

# An automated behavioral measure of mind wandering during computerized reading

Myrthe Faber<sup>1</sup>  · Robert Bixler<sup>2</sup> · Sidney K. D'Mello<sup>1,2</sup>

Published online: 8 February 2017  
© Psychonomic Society, Inc. 2017

**Abstract** Mind wandering is a ubiquitous phenomenon in which attention shifts from task-related to task-unrelated thoughts. The last decade has witnessed an explosion of interest in mind wandering, but research has been stymied by a lack of objective measures, leading to a near-exclusive reliance on self-reports. We addressed this issue by developing an eye-gaze-based, machine-learned model of mind wandering during computerized reading. Data were collected in a study in which 132 participants reported self-caught mind wandering while reading excerpts from a book on a computer screen. A remote Tobii TX300 or T60 eyetracker recorded their gaze during reading. The data were used to train supervised classification models to discriminate between mind wandering and normal reading in a manner that would generalize to new participants. We found that at the point of maximal agreement between the model-based and self-reported mind-wandering means (smallest difference between the group-level means:  $M_{\text{model}} = .310$ ,  $M_{\text{self}} = .319$ ), the participant-level mind-wandering proportional distributions were similar and were significantly correlated ( $r = .400$ ). The model-based estimates were internally consistent ( $r = .751$ ) and predicted text comprehension more strongly than did self-reported mind wandering ( $r_{\text{model}} = -.374$ ,  $r_{\text{self}} = -.208$ ). Our results also indicate that a robust strategy of probabilistically predicting mind wandering in cases with poor or missing gaze data led to improved performance on all metrics, as compared to simply discarding

these data. Our findings demonstrate that an automated objective measure might be available for laboratory studies of mind wandering during reading, providing an appealing alternative or complement to self-reports.

**Keywords** Mind wandering · Reading · Eye gaze · Machine learning

It is common for one's attention to shift toward spontaneously generated, task-unrelated thoughts. This phenomenon is called mind wandering. Numerous studies have investigated mind wandering across a range of tasks and have found that it occurs anywhere between 20%–50% of the time (Kane et al., 2007; Killingsworth & Gilbert, 2010; Schooler, Reichle, & Halpern, 2004; Smilek, Carriere, & Cheyne, 2010). Multiple studies (Feng, D'Mello, & Graesser, 2013; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997; Seibert & Ellis, 1991; Smallwood et al., 2004; Smallwood, Fishman, & Schooler, 2007; Smallwood & Schooler, 2006), including a recent meta-analysis of 49 research reports (Randall, Oswald, & Beier, 2014), have indicated that mind wandering during a task is negatively related with task performance. For instance, mind wandering is negatively correlated with text comprehension, partly because textual information is not integrated with the mental model of the text when the reader is mind wandering (Feng et al., 2013; Smallwood, 2011).

An open issue pertains to the measurement of mind wandering. Previous psychology research has primarily relied on self-reports of mind wandering. These are either freely reported by the participants throughout the task (self-caught), in response to thought probes interspersed during the task (probe-caught), or upon completion of the task (retrospective) (Smallwood & Schooler, 2015). Although self-reports provide an undoubtedly useful and valid measure of mind wandering (Smallwood et al.,

✉ Myrthe Faber  
mfaber@nd.edu

<sup>1</sup> Department of Psychology, University of Notre Dame, Notre Dame, Indiana 46556, USA

<sup>2</sup> Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA

2004; Smallwood, McSpadden, & Schooler, 2008; Smallwood & Schooler, 2006), they have several disadvantages. First, self-reports are inherently subjective. It is possible for participants to incorrectly report mind wandering, either inadvertently (e.g., mind wandering could occur outside of awareness; Smallwood & Schooler, 2006) or intentionally (e.g., social desirability biases). Second, reporting mind wandering interrupts the natural flow of the task when measured concurrently. The act of reporting itself could potentially reengage the participant, leading to underestimated rates of mind wandering. Furthermore, if a probe-caught method is used, there is a limit to the number of times that participants can be probed, because probing both can be disruptive and can too frequently lead to lower reported mind-wandering rates (Seli, Carriere, Levene, & Smilek, 2013). On the flip side, infrequent probing could lead to underestimated mind-wandering rates. Retrospective reports circumvent these issues, but are susceptible to limitations associated with memory recall and reconstruction.

Researchers have recently argued for a shift from considering thought probes as the sole identifier of mind wandering to treating them as one of many sources of data that can be leveraged to distinguish inattention from on-task behavior (Hawkins, Mittner, Boekel, Heathcote, & Forstmann, 2015). Previous work has identified some behavioral and physiological measures that are modulated by mind wandering. These include behavioral measures such as response times (McVay & Kane, 2009), physical posture (Seli et al., 2014), prosody (Drummond & Litman, 2010), reading speed (Franklin, Smallwood, & Schooler, 2011; Mills & D’Mello, 2015), and physiological measures such as brain activity (Christoff, Gordon, Smallwood, Smith, & Schooler, 2009; Mittner et al., 2014; O’Connell et al., 2009; Smallwood, Beach, Schooler, & Handy, 2008; Weissman, Roberts, Visscher, & Woldorff, 2006), peripheral physiological responses (Blanchard, Bixler, Joyce, & D’Mello, 2014; Pham & Wang, 2015; Smallwood et al., 2004), eye movements (Foulsham, Farley, & Kingstone, 2013; Frank, Nara, Zavagnin, Touron, & Kane, 2015; Reichle, Reineberg, & Schooler, 2010; Uzzaman & Joordens, 2011), eye blinks (Frank et al., 2015; Grandchamp, Braboszcz, & Delorme, 2014; Smilek et al., 2010; Uzzaman & Joordens, 2011), and pupil diameter (Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013; Smallwood et al., 2011).

Identifying these behavioral and physiological correlates of mind wandering is an important, but only a first step. The next challenge is to leverage them to build models that can detect mind wandering. One approach is to use supervised machine learning techniques (Domingos, 2012) to build a computational model of the relationship between a measure (in this case, eye gaze; see below) and instances of self-reported mind wandering (D’Mello, Duckworth, & Dieterle, 2017). The “learned” model serves as a mind-wandering detector, using a machine-readable data source (e.g., eye gaze, neural activity) to reproduce a human-provided one (e.g., self-reported mind wandering). As such, it is

possible to obtain a continuous classification of mind wandering, which can be aggregated into a proportion for a task or person.

This approach has several advantages. It provides an alternative or complement to subjective mind-wandering measures as it extrapolates the learned associations to unseen data for which no self-reports are necessary. This means that it is possible to measure mind wandering unobtrusively once the associations have been learned. The models are also typically built from a combination of machine-readable signals (henceforth called features, which is the standard terminology in machine learning), which should allow for more accurate models than the use of a single measure (Hawkins et al., 2015). This approach also leverages advances in supervised learning techniques, such as those that support nonlinear decision boundaries, ensemble learning methods, and models that favor generalizability to future data.

We developed and tested an automatic gaze-based mind-wandering detector with the aim of obtaining a valid, robust, and generalizable measure of mind wandering during computerized reading. Our approach applies supervised learning methods to eye-gaze data and self-caught mind-wandering reports. We developed our measure in the context of reading, a common context to study mind wandering, but the general method can be applied to alternate tasks (e.g., Hutt, Mills, White, Donnelly, & D’Mello, *in press*; Mills, Bixler, Wang, & D’Mello, 2016). In what follows, we discuss the key components of our measure.

### Self-caught mind-wandering reports

Supervised learning models are trained on labeled data containing instances (or cases) that are marked as “mind wandering” or “not mind wandering.” Here, we use self-caught reports of mind wandering as the labels. Although this type of reporting has its limitations (as we note below and further address in the discussion), a key advantage is that there is no limit to the number of reports, and reports can occur anywhere (i.e., they not limited by probe placement). An oft-noted disadvantage of this method is that instances of mind wandering can go unnoticed, because self-caught reporting relies on the participant’s meta-awareness. However, our method capitalizes on the associations learned from the reported instances and extrapolates them to new data, suggesting that any potential “missed” mind-wandering instances will be detected.

Another important advantage of using self-caught reports is that they (more so than probe-caught reports) maintain the temporal relationship between a stream of behavioral or physiological data and the report. That is, whereas probe-caught reports can either signal the onset of mind wandering, its continuation, or end point depending on their placement, self-caught reports tend to be associated with the point at which the participant becomes aware of the fact that they were mind

wandering, often signaling the end of an episode. Thus, across instances, windows of time before the report are therefore likely to reflect a similar process, namely mind wandering before the participant became aware that he or she was doing so.

Although probe- and self-caught reporting both disrupt the natural flow of a task, the latter occurs while a person is off-task (i.e., they realize that they are mind wandering), thereby not causing additional on-task disruptions. Further, meta-awareness of on-task behavior (i.e., being aware of what you just read) is a critical component of reading comprehension (McNamara & Magliano, 2009), so recognizing attentional lapses is instrumental to the main task. For these reasons, we consider self-caught mind wandering to be more congruous with the course of naturalistic reading than probing.

### Eye gaze correlates of mind wandering

The idea that eye gaze can be used to measure mind wandering is supported by decades of research suggesting that eye movements are modulated by ongoing cognitive processes, especially attention (Just & Carpenter, 1980; Rayner, 1998; Reichle, Pollatsek, Fisher, & Rayner, 1998). This so called eye-mind link (Just & Carpenter, 1976) breaks down when attentional focus shifts from the external environment (e.g., reading a text) to internal thoughts (e.g., what to have for dinner tonight) (Smallwood et al., 2011). Thus, mind wandering should be reflected by a decoupling between eye gaze and the reading task.

In reading, fixations (points where gaze is maintained at the same location) normally follow a regular pattern, which is modulated by lexical features such as the length and frequency of words (Rayner, 1998). These patterns tend to be more erratic during mind wandering. For instance, short fixations on low frequency words and long fixations on high frequency words are predictive of mind wandering, as they signal a decoupling between eye gaze and the text (Schad, Nuthmann, & Engbert, 2012). Although such content-dependent patterns can be useful for identifying fine-grained attentional processes, gaze data need to be highly precise to track fixations on individual words, which limits the broader applicability.

Fortunately, content-independent features of eye gaze have also been linked to mind wandering. For instance, participants tend to have fewer and longer fixations, and fixate more on off-text locations during mind wandering (Reichle et al., 2010). Similarly, saccades (rapid eye movements between fixations), within-word regressions (sum of durations of all fixations on a word), and runs (two consecutive fixations within an area of interest) are less frequent and/or slower during mind wandering (Uzzaman & Joordens, 2011). These findings demonstrate that the regular gaze pattern breaks down during mind wandering.

In addition, blink rates increase in the intervals preceding a mind-wandering report (Smilek et al., 2010). This has been

linked to the idea that the visual interruption afforded by an increased blink rate facilitates internal thought generation, which is in line with the observed decrease in fixations as noted above. Recent accounts have argued that the locus coeruleus norepinephrine (LC-NE) system controls the trade-off between on- and off-task behaviors (Mittner, Hawkins, Boekel, & Forstmann, 2016). Fluctuations in this system are measured using pupillometry (Aston-Jones & Cohen, 2005) and studies have shown that pupil diameter is significantly larger during periods of mind wandering in a word-by-word text reading paradigm (Franklin et al., 2013). Furthermore, pupil diameter and its standard error are larger when participants incorrectly respond during working memory tasks, suggesting that an increase in pupil diameter reflects a lapse in attention devoted to the task at hand (Smallwood et al., 2011). However, pupil diameter and its response to stimulation have also been found to be smaller during mind wandering (Mittner et al., 2014), so further research is necessary to shed light on these contradictory findings.

Together, these studies indicate that measures of eye gaze are related to mind wandering, suggesting that it might be possible to differentiate mind wandering from normal reading on the basis of eye gaze. We leveraged these insights by computationally modeling the relationship between eye gaze features and instances of mind wandering.

### Requirements of automated mind-wandering detection for psychological research and limitations of existing gaze-based measures

Psychological research has so far primarily employed self-reported measures of mind wandering as few alternatives have been available (but see Mittner et al., 2014, for a brain-based measure for a sustained attention task). Automatic gaze-based detection of mind wandering could provide an alternative or complementary measure, but only if it satisfies several criteria. In particular, it needs to provide a valid and reliable estimate of the occurrence of mind wandering for each participant regardless of quality of gaze data. It also needs to generalize to “new” participants whose data it has not seen before.

To date, only a few studies have attempted automatic mind-wandering detection based on eye gaze during reading (Bixler & D’Mello, 2014, 2015, 2016; D’Mello, Cobian, & Hunter, 2013; Loboda, 2014). Each study serves as a proof of concept of a gaze-based mind-wandering detector, but each has key limitations with respect to the aforementioned criteria of validity, robustness, and generalizability as discussed below.

**Reliability, convergent, and predictive validity** We first assessed the internal consistency of the gaze-based mind-wandering detector by computing odd–even reliability (a form of split-half reliability). To establish convergent validity, we

correlated the proportion of cases that the gaze-based detector denoted as mind wandering with self-reported mind-wandering proportions. Predictive validity was obtained by correlating gaze-based mind wandering with text comprehension scores, which has been shown to be negatively related with self-reported wandering (Bixler & D’Mello, 2016; Faber, Mills, Kopp, & D’Mello, 2016; Feng et al., 2013; Mills, D’Mello, & Kopp, 2015; Randall et al., 2014; Unsworth & McMillan, 2013).

We note that the present study focused on participant-level mind-wandering proportions, as state-of-the-art predictive models cannot (yet) classify individual instances of mind wandering with sufficient accuracy for psychological research (Bixler & D’Mello, 2015; Pham & Wang, 2015). These models are typically developed for engineering applications, particularly in human–computer interaction, in which the goal is for intelligent interfaces to respond to individual episodes of detected mind wandering (D’Mello, 2016; D’Mello, Kopp, Bixler, & Bosch, 2016). In those contexts, imprecise detection is permissible because the end goal is not to measure mind wandering in and of itself, but rather to influence some outcome variable of interest.

In contrast, in psychological research, the goal is usually to measure mind wandering for use as a variable for analysis. Accuracy is clearly important here, but our emphasis on overall mind-wandering proportions should not pose a limitation as most psychological studies take an aggregate of the self-reports per participant as the mind-wandering measure. This is done by either counting the number of self-caught mind-wandering instances or computing the proportion of probes for which the participants reported mind wandering. Similarly, our aim is to automatically estimate a mind-wandering proportion for each participant based on eye gaze information and show that this estimate is valid by correlating it with the number of self-reports and scores on comprehension assessments.

It is important to note that we did not expect a perfect correlation between self-caught and gaze-based mind-wandering proportions. Although both tap into the same construct (i.e., mind wandering during computerized reading), the measurements are based on different sources. Reports of self-caught mind wandering critically rely on participants’ metacognitive awareness, so lapses in attention that occur outside of this awareness are not reported. Thus, the reports only reflect *conscious* mind wandering as reported by the participant. The automated detector, on the other hand, has access to a different source, namely eye gaze data. What these eye gaze features reflect exactly (i.e., whether they reflect underlying constructs in addition to mind wandering) is unknown, and more generally, an open question in gaze-based mind-wandering research. Furthermore, personal biases (e.g., deciding not to report mind wandering out of embarrassment) affect self-reported mind wandering, whereas these mind-wandering episodes are likely to be reflected in eye gaze. As we noted

above, our measure might pick up on these unreported instances, as our model capitalizes upon learned associations between eye gaze and reported mind wandering.

We therefore expect moderate but not perfect overlap between mind-wandering proportions obtained from both sources. In general, weak to moderate correlations between physiological/behavioral measures and self-reports are quite common in psychological research, for example in the affective sciences (Barrett, 2006) and personality research (Duckworth & Kern, 2011).

**Robustness to missing, poor, and invalid gaze data** Eye gaze analyses need to be robust and automatic for mind-wandering detection. This means that eye gaze data cannot be subject to manual corrections or exclusions that rely on visual inspection of the data. Indeed, this is an important limitation in previous studies. Since the quality of eye gaze data can be poor (e.g., due to loss of signal), some studies have only used the very best data, resulting in exclusion of data points and entire participants’ data (Bixler & D’Mello, 2015; Loboda, 2014). This is obviously problematic for an automatic mind-wandering detector as it would yield selective estimates based on when gaze can be tracked for some participants and no estimates for others.

Moreover, a robust detector should be able to model poor (e.g., data from only one eye) or missing eye gaze data. Several studies have suggested that missing data might be related to the occurrence of mind wandering, as off-screen fixations are more likely when participants are not attending to the stimulus (Loboda, 2014; Reichle et al., 2010). Ignoring these instances could yield imprecise mind-wandering estimates. Hence, in contrast with previous studies (Bixler & D’Mello, 2014, 2015), we considered all the gaze data and studied the validity of mind-wandering estimates with and without inclusion of poor or missing data.

Because the quality of gaze data can vary between participants and eyetrackers, the detector needs to be robust to gaze-tracking inaccuracies. For instance, head movements and small errors in calibration can have downstream consequences for features that rely on positional information (local features; e.g., fixations on specific words), whereas other features (global features; e.g., number of fixations, mean fixation duration) are not affected as much. Previous studies have found that local features contributed little to classification accuracy of self-reported mind wandering over global features (Bixler & D’Mello, 2015, 2016). Furthermore, global features are computed independent of specific words on the screen, which aids generalizability to different texts. For these reasons, we used global features in our mind-wandering detector.

Another limitation of previous studies is that model accuracy was established using test samples with artificial base rates of mind wandering (D’Mello et al., 2013). For example, in D’Mello et al. (2013), both the training and test sets were

downsampled to contain 50% of the mind-wandering instances. It is unclear how these models would perform when the testing set reflects the original, skewed class distributions (roughly 30% mind wandering), as we did here.

**Generalizability to new participants** An automatic gaze-based mind-wandering detector needs to estimate mind wandering for “new” participants whose data it has not encountered before. The supervised learning methods adopted in this study automatically learned (from training data) relationships between eye gaze and mind wandering. They then used these relations to estimate mind-wandering proportions for new or unseen data. If the learned relations were too specific to the participants in the training data (i.e., overfitting), the detector’s performance would be very high for the training participants, but low for new participants. This is likely to occur when data from the test participants are included in the training data (e.g., Drummond & Litman, 2010). To address this, we used a leave-one-participant-out cross-validation procedure, in which the model was trained on data from all but one “held-out” participant. The model learned from the other participants’ data (training set) was applied in order to estimate the mind-wandering proportion for the held-out participant’s data (testing set). The process was repeated until all participants had been in the testing set once.

Furthermore, the data used to train the model were collected from two universities with very different student characteristics and with two eyetrackers, thereby introducing additional sources of variability that can improve model generalizability.

### Collecting data to train the model

We leveraged the data from an existing study that collected self-caught mind-wandering reports, eye gaze data, and comprehension assessments during a computerized reading task. Below, we focus on the aspects of the study that were germane to the present goal; readers are referred to Kopp, D’Mello, and Mills (2015) for full details.

### Participants

Eye gaze data was recorded for 132 of the 140 college students included in the previous analysis of this data set (Kopp et al., 2015). Ninety participants were from a highly selective private Midwestern U.S. university and 42 were from a public university in the Southern United States (gaze data for the remaining eight participants was not collected). Participants were on average 20.3 years old, 62% were females, 61.8% were Caucasian/White, 19.8% African-American/Black, 6.1% Hispanic, Latino or of Mexican origin, 8.4% Asian, and 3.8% reported “other”.

### Materials

**Text** Participants read an excerpt from a book entitled *Soap-Bubbles and the Forces Which Mould Them* (Boys, 1890). This book was chosen because it discusses a science concept that would be relatively unfamiliar to a majority of readers. The text contained around 5,700 words from the first 35 pages of the first chapter of the book. In all, 57 pages (screens of text) with an average of 100 words each were displayed on a computer screen in 36-pt Courier New typeface. The only modification to the text was the removal of images and references to them after verifying that these were not needed for comprehension.

**Eyetracking devices** Two different eyetrackers were used, one at each university. At the private Midwestern university, a Tobii TX300 set was used with a sampling frequency of 120 Hz, whereas at the public Southern university a Tobii T60 was used with a sampling frequency of 60 Hz. Tobii manufactures remote eyetrackers, so participants could read without any restrictions on head position or movement. Both eyetrackers were set to record in binocular mode.

**Trait-based mind-wandering questionnaire** After the main task, participants completed a five-item (Cronbach’s  $\alpha = .813$ ) trait-based mind-wandering questionnaire (obtained from Mrazek, Phillips, Franklin, Broadway, & Schooler, 2013). The questionnaire comprised the following items: Q1. “I have difficulty maintaining focus on simple or repetitive work”; Q2. “I do things without paying full attention”; Q3. “While reading, I find I have not been thinking about the text and must therefore read it again”; Q4. “I find myself listening with one ear, thinking about something else at the same time”; and Q5. “I mind wander during lectures or presentations”. Response options included: *almost never*, *very infrequently*, *somewhat infrequently*, *somewhat frequently*, *very frequently*, and *almost always*.

### Retrospective engagement/attention questionnaire

Participants completed a researcher-created questionnaire about their subjective experience after reading. Two of the questions pertained to engagement and attentional focus, which are related to mind wandering. Question 1, “How engaged were you while you were reading about soap bubbles?” was answered on a six-point scale of “very bored” to “very engaged.” Question 2, “While you were reading, was your attention focused on the text?” was answered on a four-point scale of “I focused completely on task unrelated thoughts” to “I stayed completely on task”.

**Comprehension assessment** A posttest consisting of 12 multiple-choice questions (four answer options) was used to assess text comprehension. The questions tapped surface-level

content covered directly in the text and did not require inference. For example, the question, “The suggestion that there is an Etruscan vase in the Louvre that depicts children blowing bubbles from a pipe was put forth by: (a) Lord Rayleigh; (b) Van der Mensbrugge; (c) Millais; (d) Plateau” had option (d) as the correct response.

### Mind-wandering reports

Mind wandering was measured using the self-caught method. Participants received the following instructions (based on Schooler et al., 2004):

Your primary task is to read the text in order to take a short test after reading. At some points during reading, you may realize that you have no idea what you just read. Not only were you not thinking about what you are actually reading, you were thinking about something else altogether. This is called “zoning out”. If you catch yourself zoning out at any time during reading, please indicate what you are thinking about at that moment during reading.

When zoning out:

If you are thinking about the task itself (e.g., how many pages are there left to read, this text is very interesting) or how the task is making you feel (e.g., curious, annoyed) but not the actual content of the text, please press the key that is labeled “task”.

OR

If you are thinking about anything else besides the task (e.g., what you ate for dinner last night, what you will be doing this weekend) please press the key that is labeled “other”.

Please familiarize yourself with where these two keys on the keyboard now so that you will know their location when you begin reading.

Please be as honest as possible about reporting zoning out. It is perfectly natural to zone out while reading. Responding that you were zoning out will in no way affect your scores on the test or your progress in this study, so please be completely honest with your reports. If you have any questions about what you are supposed to do, please ask the experimenter now.

These instructions encouraged participants to monitor their ongoing comprehension of the text rather than their thoughts. Following previous work that has shown that task-relatedness of spontaneous thoughts can modulate task performance (Stawarczyk, Majerus, Maj, Van der Linden, & D’Argembeau, 2011), we distinguished between task-related interferences (TRIs) and task-unrelated thoughts (TUTs). Note, however, that both types of reports only occurred when participants found themselves immersed in thoughts unrelated to the *content* of what they were reading and had no idea what they just read.

This contrasts with other approaches that probe participants to report the content of their thoughts, regardless of whether they were phenomenologically zoning out (e.g., Stawarczyk et al., 2011). Thus, TRIs and TUTs were conceptually similar in our study in that they both reflect subjective instances of zoning out. In line with Christoff, Irving, Fox, Spreng, and Andrews-Hanna (2016), our operationalization was intended to capture what is “arguably the key feature of mind wandering, reflected in the term itself: to wander means to ‘move hither and thither without fixed course or certain aim’” (Christoff et al., 2016, p. 719). Because both TRIs and TUTs refer to thoughts unrelated to the content of the text, were positively correlated with one another (Spearman’s  $r = .505$ ,  $p < .001$ ), and were similarly negatively correlated with comprehension scores (Spearman’s  $r = -.175$  and  $-.189$  for TRIs and TUTs, respectively,  $p < .05$ ), we combined them into a single mind-wandering category.

Participants could report mind wandering any number of times on a page; however, only data prior to the first report on a page were considered, because the act of reporting likely interfered with eye gaze. For the same reason, the eye gaze 3 s prior to each mind-wandering report was discarded because participants likely gazed at the keyboard prior to each report, thereby confounding the gaze data.

### Procedure

All instructions and experimental materials were administered via computer, and all procedures were approved by the ethics boards of both universities. Participants were first informed that their primary task was to read a text in order to take a short test after reading. They were then provided the instructions on reporting task-related interferences and task-unrelated thoughts as described above. Because these data were collected as part of a larger research project, participants were assigned to one of two list-making conditions (listing their current concerns or listing features of an automobile; for details, see Kopp et al., 2015). The present study does not differentiate between these conditions. Participants completed the Positive and Negative Affect Scale (PANAS) measure (Watson, Clark, & Tellegen, 1988), which measured the extents to which they experienced 20 emotions. The PANAS was also part of the larger research study and is not analyzed further here. After this, participants went through the calibration procedure for the eyetracker and were reminded of the main task instructions. They then began the computerized reading task. They proceeded through the text by pressing the “right arrow” key (only forward navigation was possible) and self-reported their instances of mind wandering while reading. A tone sounded in response to the keypress, to inform them that their response had been recorded. Upon completion of the reading task, they completed the PANAS once more, followed by the retrospective and trait-based mind-wandering proneness questionnaires. Finally, they were given the reading comprehension assessment and were fully debriefed on completion.

## Machine learning to build the model

An overview of the machine-learning approach is given in Fig. 1.

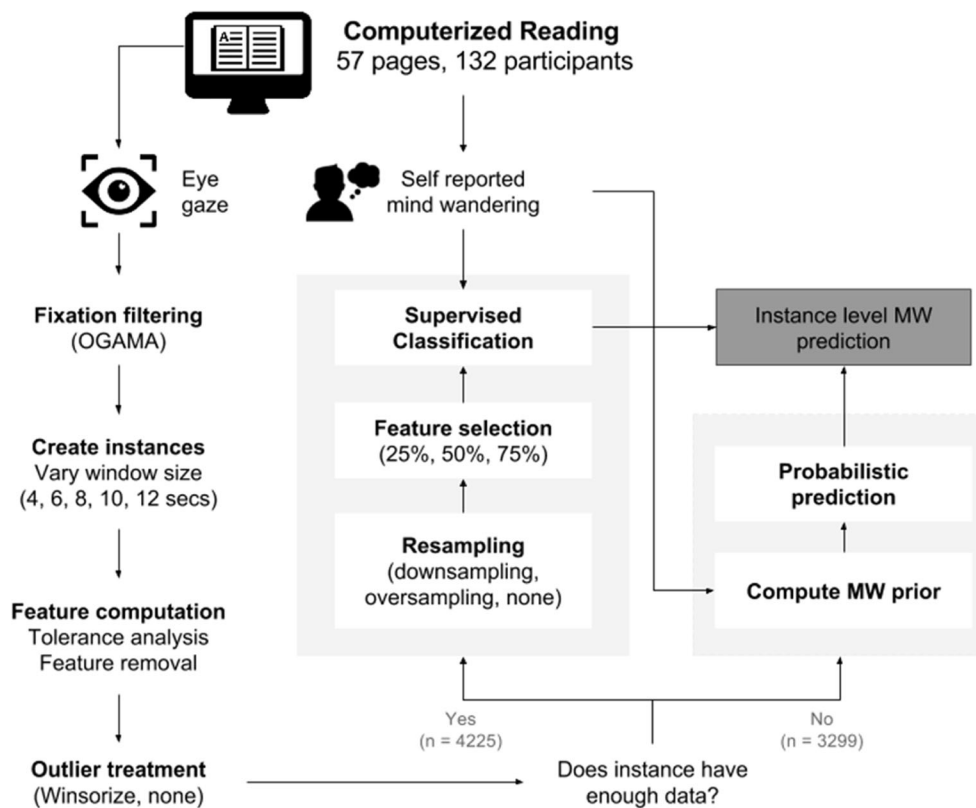
### Eye movement detection, instance creation, and feature engineering

The raw gaze data from both eyes were averaged and converted into eye movements using a dispersion-based filter with an open-source gaze analyzer software tool (OGAMA; Voßkühler, Nordmeier, Kuchinke, & Jacobs, 2008). Fixations were defined as consecutive gaze points within a range of 57 pixels (approximately 1 deg of visual angle) for longer than 100 ms, which is the shortest duration for naturalistic eye movements during reading (Holmqvist et al., 2011; Rayner, 1998). Saccades were computed from the fixations. Blinks were detected as periods during which the eyetracker lost track of both eyes for a minimum duration of 83 ms and a maximum duration of 400 ms, based on the range of blink durations during reading (Holmqvist et al., 2011).

Features were computed using the data from a specific period of time (*window*) on each computer screen of text (called a *page*). Each window ended 3 s prior to the first mind-wandering report on the page. This 3-s offset was used to avoid confounds pertaining to the keypress to submit the mind-wandering report. Data between the first mind-wandering report and the end of the page were ignored.

Training a discriminative classification model requires both instances in which participants were mind wandering and instances in which they were not. With self-caught reports, negative instances are not readily available and need to be created from the pages on which a participant did not report mind wandering. For these pages, we selected a time point corresponding to the average time at which a report occurred on the self-caught pages (16.7 s into the page for the present data set). Previous work has shown that this method is superior to other methods of selecting a window (e.g., at the end of a page, or at the same time as a randomly selected self-caught report; Bixler & D’Mello, 2015).

Next, we computed four sets of global features for each window: eye movement descriptive features, pupil diameter descriptive features, blink features, and miscellaneous gaze



**Fig. 1** Visualization of the machine-learning approach outlined in the Machine Learning section. Gaze data were processed as outlined in the first subsection. Instances with fewer than five fixations, less than 4 s of available data, or a feature that could not be computed with the available data were deemed as having insufficient data. Instances with sufficient data were used for supervised classification using the self-reports as labels, as we outline in the following subsections. For instances with

insufficient data, a probabilistic prediction was obtained using the prior probability of mind wandering (MW), based on the reasons for missing data outlined in the final subsection. Together, these steps resulted in a mind-wandering likelihood for each instance. Note that the steps in light gray were repeated for each held-out participant because we used leave-one-participant-out cross-validation

properties. The eye movement descriptive features were statistical functionals for *fixation duration*, *saccade duration*, *saccade amplitude*, *saccade velocity*, and *relative and absolute saccade angle* distributions. The *fixation duration* was the duration of each fixation in milliseconds. The *saccade duration* was measured as the number of seconds between two subsequent fixations, whereas the *saccade amplitude* was the number of pixels between two subsequent fixations. The *saccade velocity* was measured as the saccade amplitude divided by the saccade duration. The *absolute saccade angle* was the angle between the line segment between two subsequent fixations and the *x*-axis. The *relative saccade angle* was computed as the angle between two subsequent saccades. For each of these eye movement measurements, we computed the minimum, maximum, mean, median, standard deviation, skew, kurtosis, and range, thereby yielding 48 features.

For the *pupil diameter descriptive features*, the eyetracker's estimate of pupil diameter was first standardized by computing the participant-level *z* score, and then the same eight statistical functionals were computed.

The *blink features* consisted of the *number of blinks* and the *mean blink duration*.

The *miscellaneous gaze properties* consisted of the number of saccades, horizontal saccade proportion, fixation dispersion, and the fixation duration/saccade duration ratio. The *horizontal saccade proportion* was the proportion of saccades with an angle no more than 30 deg above or below the *x*-axis. *Fixation dispersion* was computed as the root mean square of the distance of each fixation to the average fixation in the window. The *fixation duration/saccade duration ratio* was the ratio of the sum of all the fixation durations to the sum of all the saccade durations in the window.

All together, 62 global gaze features were computed. We removed the mean blink duration because it had missing values for more than 10% of the instances. As expected, some features were found to be highly correlated. For instance, the numbers of fixations and saccades were strongly related (since saccades are the rapid eye movements between fixations), and different measures of centrality or dispersion tend to be correlated. To reduce multicollinearity and avoid the curse of dimensionality (Domingos, 2012), we removed the 29 features with a variance inflation factor greater than 5 (i.e.,  $R_i^2 > .80$ ; Craney & Surlles, 2002), resulting in 32 features for the model-building process.

## Supervised learning

We considered a wide array of classifiers, since there was no a priori knowledge about which classifier would be most suitable for this task. The following Waikato Environment for Knowledge Environment (WEKA; Hall et al., 2009) implementations (with default hyper parameters) were used: bagging, with REPTree as a base learner; Bayes net; naïve

Bayes; logistic regression; support vector machine; *k*-nearest neighbors; decision table; C4.5 decision tree; random forest; REPTree; and random tree.

We also varied the following four parameters known to affect classification accuracy. First, we experimented with *window sizes* of 4, 6, 8, 10, and 12 s.

Next, *outliers*, defined as values greater than 3 standard deviations from the mean, were either replaced with the corresponding value at 3 standard deviations above or below the mean (Winsorization), or were left untouched (no outlier treatment).

To address class imbalance, which was particularly problematic because mind wandering was the minority class (discussed below), the class distribution of the training set (only) was made equal through either downsampling or oversampling across five iterations. Downsampling consisted of randomly removing instances of the majority class. For oversampling, we used the Synthetic Minority Oversampling Technique (SMOTE; Chawla, Bowyer, Hall, & Kegelmeyer, 2002) to create synthetic instances of the minority class. We also considered a model based on the original class distributions. It should be noted that the class distributions in the testing set were left untouched.

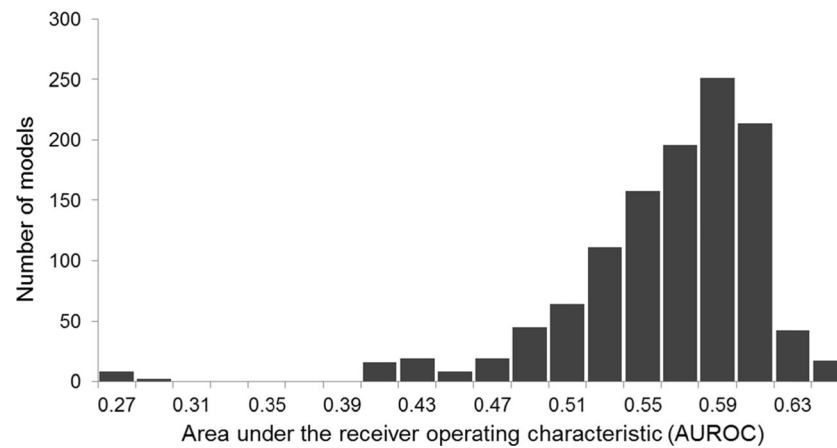
Finally, *feature selection* was applied (to the training set only) in order to select the most diagnostic features. Features were ranked higher if they were weakly correlated with other features but strongly correlated with mind-wandering reports using a correlation-based feature selection algorithm from WEKA (CFS; Hall, 1999). To avoid overfitting, feature selection was performed on a random 66% of the participants in the training set. The process was repeated five times to ameliorate variance caused by random selection of participants. The feature rankings were then averaged over these five iterations and 25%, 50%, or 75% of the top-ranked features were retained.

## Model selection

The classification models were evaluated using a leave-one-participant-out validation method to ensure that data from each participant were exclusive to either the training or test set. Using this method, the data from one participant were held aside for the test set, while the data from the remaining participants were used to train the model. The process was repeated until all the participants had been in the test set once.

Figure 2 shows a histogram of the AUROCs (areas under the receiver operating characteristic [ROC] curve) for each of the 1,170 candidate models. A total of 85.6% of the models had ROCs greater than chance (AUROC > .50). Notably, the 17 best models (each with an AUROC above .63) were all logistic regression models with a window size of 12 s. The overall best model achieved an AUROC of .64, which reflects a 28% improvement over a chance model (AUROC = .50). This model used a total of 24 features after tolerance analysis





**Fig. 2** Histograms of area under the receiver operating characteristic curves

and feature selection. The outliers in this model were Winsorized, and the training set was downsampled.

### Feature analysis

We explored how eye movements differed between mind wandering and normal reading by computing the effect size (Cohen's  $d$ ) for each feature in the final model. Table 1 lists these features in descending order of effect size. Taking into account the top 50% of these features, the results aligned with previous studies suggesting that fewer (represented by number of saccades here) and longer fixations are the key gaze signatures of mind wandering (Bixler & D'Mello, 2015; Foulsham et al., 2013; Reichle et al., 2010; Smilek et al., 2010; Uzzaman & Joordens, 2011). Furthermore, our data indicated that patterns of saccades were predictive of mind wandering: Saccade angles were smaller and encompassed a narrower range during mind wandering. Importantly, the proportion of horizontal saccades was lower during mind wandering, suggesting that the regular left-to-right reading behavior associated with normal reading breaks down during mind wandering.

We also observed that the standardized pupil diameter was smaller during mind wandering, with larger and more right-skewed diameters for normal reading. This finding is surprising, since off-task behavior is usually associated with larger pupil diameters. However, a recent study showed a similar pattern (Mittner et al., 2014), which suggests the need for further research targeting the relationship between pupil diameter and mind wandering. Similarly, in contrast with previous studies, we did not observe a difference in blink rates, which might also warrant further investigation.

### Handling unclassified instances

Instances with fewer than five fixations and/or less than 4 s of available data were excluded from the classification process, because these windows did not contain sufficient data to

compute the statistical functionals that comprise the gaze features. In all, a total of 4,225 instances from all 132 participants were used, out of a possible 7,524 instances (132 participants  $\times$  57 pages). Thus, in 44% of the cases (3,299 of the 7,524 instances), the data were insufficient for classification. Rather than simply discarding these data, we explored whether including them in the estimation process improved the validity of the measure.

We proceeded by first classifying the reason for insufficient data as (1) an insufficient amount of gaze data, (2) insufficient reading time, (3) a combination of both factors, and (4) missing data for an individual feature (e.g., no blinks in the window). Next, we computed the probability of self-reported mind wandering for each category (shown in Table 2). We leveraged the considerable variability in the likelihood of mind wandering across categories in order to generate predictions for each unclassified instance. Specifically, the mind-wandering proportions shown in Table 2 were regenerated for each held-out participant. These were used to obtain a probabilistic prediction (across 100 samples) for each page based on the reason for that page being unclassified. For example, on the basis of Table 2, there would be a 56.3% likelihood that a given page would be classified as mind wandering if the reason for it being unclassified was that the participant did not spend sufficient time ( $<4$  s) on that page. These probabilistic predictions for the unclassified instances were combined with the model-based estimates for the classified instances, thereby yielding a mind-wandering likelihood for all 7,524 instances.

### Validating the model

The model-based mind-wandering estimates were validated using three criteria: (1) a comparison between distributions of model-based and self-reported mind-wandering proportions, (2) convergent validity (correlation between model-

**Table 1** Means (with standard deviations in parentheses) and effect sizes for gaze features corresponding to instances of mind wandering versus normal reading

Feature	Mind Wandering	Normal Reading	<i>d</i>
Horizontal saccade proportion	.939 (.057)	.955 (.042)	-.312
Fixation duration median	223 (35.2)	216 (27.5)	.223
Pupil diameter skew	.096 (.322)	.158 (.219)	-.222
Number of saccades	28.3 (8.91)	30.2 (7.99)	-.218
Relative saccade angle range	356 (5.89)	357 (2.70)	-.204
Pupil diameter median	-.163 (.453)	-.092 (.263)	-.192
Absolute saccade angle mean	356 (7.32)	357 (4.37)	-.186
Saccade amplitude kurtosis	2.76 (1.86)	2.46 (1.42)	.179
Fixation duration range	522 (174)	494 (141)	.173
Relative saccade angle kurtosis	-1.93 (.662)	-2.02 (.319)	.172
Relative saccade angle median	159 (72.7)	168 (55.7)	-.149
Relative saccade angle max	358 (4.22)	359 (1.63)	-.133
Absolute saccade angle SD	145 (7.88)	146 (7.66)	-.128
Absolute saccade angle max	358 (3.86)	358 (3.51)	-.127
Saccade amplitude SD	233 (36.7)	237 (25.9)	-.109
Relative saccade angle skew	.068 (.306)	.040 (.178)	.109
Saccade duration max	1,036 (696)	964 (608)	.109
Absolute saccade angle mean	171 (23.1)	169 (19.1)	.103
Saccade duration kurtosis	7.37 (4.39)	7.65 (2.92)	-.072
Saccade amplitude median	166 (33.1)	167 (27.6)	-.058
Absolute saccade angle median	163 (43.2)	161 (37.2)	.045
Fixation dispersion mean	.456 (.034)	.457 (.020)	-.033
Pupil diameter SD	.584 (.149)	.588 (.114)	-.030
Fixation duration kurtosis	2.60 (1.97)	2.55 (1.43)	.030
Saccade velocity SD	5.01 (1.48)	4.97 (1.44)	.026
Fixation/saccade ratio	3.83 (2.46)	3.78 (1.99)	.025
Pupil diameter kurtosis	-.167 (.834)	-.184 (.520)	.024
Absolute saccade angle kurtosis	-1.41 (.369)	-1.42 (.341)	.022
Saccade velocity skew	.221 (.339)	.215 (.340)	.016
Saccade velocity kurtosis	-.572 (.576)	-.573 (.455)	.002
Saccade duration median	47.5 (47.6)	47.5 (44.3)	.001
Number of blinks	1.74 (1.62)	1.74 (1.45)	.000

Pupil diameters were first standardized at the participant level. Duration in milliseconds. Angle in degrees (maximum = 360°). Saccade amplitude in pixels. Velocity in pixels per second. *SD* = standard deviation. Max = maximum

**Table 2** Numbers of instances and mean mind-wandering (MW) proportions for classified and unclassified instances

	Classified Instances	Reason for Being Unclassified			
		Insufficient Gaze Data	Insufficient Time on Page	Insufficient Time & Gaze	Missing Feature
No. instances	4,225	1,248	1,030	894	127
Proportion of total	.561	.166	.137	.119	.017
MW Proportion	.203	.236	.563	.705	.299

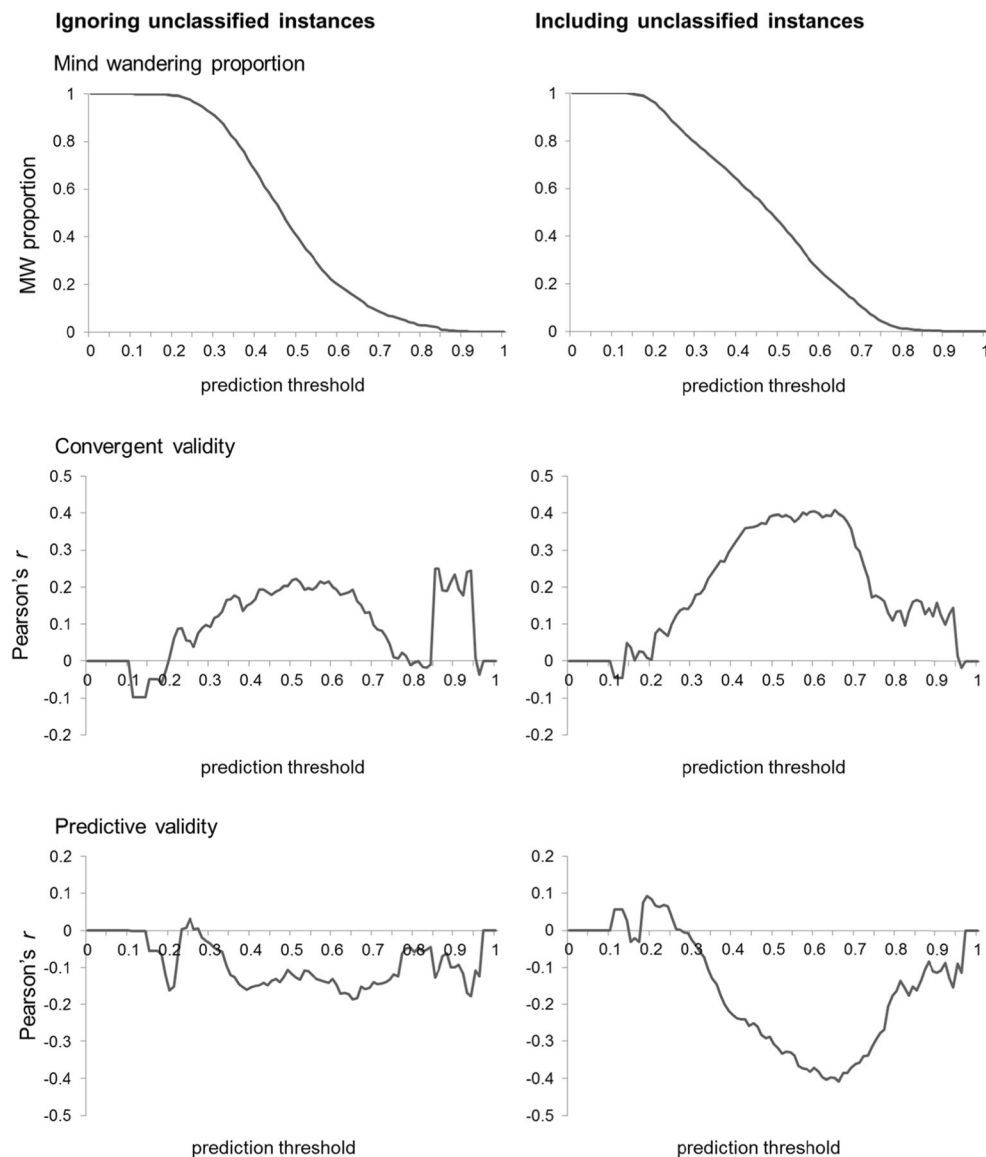
based and self-reported mind wandering), and (3) predictive validity (correlation between model-based mind-wandering proportions and performance on comprehension assessment).

The first step toward validation was to compute the estimated and self-reported mind-wandering proportions for each participant. The logistic regression model we used provides an instance-level likelihood (between 0 and 1) of mind wandering, which needed to be converted into a binary *mind-wandering* or *normal-reading* classification. This required the selection of a prediction threshold; instances with likelihoods above that threshold would be classified as mind wandering, and all other instances as normal reading. There are different ways of deciding upon this threshold. The default threshold of 0.5 can be used, or the threshold can be based on the point(s) on the ROC curve that optimally balance specificity and sensitivity, or that favor(s) one or the other on the basis of the desired application. The threshold can also be based on previous findings (e.g., using an established proportion of mind wandering during reading) or it can be chosen to optimize the relationship between the mind-wandering proportion and other measures (e.g., comprehension scores).

To illustrate, Fig. 3 shows the model-based mind-wandering proportions, convergent validity, and predictive validity at different prediction thresholds. We note that the optimal prediction threshold depends on whether we ignore or include unclassified instances and the validity metric of interest, be it the mind-wandering proportion, convergent validity, or predictive validity). Picking one threshold over another therefore entails a trade-off between one criterion versus another. Here we selected a threshold (.57) that minimized the numerical difference between the mean self-reported and model-based mind-wandering proportions at the group level ( $M_{\text{self}} = .319$ ,  $M_{\text{model}} = .310$ ). This decision came at the expense of the other validity criteria, since alternate thresholds would lead to better convergent and predictive validities (see Fig. 3).

### Distributions of mind-wandering proportions

Table 3 presents group-level mind-wandering proportions for both methods of handling unclassified instances based on the .57 prediction threshold. We note the much lower self-reported as well as model-based mind-wandering proportions



**Fig. 3** Mind-wandering (MW) proportions, convergent validity, and predictive validity for each prediction threshold in the 0-to-1 range, with increments of .01

when unclassified instances were ignored, suggesting the importance of including these cases in the analysis. The group-level model-based and self-reported mind-wandering proportions were highly similar, but this was due to our decision to select a threshold that minimized the difference between the two. More importantly, the distributions of participant-level self-reported and model-based mind-wandering proportions were also highly similar, as is shown in Figs. 4 and 5.

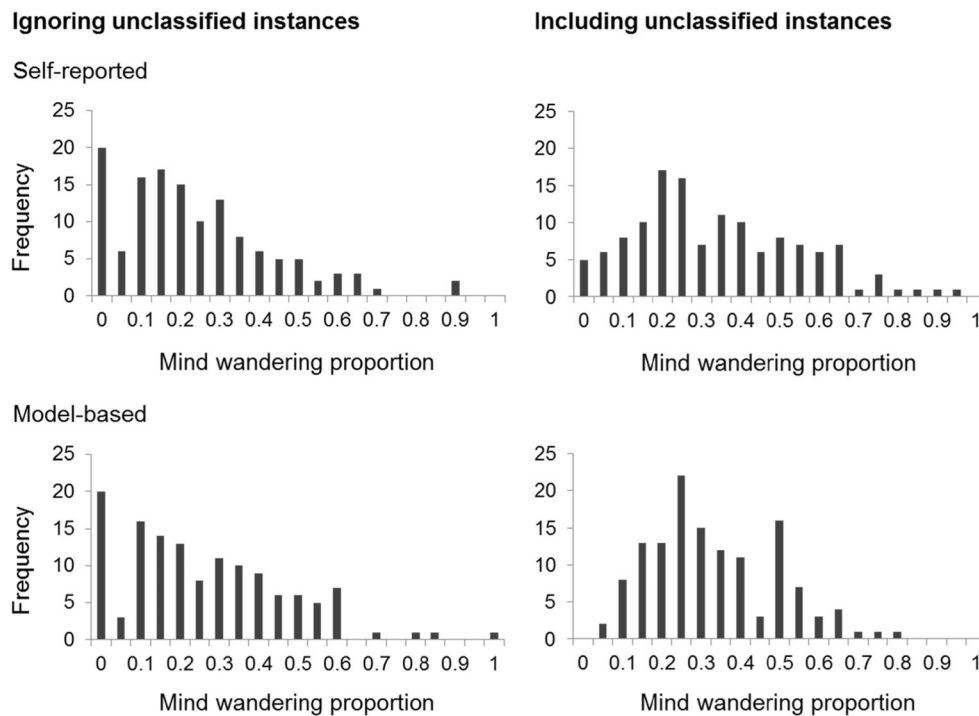
### Internal consistency reliability

To assess the internal consistency of our measure, we computed odd–even reliability by correlating each participant's model-based mind-wandering estimates for odd and even pages. Table 4 presents these correlations for both methods

of handling unclassified instances at the .57 prediction threshold. We found that reliability was higher when unclassified instances were included for both model-based and self-reported mind-wandering proportions. The fact that we observed good (cf. Cicchetti, 1994) internal consistency for the model-based proportions that included unclassified instances

**Table 3** Means and standard deviations (in parentheses) of the participant-level mind-wandering proportions at the group level

Unclassified Instances	Number of Instances	Self-Reported	Model-Based
Ignored	4,225	.217 (.190)	.244 (.202)
Included	7,524	.319 (.211)	.310 (.162)



**Fig. 4** Distributions of participant-level mind-wandering proportions for the self-reports and model-based estimates, after ignoring or including unclassified instances

suggests that this measure provides a reliable estimate of mind wandering.

### Convergent validity

We expected the model-based and self-reported mind-wandering proportions to be positively correlated, which is what we found (Table 5). The correlation was about twice as large when unclassified instances were included than when they were ignored. As expected, this correlation was moderate, because self-reports and behavioral measures seldom overlap strongly, as we discussed in the introduction.

Correlations between the model-based mind-wandering proportions and participants' retrospective engagement/attention ratings provided additional evidence for convergent validity. Again, the correlation was larger when unclassified instances were included (Table 5).

The trait-based measure of mind-wandering proneness did not correlate significantly with either the self-reported or model-based mind-wandering proportions (Table 5). Whether this incongruence reflects inaccurate self-appraisal of mind-wandering proneness or a by-product of using a self-caught measure warrants further investigation. However, previous studies have suggested that individual differences in trait-based mind wandering are related to different fluctuations in neural activity in the default mode network than are episodes of self-reported mind wandering (Kucyi & Davis, 2014), suggesting that some incongruence is to be expected.

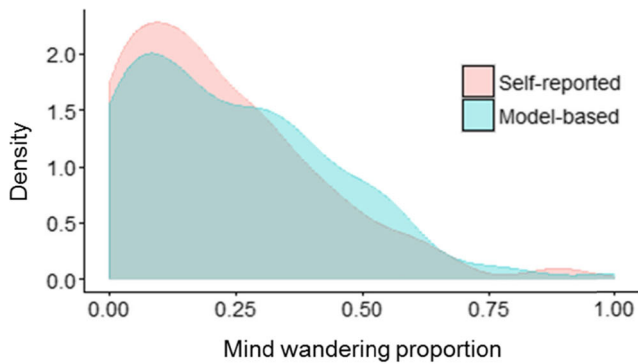
### Predictive validity

We expected that the mind-wandering proportions should be negatively correlated with text comprehension scores. As Table 6 illustrates, when we included the unclassified pages in the estimation process, the model-based measures were more strongly correlated with comprehension scores than were the self-reports ( $Z_H = 1.83$ ,  $p = .067$ ; Steiger, 1980). We do not consider this to be an artifact of the method used to estimate mind wandering for unclassified instances, because a similar pattern was reported on a different dataset when unclassified instances were discarded (Bixler & D'Mello, 2016). Instead, the stronger correlations might be due to eye gaze picking up other aspects of the reading process (i.e., fluency) beyond mind wandering. Alternatively, the model-based estimates might be more accurate than self-reports because they are not subject to the need to be aware that one is mind wandering and to other biases associated with self-reports.

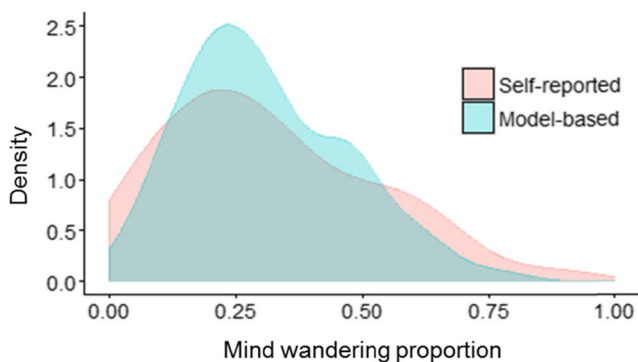
### Discussion

The goal of this study was to develop and validate an automatic objective measure of mind wandering during computerized reading, for use in psychological research as an alternative or complement to self-reports. Our results show that model-based mind-wandering proportions estimated from eye gaze data correlate with proportions of self-reported mind

### Ignoring unclassified instances



### Including unclassified instances



**Fig. 5** Density plots of the participant-level mind-wandering proportions for the self-reports and model-based estimates, after ignoring or including unclassified instances

wandering (convergent validity) and negatively predict comprehension (predictive validity).

Importantly, the measure generalizes beyond the training data by automatically estimating proportional mind-wandering scores for “new” participants from gaze alone. Along these lines, D’Mello et al. (2016) used an earlier variant of the measure to trigger real-time interventions based on predicted mind-wandering likelihoods on a *new* sample of 104 participants. The key finding was that model-based mind-wandering likelihoods negatively correlated with performance on comprehension questions that were either interspersed during reading ( $r = -.296, p < .05$ ) or appeared on a subsequent posttest ( $r = -.319, p < .05$ ). Using another variant of the

**Table 4** Odd–even reliability (Pearson’s  $r$ ) for the self-reported and model-based mind-wandering proportions after ignoring or including unclassified instances

Unclassified Instances	Self-Reported	Model-Based
Ignored	.622	.590
Included	.881	.751

model, Mills, Bixler, and D’Mello (2017) found that predicted likelihoods of mind wandering negatively correlated with scores on real-time self-explanation prompts ( $r = -.269, p = .175$ ) on a different sample of 27 participants (the nonsignificant correlation is attributed to the small sample size).

As such, our method can be implemented to produce an objective, automated measure of mind wandering in other studies. However, because we present a fully data-driven rather than prescriptive method for computationally deriving a gaze-based measure of mind wandering, the models need to be retrained for different domains. For instance, model parameters depend on the task that was used to collect the data (e.g., window length), automated feature selection identifies features that are most predictive for a specific data set, whereas the supervised classification methods learn how to associate features with mind-wandering reports, again for a given data set. We are in the process of developing domain-independent mind-wandering detection, but currently the models need to be retained on a subset of data collected in the domain of interest as in Hutt, Mills, White, Donnelly, and D’Mello (2016) and Mills et al. (2016).

The present approach also overcame several limitations with previous attempts to measure mind wandering from eye gaze data. In particular, our model was fully automatic in that it did not require manual inspection of data, unlike Loboda (2014); did not discard any cases with missing or noisy data, unlike Bixler and D’Mello (2015, 2016) and Loboda; and did not artificially manipulate the class distributions of the testing set, unlike D’Mello et al. (2013) and Bixler and D’Mello (2015). Therefore, in our view, the present work reflects the state of the art in fully automated mind-wandering detection. Taken together, our research suggests that researchers might not need to exclusively rely on self-reports for future studies on mind wandering, because objective gaze-based measurement might be a reality.

The measure has many applications beyond mind-wandering research. In many psychological studies, mind wandering is a nuisance variable rather than a variable of interest. Our approach provides an unobtrusive measure of mind wandering that can be used to partial out its confounding effect—for example, in studies on memory, visual perception, motor control, and so on. Therefore, the measure might be relevant to researchers in other fields of psychology.

Our work also has applications beyond the lab. Consumer-off-the-shelf (COTS) eyetrackers such as the Eye Tribe and the Tobii Eye X are cost-effective and mobile, which makes them suitable for research in more naturalistic settings (e.g., reading on a tablet in a classroom, library, or cafe). However, because of the lower quality of COTS eyetrackers and the decrease in experimenter control, data collected in more ecological settings are likely to be noisier than those collected in a lab setting. The present study has shown that

**Table 5** Correlations (Pearson's  $r$ ) between mind-wandering proportions and self-caught, retrospective, and trait-based mind wandering after ignoring or including unclassified instances

	Ignoring Unclassified Instances		Including Unclassified Instances	
	Self-Caught	Model-Based	Self-Caught	Model-Based
Self-caught mind wandering		.214*		.400***
How engaged were you while you were reading about soap bubbles?	.334***	.175*	.344***	.347***
While you were reading, was your attention focused on the text?	.304***	.200*	.284***	.384***
Trait-based mind wandering	.109	-.131	.133	-.021

\*\*\*  $p \leq .001$ , \*  $p < .05$

our model can provide a valid estimate of mind wandering using global gaze features, even when gaze data is of low quality or completely missing, thereby opening up promising avenues for research into mind wandering in more naturalistic settings.

Of course, these claims of generalizability to the wild need to be accompanied by a modicum of caution because in the present study, participants read one text in a lab setting using a using computerized reading paradigm that might not closely resemble naturalistic reading. Whether our approach generalizes to alternate reading tasks and texts thus remains to be explored. As a step in this direction, Hutt et al. (in press) used multiple COTS eyetracker to collect eye gaze data from roughly 14 to 30 high-school students at a time during interactions with a learning technology in their regular classroom. Using the same method used here, they were able to build a model to automatically detect mind wandering with accuracy scores that matched (and in some cases even exceeded) a model trained on data collected in a lab using another COTS eyetracker (Hutt et al., 2016).

There are multiple avenues to pursue in future research. For one, our model was trained on self-caught instances of mind wandering that are accompanied by metacognitive awareness. An open question is whether it picks up on instances of mind wandering that might not lead to the phenomenological experience of zoning out (e.g., brief or shallow lapses in attention). Another potential extension is to explore whether our model can discriminate amongst different types of mind wandering, be it with respect to content (e.g., task-unrelated thoughts versus task-related interferences) or intentionality (intentional or unintentional

mind wandering; Seli, Risko, & Smilek, 2016). It might be the case that different types of mind wandering are manifested via different signatures of eye gaze, and thus, should be distinguishable via our approach. If successful, these next-generation automated mind-wandering measures can provide novel insights into when minds begin to wander, the role of intentionality in mind wandering, and eventually on the nature of self-generated thoughts—all critical open questions in mind-wandering research (Smallwood & Schooler, 2015).

A limitation of the present model—and, in fact, of all current mind-wandering detectors—is that it is unable to classify each individual instance of mind wandering with the accuracy needed for psychological research. Although mind-wandering estimates at the participant level are valid (as we have shown here), these estimates are based on averages over instances, some of which are likely classified incorrectly—at least compared to self-reports of mind wandering. That being said, it is also possible that some of the instance-level disagreement between self-reports and the model estimates could be due to inaccurate reporting, either intentionally (e.g., due to social desirability biases) or accidentally (e.g., participants were unaware that they were mind wandering). Furthermore, self-caught reports and gaze features are likely to pick up on different aspects of mind wandering because they rely on different information sources. Given that it is unclear where the “ground truth” lies, combining objective and subjective measures of mind wandering might be the most defensible approach in the near future.

In conclusion, the last decade has witnessed unprecedented progress in advancing the science of self-generated thought (Christoff et al., 2016), and especially mind wandering (Smallwood & Schooler, 2015). However, research has been stymied by a lack of valid objective measures. In some ways we are still in the dark ages, given the almost exclusive reliance on self-reports to measure these phenomena. By showing that it is possible to develop a valid measure of mind wandering based on eye gaze that will generalize to new participants (albeit in a restricted context of computerized reading in the lab), we hope to have taken a step toward the light.

**Table 6** Predictive validity (Pearson's  $r$ ) based on the relationship between mind-wandering proportions and comprehension scores

Unclassified Instances	Self-Reported	Model-Based
Ignored	-.202*	-.134
Included	-.208*	-.374**

$N = 132$ . \*\*  $p < .001$ , \*  $p < .05$ .

**Author note** This research was supported by the National Science Foundation (NSF; Grant Nos. DRL 1235958 and IIS 1523091). Any opinions, findings and conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the NSF.

## References

- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. doi:10.1146/annurev.neuro.28.061604.135709
- Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*, 28–58. doi:10.1111/j.1745-6916.2006.00003.x
- Bixler, R., & D’Mello, S. (2014). Toward fully automated person independent detection of mind wandering. In J. Masthoff (Ed.), *User modeling, adaptation, and personalization* (pp. 37–48). Berlin: Springer.
- Bixler, R., & D’Mello, S. (2015). Automatic gaze-based detection of mind wandering with metacognitive awareness. In *User modeling, adaptation, and personalization: Proceedings of the 20th International Conference (UMAP 2012), Montreal, Canada, July 16–20, 2012* (pp. 31–43). Berlin, Germany: Springer.
- Bixler, R., & D’Mello, S. (2016). Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, *26*, 33–68. doi:10.1007/s11257-015-9167-1
- Blanchard, N., Bixler, R., Joyce, T., & D’Mello, S. (2014). Automated physiological-based detection of mind wandering during learning. In *Intelligent tutoring systems* (pp. 55–60). Berlin: Springer.
- Boys, C. V. (1890). *Soap-bubbles, and the forces which mould them*. Ithaca: Cornell University Library.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.
- Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, *106*, 8719–8724. doi:10.1073/pnas.0900234106
- Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nature Reviews Neuroscience*, *17*, 718–731. doi:10.1038/nrn.2016.113
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290. doi:10.1037/1040-3590.6.4.284
- Craney, T. A., & Surlis, J. G. (2002). Model-dependent variance inflation factor cutoff values model-dependent variance inflation factor cutoff values. *Quality Engineering*, *14*, 391–403. doi:10.1081/QEN-120001878
- D’Mello, S., Cobian, J., & Hunter, M. (2013). Automatic gaze-based detection of mind wandering during reading. In *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 364–365). Boston: International Educational Data Mining Society.
- D’Mello, S. K. (2016). Giving eyesight to the blind: Towards attention-aware AIED. *International Journal of Artificial Intelligence in Education*, *26*, 645–659. doi:10.1007/s40593-016-0104-1
- D’Mello, S. K., Duckworth, A., & Dieterle, E. (2017). *Advanced, analytic, automated (AAA) measurement of engagement during learning*. Manuscript under review.
- D’Mello, S., Kopp, K., Bixler, R. E., & 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 2015)* (pp. 1661–1669). New York, NY: ACM Press. doi:10.1145/2851581.2892329
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*, 78–87.
- Drummond, J., & Litman, D. (2010). In the zone: Towards detecting student zoning out using supervised machine learning. In *Intelligent Tutoring Systems* (pp. 306–308). Berlin: Springer.
- Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*, 259–268. doi:10.1016/j.jrp.2011.02.004
- Faber, M., Mills, C., Kopp, K., & D’Mello, S. (2016). The effect of disfluency on mind wandering during text comprehension. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-016-1153-z. Advance online publication.
- Feng, S., D’Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, *20*, 586–592. doi:10.3758/s13423-012-0367-y
- Foulsham, T., Farley, J., & Kingstone, A. (2013). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology*, *67*, 51–59. doi:10.1037/a0030217
- Frank, D. J., Nara, B., Zavagnin, M., Touron, D. R., & Kane, M. J. (2015). Validating older adults’ reports of less mind-wandering: An examination of eye movements and dispositional influences. *Psychology and Aging*, *30*, 266–278. doi:10.1037/pag0000031
- Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: Pupillometry of spontaneous thought while reading. *Quarterly Journal of Experimental Psychology*, *66*, 2289–2294. doi:10.1080/17470218.2013.858170
- Franklin, M. S., Smallwood, J., & Schooler, J. W. (2011). Catching the mind in flight: Using behavioral indices to detect mindless reading in real time. *Psychonomic Bulletin & Review*, *18*, 992–997. doi:10.3758/s13423-011-0109-6
- Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind wandering. *Frontiers in Psychology*, *5*, 31. doi:10.3389/fpsyg.2014.00031
- Hall, M. (1999). *Correlation-based feature selection for machine learning* (PhD thesis). Department of Computer Science, University of Waikato, Hamilton, New Zealand.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, *11*, 10–18.
- Hawkins, G. E., Mittner, M., Boekel, W., Heathcote, A., & Forstmann, B. U. (2015). Toward a model-based cognitive neuroscience of mind wandering. *Neuroscience*, *310*, 290–305. doi:10.1016/j.neuroscience.2015.09.053
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Hutt, S., Mills, C., Bosch, N., Krasich, K., Brockmole, J. R., & D’Mello, S. K. (in press). Out of the fr-eye-ing pan: Toward gaze-based, attention-aware cyberlearning in classrooms. *Journal*.
- Hutt, S., Mills, C., White, S., Donnelly, P. J., & D’Mello, S. K. (2016). The eyes have it: Gaze-based detection of mind wandering during learning with an intelligent tutoring system. In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining 2016* (pp. 86–93). Boston: International Educational Data Mining Society.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480. doi:10.1016/0010-0285(76)90015-3

- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354. doi:10.1037/0033-295X.87.4.32
- Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when an experience-sampling study of working memory and executive control in daily life. *Psychological Science*, *18*, 614–621.
- Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*, 932–932. doi:10.1126/science.1192439
- Kopp, K., D’Mello, S., & Mills, C. (2015). Influencing the occurrence of mind wandering while reading. *Consciousness and Cognition*, *34*, 52–62. doi:10.1016/j.concog.2015.03.003
- Kucyi, A., & Davis, K. D. (2014). Dynamic functional connectivity of the default mode network tracks daydreaming. *NeuroImage*, *100*, 471–480. doi:10.1016/j.neuroimage.2014.06.044
- Loboda, T. D. (2014). *Study and detection of mindless reading*. Pittsburgh: University of Pittsburgh.
- McNamara, D. S., & Magliano, J. P. (2009). Self-explanation and metacognition: The dynamics of reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 60–81). New York: Routledge.
- McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 196–204. doi:10.1037/a0014104
- Mills, C., Bixler, R., Wang, X., & D’Mello, S. (2016). Automatic gaze-based detection of mind wandering during film viewing. In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining 2016* (pp. 30–37). Boston: International Educational Data Mining Society.
- Mills, C., & D’Mello, S. (2015). Toward a real-time (day) dreamcatcher: Detecting mind wandering episodes during online reading. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 69–76). Boston, MA: International Educational Data Mining Society.
- Mills, C., D’Mello, S. K., & Kopp, K. (2015). The influence of consequence value and text difficulty on affect, attention, and learning while reading instructional texts. *Learning and Instruction*, *40*, 9–20. doi:10.1016/j.learninstruc.2015.07.003
- Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the brain takes a break: A model-based analysis of mind wandering. *Journal of Neuroscience*, *34*, 16286–16295. doi:10.1523/JNEUROSCI.2062-14.2014
- Mittner, M., Hawkins, G. E., Boekel, W., & Forstmann, B. U. (2016). A neural model of mind wandering. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2016.06.004
- Mrazek, M. D., Phillips, D. T., Franklin, M. S., Broadway, J. M., & Schooler, J. W. (2013). Young and restless: Validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Frontiers in Psychology*, *4*, 560. doi:10.3389/fpsyg.2013.00560
- O’Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A., Foxe, J. J., & Kelly, S. P. (2009). Uncovering the neural signature of lapsing attention: Electrophysiological signals predict errors up to 20 s before they occur. *Journal of Neuroscience*, *29*, 8604–8611. doi:10.1523/JNEUROSCI.5967-08.2009
- Pham, P., & Wang, J. (2015). AttentiveLearner: Improving mobile MOOC learning via implicit heart rate tracking. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial intelligence in education* (Vol. 9112, pp. 367–376). Berlin: Springer.
- Randall, J. G., Oswald, F. L., & Beier, M. E. (2014). Mind-wandering, cognition, and performance: A theory-driven meta-analysis of attention regulation. *Psychological Bulletin*, *140*, 1411–1431. doi:10.1037/a0037428
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. doi:10.1037/0033-2909.124.3.372
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157. doi:10.1037/0033-295X.105.1.125
- Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, *21*, 1300–1310. doi:10.1177/0956797610378686
- Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). “Oops!”: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*, 747–758. doi:10.1016/S0028-3932(97)00015-8
- Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your mind wanders weakly, your mind wanders deeply: Objective measures reveal mindless reading at different levels. *Cognition*, *125*, 179–194.
- Schooler, J. W., Reichle, E. D., & Halpern, D. V. (2004). Zoning out while reading: Evidence for dissociations between experience and metacognition. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 203–226). Cambridge: MIT Press.
- Seibert, P. S., & Ellis, H. C. (1991). Irrelevant thoughts, emotional mood states, and cognitive task performance. *Memory & Cognition*, *19*, 507–513.
- Seli, P., Carriere, J. S. A., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology*, *4*, 430. doi:10.3389/fpsyg.2013.00430
- Seli, P., Carriere, J. S. A., Thomson, D. R., Cheyne, J. A., Martens, K. A. E., & Smilek, D. (2014). Restless mind, restless body. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 660–668. doi:10.1037/a0035260
- Seli, P., Risko, E. F., & Smilek, D. (2016). On the necessity of distinguishing between unintentional and intentional mind wandering. *Psychological Science*, *27*, 685–691. doi:10.1177/0956797616634068
- Smallwood, J. (2011). Mind-wandering while reading: Attentional decoupling, mindless reading and the cascade model of inattention. *Linguistics and Language Compass*, *5*, 63–77. doi:10.1111/j.1749-818X.2010.00263.x
- Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *Journal of Cognitive Neuroscience*, *20*, 458–469.
- Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., & Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS ONE*, *6*, e18298. doi:10.1371/journal.pone.0018298
- Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O’Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, *13*, 657–690. doi:10.1016/j.concog.2004.06.003
- Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review*, *14*, 230–236. doi:10.3758/BF03194057
- Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition*, *36*, 1144–1150. doi:10.3758/MC.36.6.1144
- Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*, 946–958. doi:10.1037/0033-2909.132.6.946
- Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*, 487–518. doi:10.1146/annurev-psych-010814-015331



- Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010). Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychological Science, 21*, 786–789. doi:10.1177/0956797610368063
- Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D'Argembeau, A. (2011). Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica, 136*, 370–381. doi:10.1016/j.actpsy.2011.01.002
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251. doi:10.1037/0033-2909.87.2.245
- Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 832–842. doi:10.1037/a0029669
- Uzzaman, S., & Joordens, S. (2011). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition, 20*, 1882–1886. doi:10.1016/j.concog.2011.09.010
- Voßkühler, A., Nordmeier, V., Kuchinke, L., & Jacobs, A. M. (2008). OGAMA (Open Gaze and Mouse Analyzer): Open-source software designed to analyze eye and mouse movements in slideshow study designs. *Behavior Research Methods, 40*, 1150–1162. doi:10.3758/BRM.40.4.1150
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070. doi:10.1037/0022-3514.54.6.1063
- Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience, 9*, 971–978. doi:10.1038/nrn1727