for real-world generalizability by collecting data in classrooms rather than the lab. We recently tested our implementation on 135 students (total) in a noisy computer-enabled high-school classroom where eye-gaze of entire classes of students was collected during their normal class periods [20]. Using a similar approach to the present work, we used the data to build and validate a student-independent gaze-based mind wandering detector. The resultant mind wandering detection accuracy (F_1 of 0.59) was substantially greater than chance (F_1 of 0.24) and outperformed earlier work on the same domain [21]. The next step is to develop interventions that redirect attention and correct learning deficiencies attributable to mind wandering and to test the interventions in real-world environments. By doing so, we hope to advance our foundational vision of developing next-generation technologies that enhance the process and products of learning by "attending to attention."

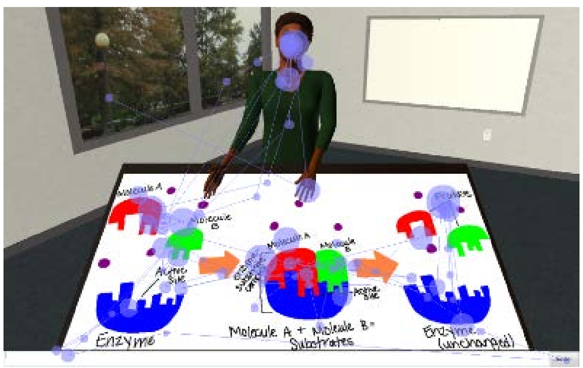

Figure 3: Guru Tutor interface overlaid with eye-gaze obtained via the EyeTribe

6. Acknowledgements

This research was supported by the National Science Foundation (NSF) (DRL 1235958 and IIS 1523091). The authors are grateful to Kris Kopp and Jenny Wu for their contributions to the study. Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

7. REFERENCES

[1] Anderson, J.R. 2002. Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science*, *26* (1), 85-112.

[2] Baird, B., Smallwood, J., Mrazek, M.D., Kam, J.W., Franklin, M.S. and Schooler, J.W. 2012. Inspired by distraction mind wandering facilitates creative incubation. *Psychological Science*, *23* (10), 1117-1122.

[3] Bixler, R. and D'Mello, S.K. 2016. Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling & User-Adapted Interaction*, *26*, 33-68.

[4] Boys, C.V. 1895. *Soap bubbles, their colours and the forces which mold them*. Society for Promoting Christian Knowledge.

[5] Conati, C., Aleven, V. and Mitrovic, A. 2013. Eye-Tracking for Student Modelling in Intelligent Tutoring Systems. In Sottilare, R., Graesser, A., Hu, X. and Holden, H. eds. *Design Recommendations for Intelligent Tutoring Systems - Volume 1: Learner Modeling*, Army Research Laboratory, Orlando, FL.

[6] Conati, C. and Merten, C. 2007. Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. *Knowledge-Based Systems*, *20* (6), 557-574.

[7] D'Mello, S., Olney, A., Williams, C. and Hays, P. 2012. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies*, *70* (5), 377-398.

[8] D'Mello, S.K. 2016. Giving Eyesight to the Blind: Towards attention-aware AIED. *International Journal of Artificial Intelligence In Education*, *26* (2), 645-659.

[9] D'Mello, S.K., Blanchard, N., Baker, R., Ocumpaugh, J. and Brawner, K. 2014. I feel your pain: A selective review of affect-sensitive instructional strategies. In Sottilare, R., Graesser, A., Hu, X. and Goldberg, B. eds. *Design Recommendations for Adaptive Intelligent Tutoring Systems: Adaptive Instructional Strategies (Volume 2)*, US Army Research Laboratory, Orlando, FL.

[10] D'Mello, S.K., Kopp, K., Bixler, R. and Bosch, N. 2016. Attending to attention: Detecting and combating mind wandering during computerized reading In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2016)*, ACM, New York.

[11] Drummond, J. and Litman, D. 2010. In the zone: Towards Detecting student zoning out using supervised machine learning. In Aleven, V., Kay, J. and Mostow, J. eds. *Intelligent Tutoring Systems.*, Springer-Verlag, Berlin / Heidelberg.

[12] Eastwood, J.D., Frischen, A., Fenske, M.J. and Smilek, D. 2012. The unengaged mind: Defining boredom in terms of attention. *Perspectives on Psychological Science*, *7* (5), 482-495.

[13] Faber, M., Bixler, R. and D'Mello, S.K. in press. An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*.

[14] Franklin, M.S., Broadway, J.M., Mrazek, M.D., Smallwood, J. and Schooler, J.W. 2013. Window to the Wandering Mind: Pupillometry of Spontaneous Thought While Reading. *The Quarterly Journal of Experimental Psychology*, *66* (12), 2289-2294.

[15] Gluck, K.A., Anderson, J.R. and Douglass, S.A. 2000. Broader Bandwidth in Student Modeling: What if ITS Were "Eye" TS? In Gauthier, C., Frasson, C. and VanLehn, K. eds. *Proceedings of the 5th international conference on intelligent tutoring systems*, Springer, Berlin.

[16] Graesser, A., Lu, S., Olde, B., Cooper-Pye, E. and Whitten, S. 2005. Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition*, *33*, 1235-1247.

[17] Hanley, J.A. and McNeil, B.J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143* (1), 29-36.

[18] Harley, J.M., Lajoie, S.P., Frasson, C. and Hall, N.C. in press. Developing Emotion-Aware, Advanced Learning Technologies: A Taxonomy of Approaches and Features. *International Journal of Artificial Intelligence In Education*.

[19] Hegarty, M. and Just, M. 1993. Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32* (6), 717-742.

[20] Hutt, S., Mills, C., Bosch, N., Krasich, K., Brockmole, J.R. and D'Mello, S.K. in review. Out of the Fr-Eye- ing Pan: Towards Gaze-Based Models of Attention during Learning with Technology in the Classroom.

[21] Hutt, S., Mills, C., White, S., Donnelly, P.J. and D'Mello, S.K. 2016. The Eyes Have It: Gaze-based Detection of Mind Wandering during Learning with an Intelligent Tutoring System. In *Proceedings of the 9th International Conference on Educational Data Mining (EDM 2016)*, International Educational Data Mining Society.

[22] Jaques, N., Conati, C., Harley, J.M. and Azevedo, R. Year. Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System. In *Intelligent Tutoring Systems*, (2014), Springer, 29-38.

[23] Kardan, S. and Conati, C. 2012. Exploring gaze data for determining user learning with an interactive simulation. In Carberry, S., Weibelzahl, S., Micarelli, A. and Semeraro, G. eds. *Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization (UMAP 2012)*, Springer, Berlin.

[24] Killingsworth, M.A. and Gilbert, D.T. 2010. A wandering mind is an unhappy mind. *Science*, *330* (6006), 932-932.

[25] Kintsch, W. 1998. *Comprehension: A paradigm for cognition*. Cambridge University Press, New York.

[26] Kopp, K., D'Mello, S. and Mills, C. 2015. Influencing the occurrence of mind wandering while reading. *Consciousness and Cognition*, *34* (1), 52-62.

[27] Mills, C. and D'Mello, S.K. 2015. Toward a Real-time (Day) Dreamcatcher: Detecting Mind Wandering Episodes During Online Reading. In Romero, C., Pechenizkiy, M., Boticario, J. and Santos, O. eds. *Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015)*, International Educational Data Mining Society.

[28] Mooneyham, B.W. and Schooler, J.W. 2013. The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67* (1), 11.

[29] Muir, M. and Conati, C. 2012. An analysis of attention to student–adaptive hints in an educational game. In Cerri, S.A., Clancey, W.J., Papadourakis, G. and Panourgia, K. eds. *Proceedings of the International Conference on Intelligent Tutoring Systems*, Springer, Berlin.

[30] Olney, A., D'Mello, A., Person, N., Cade, W., Hays, P., Williams, C., Lehman, B. and Graesser, A. 2012. Guru: A computer tutor that models expert human tutors. In Cerri, S., Clancey, W., Papadourakis, G. and Panourgia, K. eds. *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*, Springer-Verlag, Berlin/Heidelberg.

[31] Olney, A., Risko, E.F., D'Mello, S.K. and Graesser, A.C. 2015. Attention in educational contexts: The role of the learning task in guiding attention. In Fawcett, J., Risko, E.F. and Kingstone, A. eds. *The Handbook of Attention*, MIT Press, Cambridge, MA.

[32] Pham, P. and Wang, J. 2016. Adaptive Review for Mobile MOOC Learning via Implicit Physiological Signal Sensing. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016)*, ACM, New York, NY.

[33] Pham, P. and Wang, J. 2015. AttentiveLearner: improving mobile MOOC learning via implicit heart rate tracking. In *International Conference on Artificial Intelligence in Education*, Springer, Berlin Heidelberg.

[34] Randall, J.G., Oswald, F.L. and Beier, M.E. 2014. Mind-wandering, cognition, and performance: A theory-driven meta-analysis of attention regulation. *Psychological Bulletin*, *140* (6), 1411-1431.

[35] Rapp, D.N. 2006. The value of attention aware systems in educational settings. *Computers in Human behavior*, *22* (4), 603-614.

[36] Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124* (3), 372-422.

[37] Reichle, E.D., Reineberg, A.E. and Schooler, J.W. 2010. Eye movements during mindless reading. *Psychological Science*, *21* (9), 1300.

[38] Risko, E.F., Buchanan, D., Medimorec, S. and Kingstone, A. 2013. Everyday attention: mind wandering and computer use during lectures. *Computers & Education*, *68* (1), 275-283.

[39] Roda, C. and Thomas, J. 2006. Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior*, *22* (4), 557-587.

[40] Rowe, J., Mott, B., McQuiggan, S., Robison, J., Lee, S. and Lester, J. Year. Crystal island: A narrative-centered learning environment for eighth grade microbiology. In *Workshop on Intelligent Educational Games at the 14th International Conference on Artificial Intelligence in Education, Brighton, UK*, (2009), 11-20.

[41] Shute, V.J., Ventura, M., Bauer, M. and Zapata-Rivera, D. 2009. Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In Ritterfeld, U., Cody, M. and Vorderer, P. eds. *Serious games: Mechanisms and effects*, Routledge, Taylor and Francis, Mahwah, NJ.

[42] Sibert, J.L., Gokturk, M. and Lavine, R.A. 2000. The reading assistant: eye gaze triggered auditory prompting for reading remediation. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, ACM, New York, NY.

[43] Smallwood, J., Davies, J.B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R. and Obonsawin, M. 2004. Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, *13* (4), 657-690.

[44] Smallwood, J., Fishman, D.J. and Schooler, J.W. 2007. Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review*, *14* (2), 230-236.

[45] Smallwood, J. and Schooler, J.W. 2015. The science of mind wandering: empirically navigating the stream of consciousness. *Annu. Rev. Psychol*, *66*, 487-518.

[46] Stewart, A., Bosch, P., Chen, H., Donnelly, P.J. and D'Mello, S.K. 2016. Where's Your Mind At? Video-Based Mind Wandering Detection During Film Viewing. In Aroyo, L., D'Mello, S., Vassileva, J. and Blustein, J. eds. *Proceedings of the 2016 ACM on International Conference on User Modeling, Adaptation, & Personalization (ACM UMAP 2016)*, ACM, New York.

[47] Szpunar, K.K., Khan, N.Y. and Schacter, D.L. 2013. Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110* (16), 6313-6317.

[48] Wang, H., Chignell, M. and Ishizuka, M. 2006. Empathic tutoring software agents using real-time eye tracking. In *Proceedings of the 2006 symposium on Eye tracking research &applications*, ACM, New York.