

# Identifying Learning Conditions that Minimize Mind Wandering by Modeling Individual Attributes

Kristopher Kopp<sup>1</sup>, Robert Bixler<sup>2</sup>, and Sidney D'Mello<sup>1,2</sup>

<sup>1</sup>Department of Psychology

<sup>2</sup>Computer Science, University of Notre Dame, South Bend, IN 46556, USA  
{kkopp, rbixler, sdmello}@nd.edu

**Abstract.** The propensity to involuntarily disengage by zoning out or mind wandering (MW) is a common phenomenon that has negative effects on learning. The ability to stay focused while learning from instructional texts involves factors related to the text, to the task, and to the individual. This study explored the possibility that learners could be placed in optimal conditions (task and text) to reduce MW based on an analysis of individual attributes. Students studied four texts which varied along dimensions of value and difficulty while reporting instances of MW. Supervised machine learning techniques based on a small set of individual difference attributes determined the optimal condition for each participant with some success when considering value and difficulty separately (kappas of .16 and .24; accuracy of 59% and 64% respectively). Results are discussed in terms of creating a learning system that prospectively places learners in the optimal condition to increase learning by minimizing MW.

**Keywords:** engagement, mind wandering, affect, machine learning.

## 1 Introduction

Advances in research on intelligent tutoring systems (ITSs) have recently intertwined aspects of the cognitive sciences with the affect sciences [1,2,3,4]. ITSs have evolved from systems that emphasize modeling student cognition [5,6] to systems that detect and respond to student affect as well [7,8,9]. One related area of interest is learner engagement. Engagement has been defined as a state of involvement in some activity or task with focused attention and intense concentration [7]. Engagement is a necessary condition to learning since learners have to attend to information in order to learn. It is not uncommon, however, for students to experience involuntary lapses in attention and suddenly realize that they were thinking about things totally unrelated to the learning content. Such *mind wandering* (MW) activities can be detrimental to learning [10,11], so it is important to develop systems that can sustain engagement by reducing the propensity of MW behaviors. The goal of this paper is to take steps towards developing a preventative system with the ability to place students in an optimal learning condition that would result in the least amount of MW based on measures of individual difference attributes.

## 1.1 Related Works

Recently, researchers have been interested in the relationship between affect and learning. D’Mello [2] conducted a meta-analysis of 24 studies that investigated the influence of student affective states on learning. Basic affective states, such as anger, fear, happiness, etc. [12], are considered to have specific and culturally unanimous qualities to them that make them rather distinguishable and easy to detect. However, it is the non-basic affective states (e.g., confusion, boredom and engagement) that were more frequent during learning with ITSs. For example, Craig and colleagues [13] identified significant and positive relationship between confusion and learning when interacting with an ITS.

Similarly, Baker and colleagues [7] observed the presence of non-basic affective states of students while they interacted with various ITSs. One of their main findings was that when boredom occurred, it was difficult to get the students to re-engage in the learning task. Instead, students experiencing boredom exhibited a propensity to engage in behaviors such as “gaming the system.” They also found that confusion and engagement were the most prevalent states and better precursors to learning than boredom since those who chose to game the system do not learn.

The studies mentioned above are just a few examples of research identifying affective states during interactions with ITSs and the different types of repercussions they can have. Research along these lines has led to the development of **Reactive** affect-sensitive ITSs that attempt to sense affective states that could have an effect on learning and respond accordingly [1], [14,15]. One of the early examples of this type of system is Affective AutoTutor [16] which detects specific emotions (i.e., boredom, confusion) based on conversational modeling, facial cues, and body language and alters the dynamics of the tutoring session to react to the learner through dialog moves designed to address specific affective states.

With respect to mind wandering, Drummond and Litman [17] attempted to identify episodes of “zoning out” while students were engaged in a spoken dialog with an ITS. Students were periodically interrupted to complete a short survey to indicate the extent to which they were focusing on the task (low zoning out) or on other thoughts (high zoning out). J48 decision trees trained on acoustic-prosodic features extracted from the students’ utterances yielded 64% accuracy in discriminating high vs. low zone-outs. The next step in this line of research would be for the ITS to respond when zone-outs are detected. A system called GazeTutor [8] attempted this by using eye tracking to assess a lack of attention and responded with interventions to re-engage learners. Thus, based on affect detection methodologies, systems are able to identify and respond to affective states to increase learning.

## 1.2 The Current Project

An alternative to reacting to affective states as they arise is to implement **Proactive** strategies that attempt to create or foster affective states that would be beneficial for learning. Here, we focus on engagement since it is a necessary condition for learning. Engagement is considered to have three components: a cognitive, an affective, and a behavioral component [18]. The affective and behavioral components have been extensively studied in previous ITS research (e.g., [19,20]); hence, our present

emphasis is on the cognitive component, specifically momentary lapses of attention or MW which has been shown to have a detrimental influence on learning under various conditions [10].

Our approach is motivated by the assumption that engagement emerges from an intersection of factors related to the learning task itself (e.g., task difficulty), factors related to the perceptions of the learning activity (e.g., task value), and factors related to the individual performing the task (e.g., abilities and traits) [21]. The unique interaction will differ among individuals depending on their own unique traits. The purpose of our overall project is to investigate whether or not we can capitalize on this interaction and place students into an ideal learning condition (i.e., influenced by text and task factors) based on the factors related to the learner (i.e., abilities and traits) that would lead to the least amount of MW.

As an initial step in this direction, we first considered the possibility of using machine learning techniques to predict the learning condition that was optimal in terms of minimizing MW for a specific learner based on his or her attributes. To do this, we collected a large data set where students were asked to study about scientific research methods from instructional texts. During learning, students were asked to report incidents of MW using standard probe-based methods [10]. Each student was exposed to four conditions that varied in combinations of difficulty (easy or difficult) and value (high or low value) of the text. Students also completed multiple measures of individual attributes. Ideal conditions were identified for each student as defined by the least proportion of MW reports. Supervised machine learning was used to predict the ideal condition for each student using their individual attributes as features.

## 2 Data Collection

### 2.1 Participants

Undergraduate students ( $N = 187$ ) from two U.S. universities participated for course credit. 105 students were recruited from a medium-sized private mid-western university while 82 were from a large public university in the mid-south. The average age was 19.7 years ( $SD = 2.65$ ).

### 2.2 Texts and Task Context

Students learned from four different texts, on a computer screen, on research methods topics (i.e., experimenter bias, replication, causality, and dependent variables). The texts contained 1500 words on average ( $SD = 10$ ) and were split into 30-36 pages. The difficulty manipulation consisted of presenting either an easy or a difficult version of each text. Texts were made more difficult by replacing words and sentences with more complex versions while retaining content, length, and semantics. The value manipulation was modeled after a common strategy used by instructors during review sessions before exams. Specifically, value was manipulated based on the weight assigned to each text on a subsequent posttest. Questions corresponding to the “high-value” texts counted three times more toward the test score than questions

for the “low-value” texts. Students were made aware of this before reading each text. Thus students saw all four texts with 1 text in each one of the 4 conditions: 2 (difficulty: easy vs. difficult)  $\times$  2 (value: high vs. low). The success of the manipulations was confirmed with self-reports of the perceived difficulty and perceived value of the texts (see [22]).

### 2.3 Measures

**Mind Wandering** was measured through auditory probes, a standard and validated method for collecting online MW report [10]. Nine pseudorandom pages in each text were identified as “probe pages.” When a student encountered a probe page, an auditory probe (i.e., a beep) was triggered at a randomly chosen time interval 4 to 12 seconds from the time the page appeared. Students were instructed to indicate if they were MW or not by pressing keys marked “Yes” or “No,” respectively. The MW rate for each text was then obtained by computing the proportion of “Yes” responses to probes.

**Individual Attribute** measures were collected for use as features in our models. The following measures were collected: (a) performance scores of the Nelson Denny self-paced *reading comprehension test* [23], (b) median sentence reading time of the Nelson Denny test as a measure of *reading fluency*, (c) performance on the *reading span* test as a measure of *working memory ability* [24], (d) *interest in research methods*, measured using a Topic Interest Scale adapted from Linnenbrink-Garcia et al. [25], (e) the Boredom Proneness scale measured the participant’s *trait behavior of general boredom* [26], (f, g) the Academic Boredom Survey [27] measured traits specific to *boredom in academic situations* when overwhelmed and underwhelmed (considered separately), (h) self-reported ACT/SAT scores from each participant as a measure of *scholastic aptitude*, and (i) pretest performance on an assessment of the target concepts as *prior knowledge*. Scores of all measures were standardized by school to alleviate any large discrepancies due to population differences between schools.

### 2.4 Procedure

First, students filled out a brief demographic survey and completed the Nelson Denny test. Second, students completed one of two multiple choice pretests (counterbalanced between pre and posttest across all students) comprised of 24 deep-reasoning questions. Students were then given the topic interest measure. Students next received instruction on the learning task and how to respond to the MW probes based on instructions taken from previous task studies [28]. All students studied four texts (one at a time) for an average of 32.4 mins ( $SD = 9.09$ ) on a page-by-page basis, using the space bar to navigate forward. The name of the topic and the corresponding weight of the test questions (value manipulation) were explicitly presented before each text. After students studied all four texts, they were presented with the remaining 24 item posttest. They then completed several additional measures: the boredom proneness scale; the academic boredom survey; and a reading span test.

### 3 Supervised Classification

Our principal goal was to assess our ability to place a student in a learning condition that would result in the least amount of MW reports. Each data point corresponded to one participant and was labeled with the conditions (difficulty and value) of the text with lowest rate of MW resulting in 187 data points. We then attempted to predict this optimal condition using nine measures of individual attributes as features using supervised machine learning.

#### 3.1 Model Building

The WEKA machine learning software tool’s [29] implementations of 34 machine learning algorithms were used to build models predicting which text condition (difficulty and value) led to optimal values of MW reports. There were two additional parameters for the classification task. The first parameter was a *threshold* for the difference between the standardized MW rate for the best and worst condition. For each data point, if this difference was above the threshold the data point was included in the data set. This allowed us to consider only those who reported a meaningful difference of MW between conditions. Values used for this threshold included 0, 0.25, and 0.5 standard deviations. The second parameter was the *classification task*. In addition to classifying across all four conditions, we collapsed difficulty across value and vice versa, resulting in two additional classification tasks: classifying difficult texts from easy texts, and high value texts from low value texts. This resulted in 408 models (4 classification task  $\times$  3 difference threshold  $\times$  34 classifiers) and the classifier that yielded the best model for each parameter combination was retained for analysis.

#### 3.2 Model Validation

Models were evaluated using leave-one-student-out cross validation. The model was trained on all but one student, which was then used to predict the best text condition for the remaining student. This process was repeated until each student had been classified in this way. This method ensures generalizability across students because each of the training and testing sets are student-independent. The Kappa statistic was taken as the measure of classifier accuracy since it is less sensitive to variations in data distribution.

## 4 Results

We first assessed any differences to assigned conditions across all three classification tasks for the threshold value of 0. When considering all four conditions of difficulty  $\times$  value, 26% of the students reported the least amount of MW in the easy and low value condition, 28% reported in the easy and high value condition, 28% reported in the difficult and low value, and 18% reported in the difficult and high value condition. When considering value and difficulty separately, 53% of the students reported the

least amount of MW in the low value condition and 57% in the easy condition when considering text difficulty. Thus, these differences indicate that there is not one single, optimal condition for all students.

#### 4.1 Classification Accuracy

We first analyzed models that were built in an attempt to place individuals into one of the four ideal conditions (i.e., easy and low value, easy and high value, difficult and low value, difficult and high value) based on the nine individual attribute measures (i.e., features). As can be seen in Table 1, the best classification (i.e., highest kappa) occurred when we discriminated .25 standard deviations between the highest and least amount of MW reports between conditions with a Decision Stump classifier.

In addition to attempting to classify according to the four conditions, we collapsed MW reports across value and then difficulty and assessed each separately. As can be seen in Table 1, when collapsing across value conditions, the best classification occurred when we discriminated .25 standard deviations between the highest and least amount of MW reports between conditions with a Simple Logistic Classifier. Similarly, when collapsing across difficulty conditions, the best classification occurred when we discriminated .5 standard deviations between the highest and least amount of MW reports between conditions with a Decision Stump Classifier.

**Table 1.** Classification results

<b>Classification Task</b>	<b>Classification Threshold</b>	<b>Kappa</b>	<b>Observed Accuracy</b>	<b>Expected Accuracy</b>	<b>N</b>
<i>Difficulty × Value</i>	0	.03	.27	.25	187
	<b>.25</b>	<b>.11</b>	<b>.34</b>	<b>.26</b>	<b>141</b>
	.5	.06	.31	.26	98
<i>Value</i>	0	.01	.51	.50	187
	<b>.25</b>	<b>.16</b>	<b>.59</b>	<b>.51</b>	<b>141</b>
	.5	.13	.56	.50	98
<i>Difficulty</i>	0	.05	.54	.51	187
	.25	.05	.55	.52	141
	<b>.5</b>	<b>.24</b>	<b>.64</b>	<b>.53</b>	<b>98</b>

Note: The kappa value is calculated using the formula  $(\text{Observed Accuracy} - \text{Expected Accuracy}) / (1 - \text{Expected Accuracy})$ , where Observed Accuracy is equivalent to recognition rate and Expected Accuracy is estimated from the marginal probabilities in the confusion matrix.

#### 4.2 Features

We next considered the correlations between the performance on individual attribute measure (i.e., features) and placement in the optimal conditions of the value and difficulty classification tasks. For value, the conditions were dummy coded as low = 0 and high = 1. For difficulty, easy = 0 and difficult = 1. As can be seen in Table 2, there are some similarities and some differences with respect to the features that correlate with optimal classification for each classification task. With regard to the

highest correlations, for value, for students who have a higher propensity to experience boredom during academic situations that are underwhelming, the low value condition would be the optimal condition for the least amount of mind wandering. On the other hand, for difficulty, students with the propensity to experience boredom during overwhelming situations would benefit from having more difficult text. Additionally, the topic interest measure shows that a student may benefit from a more difficult text if they have a high amount of interest in the topic.

**Table 2.** Correlations (pearson  $r$ 's) between performance on the individual attribute measures (i.e., features) optimal conditions of classification task dummy coded for value (low = 0 and high = 1) and difficulty (easy = 0 and difficult = 1)

Individual Attribute	Value ( $n = 141$ )	Difficulty ( $n = 98$ )
Working Memory	.10	-.03
Academic Boredom (Overwhelmed)	-.03	.20
Academic Boredom (Underwhelmed)	-.16	-.04
General Boredom	-.05	.09
Prior Knowledge	-.03	-.09
Reading Fluency	.10	.09
Reading Comprehension	.01	.01
Topic Interest (Research Methods)	-.05	.18
Scholastic Aptitude	-.04	.01

## 5 Discussion

The negative influence of mind wandering (MW) on learning coupled with the frequency of MW suggest that educational technologies could benefit by prospectively selecting learning conditions to reduce the incidence of MW. As an initial step in this direction, our hypothesis was that it was possible to determine an optimal learning condition that would lead to a lowered rate of MW based on a relatively modest set of nine individual attributes.

There was not a single condition that was optimal for all students, which suggests that even though *on average* one condition might yield lower MW rates than others, assigning every student to the same condition is not an optimal strategy since individual differences matter. We attempted to capitalize on those differences and our results show that it is possible to determine the condition that leads to the lowest rate of MW for an individual by considering that individual's trait attributes. Removing students with stable MW rates across all conditions improved our kappas from .03, .01 and .05 to .11, .16 and .24, for difficulty  $\times$  value, value, and difficulty, respectively. This method of participant removal is justified because a participant whose MW rate does not change across condition does not add any meaningful variability to model. Furthermore, individuals who do not have different rates of MW

across conditions could not possibly have their MW rate lowered by altering condition no matter their individual attributes.

We acknowledge that classification rates were modest, even for the best models. However, one needs to consider the difficulty of the task in that we are attempting to prospectively predict a task condition that yielded the lowest rates of MW from a set of sparse individual difference measures alone, despite the fact that MW is an extremely complex and elusive state that is likely influenced by numerous additional factors. Furthermore, we have some confidence in the generalizability of our results because we employed a leave-one-subject-out validation method and our data included students from a medium-sized private mid-western university and a large public mid-south university with very different characteristics.

The usefulness of this research depends on how it can act to influence future designs of ITSs that intend to increase learner engagement by minimizing off-task thought. It may be of interest for designers of these systems to be able to predict mind wandering behaviors from attributes of the learner in order to advance preventative technologies. From our results, it was difficult to accurately predict conditions when including students who did not deviate in their MW behaviors across conditions in a meaningful way. This work does show, however, that it is possible to predict optimal conditions for those who show some contrast of mind wandering behaviors between different learning conditions. It may be that ITSs would benefit from initially targeting those whose mind wandering behaviors are somewhat different under different learning conditions.

There were some limitations of this work. First, the method of tracking MW through auditory probes is subject to students providing an incorrect rate of MW. An incorrectly reported rate of MW would result in our models being trained on data which was not completely correct. This would ultimately make classification more difficult. However, many studies have used this method of measuring MW as there are no alternatives to tracking this highly internal phenomenon (see [10] for a review), so we are confident that we are adhering to state of the art methods. Second, these findings are based on a task that requires studying texts on research methods. Future studies may consider incorporating other topics and other modes of information delivery to ensure generalizability. Furthermore, the present study was conducted in a laboratory context, so replication in more ecological learning situations is warranted.

This paper reports a first step towards a proactive learning system to reduce the rates of MW. The present work demonstrated the ability to select the best condition of easy and difficult text or high and low value for a learner to have the lowest rate of MW based on the learner's individual attributes. Our approach generalizes to individuals due to the method of validation and the diversity of the students. The next step is to use the best models in a personalized learning environment that optimizes the potential for the least amount of mind wandering during a learning session by personalizing the experience based on the measures of individual differences. For example, for each learner, the environment can prescribe conditions that minimize MW. MW and learning associated with this personalized environment can then be compared to control conditions (e.g., randomly assigning learners to condition or assigning all learners to the condition that resulted in the lowest MW overall). Whether, the proposed approach outperforms these alternatives awaits further research.



**Acknowledgment.** This research was supported by the National Science Foundation (NSF) (ITR 0325428, HCC 0834847, DRL 1235958). 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.

## References

1. Calvo, R.A., D’Mello, S.K.: Frontiers of affect-aware learning technologies. *IEEE Intelligent Systems* 27(6), 86–89 (2012)
2. D’Mello, S.K.: A Selective Meta-analysis on the Relative Incidence of Discrete Affective States during Learning with Technology. *Journal of Educational Psychology* (2013)
3. D’Mello, S.K., Graesser, A.C.: Feeling, Thinking, and Computing with Affect-Aware Learning Technologies. In: Calvo, R.A., D’Mello, S.K., Gratch, J., Kappas, A. (eds.) *Handbook of Affective Computing*. Oxford University Press (in press)
4. Picard, R.W.: *Affective Computing*. MIT Press, Cambridge (1997)
5. Graesser, A.C., Olney, A., Haynes, B.C., Chipman, P.: AutoTutor: A cognitive system that simulates a tutor that facilitates learning through mixed-initiative dialogue. In: Forsythe, C., Bernard, M.L., Goldsmith, T.E. (eds.) *Cognitive Systems: Human Cognitive Models in Systems Design*. Erlbaum, Mahwah (2005)
6. VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., Rosé, C.P.: When are tutorial dialogues more effective than reading? *Cognitive Science* 31(1), 3–62 (2007)
7. Baker, R., D’Mello, S., Rodrigo, M., Graesser, A.: Better to be frustrated than bored: The incidence, persistence, and impact of learners’ cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies* 68(4), 223–241 (2010)
8. D’Mello, S., Olney, A., Williams, C., Hays, P.: Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies* 70(5), 377–398 (2012)
9. Woolf, B., Bursleson, W., Arroyo, I., Dragon, T., Cooper, D., Picard, R.: Affect-aware tutors: Recognizing and responding to student affect. *International Journal of Learning Technology* 4(3/4), 129–163 (2009)
10. Mooneyham, B.W., Schooler, J.W.: The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale* 67(1), 11 (2013)
11. Szpunar, K.K., Moulton, S.T., Schacter, D.L.: Mind Wandering and education: From the classroom to online learning. *Frontiers in Psychology*, 4 (2013)
12. Ekman, P.: An argument for basic emotions. *Cognition & Emotion* 6(3-4), 169–200 (1992)
13. Craig, S., Graesser, A., Sullins, J., Gholson, B.: Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media* 29(3), 241–250 (2004)
14. Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Alevan, V., Kusbit, G., Ocumpaugh, J., Rossi, L.: Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. In: *Proceedings of the 5th International Conference on Educational Data Mining*, pp. 126–133 (2012)
15. Conati, C., Maclaren, H.: Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction* 19(3), 267–303 (2009)
16. D’Mello, S.K., Jackson, G.T., Craig, S.D., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., el Kaliouby, R., Picard, R., Graesser, A.C.: AutoTutor detects and responds to learners affective and cognitive states. In: *Workshop on Emotional and Cognitive Issues at the International Conference on Intelligent Tutoring Systems* (2008)

17. Drummond, J., Litman, D.: In the Zone: Towards Detecting Student Zoning Out Using Supervised Machine Learning. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010, Part II. LNCS, vol. 6095, pp. 306–308. Springer, Heidelberg (2010)
18. Fredricks, J.A., Blumenfeld, P.C., Paris, A.H.: School engagement: Potential of the concept, state of the evidence. *Review of Educational Research* 74(1), 59–109 (2004)
19. Pekrun, R., Linnenbrink-Garcia, L.: Academic emotions and student engagement. In: *Handbook of Research on Student Engagement*, pp. 259–282. Springer, US (2012)
20. Gregory, A., Allen, J.P., Mikami, A.Y., Hafen, C.A., Pianta, R.C.: Effects of a professional development program on behavioral engagement of students in middle and high school. *Psychology in the Schools* 51(2), 143–163 (2014)
21. Snow, C.: *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Corporation, Santa Monica (2002)
22. Mills, C., D’Mello, S.K.: How Do Extrinsic Value and Difficulty Impact Engagement: An Experimental Approach (in prep.)
23. Brown, J.I.: *The Nelson-Denny Reading Test* (1960)
24. Daneman, M., Carpenter, P.A.: Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19(4), 450–466 (1980)
25. Linnenbrink-Garcia, L., Durik, A.M., Conley, A.M., Barron, K.E., Tauer, J.M., Karabenick, S.A., Harackiewicz, J.M.: Measuring Situational Interest in Academic Domains. *Educational and Psychological Measurement* 70(4), 647–671 (2010)
26. Farmer, R., Sundberg, N.D.: Boredom proneness—the development and correlates of a new scale. *Journal of Personality Assessment* 50(1), 4–17 (1986)
27. Acee, T.W., Kim, H., Kim, H.J., Kim, J.I., Chu, H.N.R., Kim, M., Cho, Y., Wicker, F.W.: Academic boredom in under- and over-challenging situations. *Contemporary Educational Psychology* 35(1), 17–27 (2010)
28. Feng, S., D’Mello, S., Graesser, A.C.: Mind Wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, 1–7 (2013)
29. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter* 11(1), 10–18 (2009)
30. Kononenko, I.: Estimating attributes: Analysis and extensions of RELIEF. In: Bergadano, F., De Raedt, L. (eds.) *ECML 1994*. LNCS, vol. 784, pp. 171–182. Springer, Heidelberg (1994)