Eye-Tracking to Predict User Cognitive Abilities and Performance for User-Adaptive Narrative Visualizations

Oswald Barral†, Sébastien Lallé†, Grigorii Guz, Alireza Iranpour, Cristina Conati
The University of British Columbia
Vancouver, BC, Canada
{obarral, lalles, alireza.iranpour, g.guz, conati}@cs.ubc.ca

ABSTRACT

We leverage eye-tracking data to predict user performance and levels of cognitive abilities while reading magazine-style narrative visualizations (MSNV), a widespread form of multimodal documents that combine text and visualizations. Such predictions are motivated by recent interest in devising user-adaptive MSNVs that can dynamically adapt to a user's needs. Our results provide evidence for the feasibility of real-time user modeling in MSNV, as we are the first to consider eye tracking data for predicting task comprehension and cognitive abilities while processing multimodal documents. We follow with a discussion on the implications to the design of personalized MSNVs.

CCS CONCEPTS

• **Human-centered computing ~ User models; Information visualization** • Computing methodologies ~ Classification and regression trees

KEYWORDS

Eye-tracking; adaptive visualizations; narrative visualizations; user model; user characteristics; classification

ACM Reference format:

Oswald Barral, Sébastien Lallé, Grigorii Guz, Alireza Iranpour, and Cristina Conati. 2020. Eye-Tracking to Predict User Cognitive Abilities and Performance for User-Adaptive Narrative Visualizations. In *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI'20). ACM, Virtual event, The Netherlands, 11 pages.* https://doi.org/10.1145/3382507.3418884

1 Introduction

Eye-tracking has been extensively used in cognitive psychology for understanding various aspects of human cognition (e.g., [29,30]), as well as in human computer interaction (HCI) for off-line evaluation of interface design or as an alternative form of intended user input (e.g., [38]). In recent years, however, eye-tracking has also been investigated as a source of information to infer relevant user states to personalize interaction with specific applications. Examples range from relevance feedback for information retrieval (e.g., [12]) to cognitive load assessment for emergency situations (e.g., [3]), learning gains with educational software (e.g., [9]), or automatic detection of personality traits [8].

Due to its perceptual nature, Information Visualization (InfoVis) processing is another such task where eye tracking has proven very useful for capturing user states in order to deliver personalization. Specifically, eye-tracking has been used to predict, in real-time, a user's performance in terms of task speed [56], as well as short-term states such as confusion [34] and interest [52]. These findings are valuable because they indicate that eye-tracking can be used to dynamically adapt the visualization to the user's needs by providing information about *when to adapt* (e.g., when the user is predicted to be slow to complete the task or is confused). In our previous work we have used eye-tracking data to predict, in real-time, some of the user's long-term cognitive abilities known to impact visualization processing, e.g., perceptual speed and verbal working memory (WM) [15,56]. These findings provide evidence about which relevant user abilities warrant adapting to (i.e., *what to adapt to*).

In this paper, we extend previous work by studying whether eye-tracking can also be used to drive adaptation in the context of visualizations embedded in narrative text, a form of multimodal document known as Magazine Style Narrative Visualization [51], or MSNV (e.g., Fig. 1). Combining text and visualizations, as it is done in MSNVs, is a widespread approach to convey trends and patterns out of complex information and can be found in newspapers, scientific papers, textbooks or blogs, among others (e.g., [37,51]). Processing MSNVs, however, can be challenging due to the need to split attention between the two information sources, a phenomenon known as the split-attention effect, which can increase cognitive load and reading time [4]. We have recently shown that these difficulties can be exacerbated depending on the user's levels of cognitive abilities such as *reading proficiency* and *verbal WM* [59]. We have also found that gaze-driven guidance in MSNVs can improve comprehension, but solely for users with *low visualization literacy* [36]. Altogether, these findings suggest that personalization to specific user cognitive abilities may be beneficial in tasks involving MSNVs.

†Authors with equal contribution

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ICMI '20, October 25–29, 2020, Virtual event, The Netherlands
© 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-7581-8/20/10...$15.00
https://doi.org/10.1145/3382507.3418884

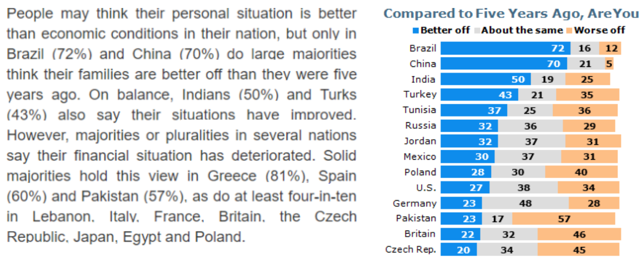

Figure 1. Sample MSNV used in this paper.

In this paper, we provide evidence for the feasibility of building eye-tracking-based user models that can be leveraged to provide such personalized support. We do so by training a set of classifiers on eye-tracking data to predict:

- Two measures of task performance, namely, *task completion time*, and *task accuracy* in terms of MNSV comprehension, both relevant to decide *when to adapt*.
- Three of the cognitive abilities identified in previous work as relevant to decide *what to adapt to* in MSNVs (*reading proficiency, visualization literacy* and *verbal WM*).

Our results extend previous work on user modeling as follows:

First, our results are the first to show the feasibility of predicting *task accuracy* for InfoVis, as previous work only predicted task completion time [56]. This holds a large practical contribution because prediction of task completion time alone is insufficient to drive personalization, as a user may be slower but more accurate. Some works have used eye-tracking to predict *reading comprehension* [16,17], but only with text. We are, thus, the first to extend these findings to multimodal documents.

Second, although previous work has already shown the feasibility of predicting *verbal WM* and *task completion time* from eye-tracking [15,56], this was done only for stand-alone visualizations. We replicate these results in a context that is substantially different due to the combination of two modalities (text and visualization), thus providing evidence on the generality of these findings. This as an important contribution because generalization often remains unstudied in applied user modeling.

Third, our results show for the first time the feasibility of predicting *reading proficiency* and *visualization literacy*. The prediction of reading proficiency is especially compelling, as this cognitive ability is different in nature from those predicted before. Namely, reading ability relates to vocabulary size and comprehension, whereas previous work focused on perceptual and memory abilities [15,56]. Thus, the fact that eye-tracking can predict this new ability extends our understanding of the role that eye-tracking data can play in modeling long-term user traits. Previous work has predicted reading abilities while reading text [64], but not reading proficiency nor in multimodal documents.

2 Related Work

2.1 Eye-Tracking for User Modeling

Eye-tracking has been used for gaze-driven adaptation, i.e., adaptation which reacts to specific user gaze patterns, in several domains (see [38] for an overview). Examples include gaze-based prompts to refocus student attention in educational software when they look away from the screen [21], adaptive online advertisements based on what information users look at in e-commerce webpages [2], and adaptive maps that dynamically place the legend next to where the user was looking [25]. Furthermore, research has shown that eye-tracking can reveal more about the users than where they look. Specifically, eye-tracking has been used for real-time user modeling tasks, by predicting a user's *task performance*, *short-term states* and *long-term characteristics* that can drive the delivery of adaptive support.

Predicting task performance. Eye-tracking has been used to predict different type of task performance, which provides valuable information about *when* to adapt, i.e., when the user is predicted with low performance. Specifically, previous work successfully used eye-tracking to predict a user's task speed in visual search tasks [11] and during processing of bar and radar charts [56]. Predicting task speed, however, is not always sufficient to drive adaptation, as there is often a trade-off between task time and accuracy, e.g., a user may be slower but more accurate. Other work did predict task comprehension in real time using eye-tracking, but only when reading text with no visualization [16,17], or requiring user to read aloud [18]. There also exists research on predicting other domain-specific measures of task performance from gaze, such as learning gains with educational software [9,31] and problem-solving outcomes in puzzle games [22]. We contribute to these works by showing that, in addition to task speed, we can also predict task comprehension from eye-tracking data in multimodal documents that combine text and visualizations, an important finding for our goal of driving adaptation in such documents.

Predicting user states and characteristics. In addition to task performance, previous work has used eye-tracking to predict a variety of short-term cognitive and affective states. For instance, eye-tracking has been used to predict emotions [28,63] and mind wandering [20,43] in educational settings and video watching activities, intention in games [53], interest when browsing webpages [1], and cognitive load in route planning and document editing tasks [26]. In InfoVis, eye-tracking has been used to predict user confusion [34], interest in the visualized data [52], and levels of familiarity with the visualization [35].

Eye-tracking has also been used to predict long-term cognitive abilities and personality traits relevant to adaptation. Specifically, previous work has leveraged eye-tracking to predict cognitive abilities known to influence visualization processing, such as perceptual speed, verbal WM, visual WM, visual scanning, and cognitive style [15,24,49,55,57]. These predictions were obtained during processing of simple bar, line and radar charts [24,55,57] as well as with a decision-making interface that features two visualizations (map and deviation chart) [15]. In text reading tasks, eye-tracking has also been used to predict user reading ability [64], which in turn impacts reading comprehension [13]. Other works have predicted personality traits in educational settings [47] and during video watching [8]. Song et al. [54] have used eye-tracking to infer a user's device preferences and then tailor shopping recommendations accordingly. We extend these works by predicting from eye-tracking data two long-terms abilities which have been shown to influence MSNV processing [59] but have never before been considered for user modeling, namely visualization literacy (i.e., the cognitive ability to use common data visualizations in an efficient and confident manner, [10]) and reading proficiency (i.e., the vocabulary size and reading comprehension ability [41]).

2.2 Eye-tracking in Multimodal Documents

The above works on eye-tracking for user modeling and adaptation, have, so far, largely ignored the context of multimodal documents, with the notable exception of instructional material that combines text and pictures (but not visualizations) [28,33,48]. In particular, these works have leveraged eye-tracking to predict learning outcome [48], emotions [28] and motivational goals [33]. Other works have used eye-tracking for off-line evaluation of how users process webpages with text and pictures [39,40], with no attempt, so far, to use these data for on-line user modeling.

While multimodal documents with visualizations, such as narrative visualizations, are very common in real-world sources (as stated in the introduction), only our previous work [36] has studied gaze-driven adaptation in this context, by providing highlighting cues based on the user's reading patterns as captured by an eye-tracker. We found in [36] that these cues can improve user comprehension of the MSNVs, but only in users with lower levels of visualization literacy, whereas users with higher levels of literacy experienced a drop in comprehension. This work also did not attempt to predict visualization literacy from the user's data.

In our recent work we further studied the impact of cognitive abilities on task performance and gaze behaviors when reading non-adaptive MSNVs [59]. We found that verbal WM (i.e., the amount of verbal information that can be temporarily maintained in WM [60]) affects overall reading speed and processing of the text, with low verbal WM users taking significantly more time and looking significantly longer at the MSNV text than their counterparts. We also found that reading proficiency affects task speed and processing of the key components of the MSNV visualization, with low reading proficiency users being significantly slower due to more transitions to and from the labels and datapoints in the visualization. Altogether, these findings show that users may benefit from adaptive support tailored to their levels of cognitive abilities and expected performance when reading MSNVs. In this paper, we focus on building an eye-tracking-based user model that can drive such real-time personalization.

3 Dataset

Below we provide a brief summary of the user study that generated the data used in this paper; for a more extensive description refer to [59]. 56 subjects (32 female, age ranging from 19 to 69) participated in the study, which started with a Tobii T-120 eye-tracker calibration, and baseline pupil size collection. The study was administered in a windowless room with uniform lighting to avoid tracking issues. Subjects were then given the task of reading 15 MSNVs in a web browser (order fully randomized), with no time constraints, and their gaze was tracked the whole time. Each MSNV consisted of a self-contained excerpt of a real-world document and one accompanying bar chart, one of the most used visualizations in real-world sources [45], see Fig. 1. After reading each MSNV, participants were asked to answer three comprehension questions on what they read. Next, each participant took a battery of well-established psychological tests to measure some of their cognitive abilities, including the Bar Chart Test for visualization literacy [10], the OSPAN Test for verbal WM [60], and the X_Lex Test for reading proficiency[1] [41]. These tests measure long-term abilities that do not vary over the course of the study.

4 Classification Experiments

We leverage the eye-tracking data collected while users processed the MSNVs to build classifiers that can predict binary labels of the following five measures (summary statistics in Table 1):

- **Two task performance measures**, namely task completion time in seconds (*TaskSpeed*), and task accuracy (prop. correct) of the comprehension questions (*TaskComp*).
- **Three cognitive abilities:** The test score for reading proficiency (*ReadP*), visualization literacy (*VisLit*), and verbal WM (*VerWM*), i.e., the 3 abilities known to impact MSNV processing and thus warrant adapting to (cf. section 2.2).

The binary labels are generated by dividing participants into High and Low groups for each measure, based on a median split (scores falling on the median are assigned to the smaller group). We target binary labels instead of the raw test/performance scores as it is a common practice in related work (e.g., [46,52,56,62]) given that an adaptive system generally needs to decide who receives and who does not receive adaptation, which makes binary classification best suited for this application.

Table 1. Summary statistics of the target measures.

Measure	Range	Mean	Median	SD	Min	Max
TaskSpeed	-	56	48	32	8	296
TaslkComp	0;1	.69	.67	.3	0	1
ReadP	0;100	84	85	10	55	96
VisLit	-2;2	.3	.47	.7	-2	1
VerWM	0;6	5	5	1.1	2	6

[1] Standard reading proficiency tests such as the IELTS and TOEFL require 1h+ to administer, which was not feasible in the scope of our user study. Instead, we use the X_Lex test which measures English vocabulary size, and can be done in under 5 minutes. This test has been shown to reliably approximate reading proficiency [42].

4.1 Eye-Tracking Features for Prediction

We use the EMDAT Python library (*github.com/ATUAV/EMDAT*) to process the eye-tracking data. EMDAT produces a battery of eye-tracking metrics specified over the entire display, and over specific Areas of Interest (AOIs). We defined 7 AOIs shown in Fig. 2, which encode the main elements in the text side of the MSNVs (reference sentences, and remaining text), as well as in the visualization side of the MSNVs (relevant bars, non-relevant bars, legends, labels, all of the vis). These AOIs are meant to capture the specific gaze processing generated by each modality, as well as gaze transitions between modalities and key elements within each modality. The resulting features are listed in Table 2 and include:

- **Gaze features** (Table 2.a) computed over the user's fixations (gaze maintained at one point on the screen) and saccades (quick eye movement between two fixations).
- **Pupil and head distance features** (Table 2.b), computed over the user's pupil size (baseline-adjusted; Iqbal et al. 2005) and distance from both eyes to the screen.
- **Areas of interest (AOI) features** (Table 2.c), i.e., gaze, pupil and head distance features computed over each of the salient regions of an MSNV as well as transitions between AOIs.

These features are mostly comparable to the ones used in related user modeling work in InfoVis [15,56] and HCI [11,17,20, 29]. We also consider features related to the pupil dilation and head distance to the screen when looking at a specific AOIs (last row in Table 2), which, to the best of our knowledge, have not been used for gaze-based user modeling.

Table 2. Summary of eye-tracking features.

a) Overall Gaze Features (21)
- Fixation rate & Fixation duration (mean, SD, Max) - Saccade Duration, Distance, & Velocity (mean, SD, max) - Absolute and Relative saccade angles (mean, SD, rate)
b) Pupil width and Head distance Features (16)
- Pupil width / head distance (mean, SD, min, max) - Pupil width / head distance at first and last recorded fixation - Pupil dilation velocity (mean, SD, min, max)
c) AOI Features (34 x 7 AOI = 238)
- Fixation Rate, Fixation duration in AOI (mean, SD, max) - Time to first fixation in AOI - Proportion of time, Proportion of fixations in AOI - Number, Prop. of transitions from this AOI to every AOI - Pupil width, pupil change velocity, and head distance when looking at AOI (mean, SD, min, max)

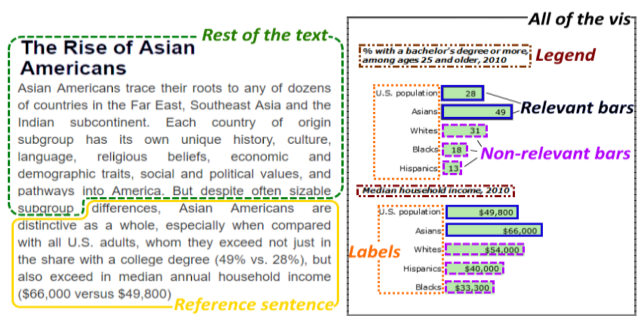

Figure 2. Set of AOIs defined over a sample MSNV.

4.2 Machine Learning Setup

***Within-task* prediction.** The dataset consists of 840 tasks (15 MSNVs x 56 subjects, see Section 3). Following [15,56], we investigate at which point *within a task* we can best predict our target users' cognitive abilities and task performance, to gauge how early adaptive interventions could be triggered based on these predictions. We consider each task as an independent datapoint, thus the dataset for within-task prediction consists of the 840 individual MNSV tasks. We evaluate prediction accuracy after each second of interaction within a given task, by feeding to the classifiers the corresponding eye-tracking data, up to 29 second. We stop at 29 seconds because this represents half of the average task time in our dataset and, beyond this point, making predictions to trigger adaptive support would not be very timely in an online scenario. For within-task predictions, the classifiers are trained on data from all the individual tasks in the training set, regardless of the order in which they were performed. To make predictions at each time t (from 1 to 29 secs) for task n, a classifier is trained on feature vectors built from user gaze data from the beginning up to time t, for all the other tasks in the training set, and prediction for task n is made by feeding to the classifier the feature vector built from the corresponding gaze data from 0 to t.

Across-tasks prediction. Predictions within task are relevant to predict task performance for that task, as well as cognitive abilities in a situation where a user is just performing that one task. However, cognitive abilities usually do not change in the short term. In our case, a given user has the same level of reading ability, vis literacy and verbal WM across all 15 tasks. Thus, we also explore how accuracy in predicting these user characteristics evolves overtime *across-tasks*, namely as users work through a sequence of 15 tasks, as they did in the original user study. For this across-tasks prediction, the dataset consists of 56 task sequences, and we evaluate prediction accuracy at the end of each task within a sequence. For making across-tasks predictions at the end of the n^{th} task in sequence s, the feature vector is built from user gaze data during this task and all its preceding tasks in the sequence. The classifiers are trained on eye-tracking data from the task sequences in the training set, looking at data from the beginning of the first task to the end of n^{th} task in each sequence.

For both within-task and across-tasks classifications, we test four different classification algorithms available in the Caret package [32] in R: Logistic Regression (LR), Random Forest (RF), Support Vector Machines with linear kernel (SVM), and eXtreme Gradient Boosting (XGB). We selected LR and RF because, in previous work, they produced good results for predicting various user states during visualization processing, e.g., [15,56]. We use SVM as this approach is commonly used for eye-tracking-based user modeling, e.g., [19,27,65]. Lastly, we look at XGB due to the strong performances obtained by this recent algorithm in a variety of prediction tasks, performing comparably well to deep learning approaches [23,44]. We chose not to experiment with deep learning methods because of the small size of our dataset (840 datapoints for within-task predictions, 56 for across-tasks).

To train the classifiers for each time-step (29 time-steps within-task, 15 time-steps across-tasks), we perform 10-fold nested cross-validation over users so that users in the test fold never appear in the training fold. Predictions for users in the test fold at a given time step is solely made based on classifiers trained on users in the train fold at the exact same time step so as to use eye-tracking features built over the same amount of interaction. Nested cross-validation is used for hyper-parameter tuning as well as for Correlation Feature Selection [32] to remove highly correlated features. The cross-validation process is repeated 10 times (runs) for reproducibility purposes. We report classification performance in terms of accuracy averaged over the 10 folds and the 10 runs and use a majority-class classifier as a baseline.

5 Results

We first report results on predicting task performance measures *within task* (Section 5.1), followed by results on predicting cognitive abilities *within* and *across* tasks (Sections 5.2 and 5.3). Then, we discuss the implications of the results for adaptation by analyzing class accuracies and feature importance (Section 5.4).

5.1 Predicting Measures of Task Performance

We evaluate the performance of the classifiers at predicting task accuracy (TaskComp) and task speed (TaskSpeed) within tasks by running, for each measure, a one-way repeated-measures ANOVA with accuracy as the dependent variable, classifier type (5 with the baseline) as the factor, and time-step (29) as the repeated measure. To account for the multiple ANOVAs we run (one per prediction target), we adjust all the obtained p-values using the Benjamini & Hochberg [7] method to control for the false discovery rate (FDR).

For both measures, we found a significant main effect of classifier on accuracy (*statistical significance* is reported at $p < .05$ after adjustment). We investigate these main effects by running post-hoc pairwise comparisons (t-tests adjusted for FDR), with results shown in the 2nd column of Table 3. Classifiers are ordered from left to right by prediction accuracy; ">" indicates statistically significant difference, while "=" indicates statistical equivalence.

Overall, most of the classifiers outperform the baseline for TaskComp and TaskSpeed, indicating the feasibility of predicting user performance from eye-tracking data in MSNVs. Both LR and RF generate strong performance, which aligns with previous results on predicting cognitive abilities from gaze with bar and radar charts [15,56]. Next, we examine how early these classifiers can predict task performance. Classification accuracy at each evaluated time-step is shown in Figure 3 for TaskComp, and Figure 4 for TaskSpeed, for the best classifiers only. We also report the accuracy at the best time-step in the third column of Table 3.

Table 3. Comparisons of the classifiers, as well as accuracy at the best time-step for the best classifier (RF).

Perform.	Ranking of classifiers	Acc. at Best Step		
		Step	Acc.	Base
TASKCOMP	**RF** > XGB = LR > Base > SVM	23 sec	.67	.58
TASKSPEED	**RF = LR** > XGB > SVM > Base	29 sec	.73	.60

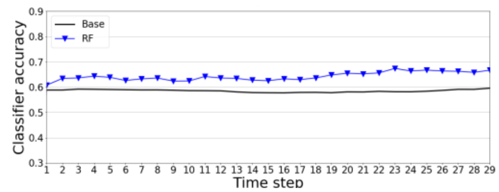

Figure 3. Results for task comprehension.

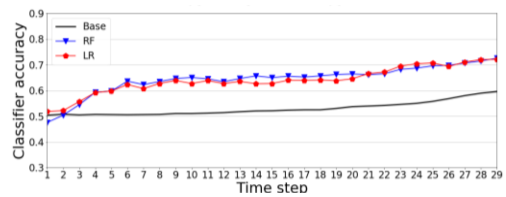

Figure 4. Results for task speed.

TaskComp: Figure 3 shows that RF achieves a peak accuracy of .67 at 23 sec, which represents a 15.5% improvement with respect to the baseline (.58) and seems to stabilize past this point. This indicates that observing 23 sec (i.e., within the first half of the interaction) is sufficient to identify if a user is understanding the content of the MSNVs, which leaves substantial time in the task to provide adaptive support. As discussed in the related work, there is no other work that predicted task comprehension in visualization tasks, nor with multimodal documents. There has been work on predicting text reading comprehension [16,17], though without improvements over a majority-class baseline due to a strong class imbalance. The fact that we do obtain significant improvement over the baseline suggests that the visualization generates additional useful gaze patterns for comprehension prediction, in addition to the text.

TaskSpeed: Figure 4 shows that prediction performance increases consistently as we accumulate more data *within a task*, such that the best accuracy (.73) is obtained at the last time-step (29 sec). However, the baseline is also increasing over time (at 29 secs the baseline is .6), due to the fact that users finish some of their tasks in under 29 seconds, especially in simpler MSNVs. To account for this increasing baseline, we also examined the improvements in accuracy obtained by RF and LR over the baseline, and found that the highest improvement is obtained 10 seconds in the task by RF (.65 accuracy against a .51 baseline, representing a 27.4% increase). These results show that even early in the task, RF substantially outperforms the baseline. Noteworthy, this accuracy at 10 seconds is similar to the ones obtained in Steichen et al. [56] (.65 over a .51 baseline) on predicting task speed, albeit during simple analytic tasks with bar and radar charts, while our peak accuracy of .73 is substantially higher.

Note that there is often a trade-off between task speed and task accuracy as users may sacrifice speed for accuracy and vice versa. Thus, predicting both measures allows us to identify those users who may be both slow and inaccurate, which are the ones that may benefit the most from adaptive support. We will further examine this tradeoff in Section 5.4.1.

5.2 Predicting Cognitive Abilities Within Task

To examine the feasibility of predicting user cognitive abilities within a task, we perform, for each cognitive ability, the same statistical procedure as in the previous section (5.1).

We found a significant main effect of classifier on accuracy for each cognitive ability. Post-hoc pairwise comparisons are reported in Table 4, which shows that no classifier beats the baseline at predicting VerWM. However, LR significantly outperforms all other classifiers, including the baseline, at predicting VisLit and ReadP. These results show, for the first time, the feasibility of predicting these two cognitive abilities from eye-tracking data. The fact that LR is the best classifier is consistent with previous results on predicting other cognitive abilities in InfoVis using a comparable dataset size and feature set [56]. Focusing on LR as the best classifier, we examine at which time-step we achieve highest accuracy for VisLit and ReadP, to ascertain how early, during a task, these abilities can be predicted. Figure 5 shows classification accuracy over the evaluated time-steps for ReadP, and Figure 6 for VisLit, Table 4 (3rd column) reports the LR peak accuracy and the corresponding time-step.

ReadP: Figure 5 indicates that classification accuracy improves steadily for most of the first 29 seconds of the task, reaching .60 at 4 sec, and eventually peaking at (.67) at 26 sec. This finding means that the best prediction can still be made in the first half of the task on average, thus leaving substantial time for adaptation. The fact that accuracy increases in the first 26 seconds may be due to the fact that many users start the tasks by reading the text, and as they keep reading, they generate more gaze data over the text making it more likely to capture ReadP.

VisLit: Figure 6 shows that peak accuracy (.58) is reached at the first second, and that classification accuracy drops past 2 seconds. In this case, the fact that the very first second of data yields the best accuracy is consistent with previous findings on the prediction of perceptual and memory abilities necessary to process visualizations [15,56]. This might be due to the fact that these abilities influence how users discover a new visualization layout at hand, i.e., they might be faster in locating where to start, thus generating useful early patterns for prediction, although further analysis is needed to ascertain this. The fact that for ReadP the best prediction occurs past few seconds of interaction, but for VisLit it is achieved right at the start, indicates that abilities related to reading and comprehension such as ReadP require more accumulated data, while abilities involving perception and memory such as VisLit and the ones in [15,56] require less accumulated data.

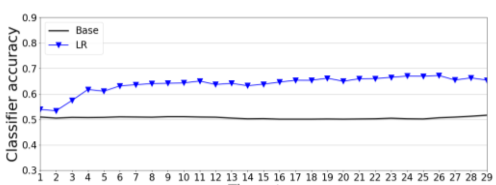

Figure 5. Results for ReadP within task.

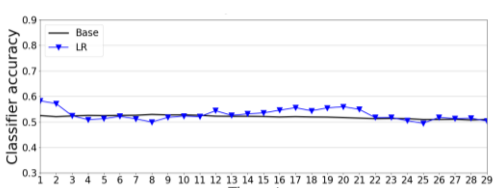

Figure 6. Results for VisLit within task

Table 4. Comparisons of the classifiers within task as well as accuracy at the best time-step for the best classifier.

Cog. Ability	Ranking of classifiers	Acc. at Best Step		
		Step	Acc.	Base
ReadP	**LR** > SVM > RF = XGB > Base	26 sec	.67	.51
VisLit	**LR** > Base = XGB > RF = SVM	1 sec	.58	.52
VerWM	**Base** > SVM = LR > XGB = RF	-	-	-

5.3 Predicting Cognitive Abilities Across Tasks

As in the previous section, we compare the classifiers across-tasks by running, for each cognitive ability, a one-way repeated-measures ANOVA with accuracy as the dependent variable, classifier (5) as the factor, and time-step (15) as the repeated measure. We found a significant main effect of classifier for each cognitive ability. Post-hoc pairwise comparisons (see Table 5) reveal that at least one classifier outperforms the baseline for each cognitive ability. These findings show that it is feasible to predict all of our target abilities by accumulating data across tasks.

ReadP: As shown in Table 5 and Figure 7, peak accuracy (.67) is reached after observing 4 tasks, which is similar to the best within-task accuracy (.67) reported in the previous section. This suggests that this ability can be predicted regardless of the amount of interaction with the visualizations which is promising for the generalization of the prediction of this ability across MSNV tasks.

VisLit: Peak accuracy (.66) is obtained by RF at task 3, as shown (see Table 5 and Figure 8). This peak accuracy is substantially higher as compared to the best *within task* accuracy (.58), with a 14% increase. This suggests that VisLit can be better predicted when more instances of MSNV processing are available, likely because more MSNV tasks allow to capture more of the sparse relevant gaze behaviors otherwise not sufficient when considering only a single task as we will highlight in Section 5.4.2.

VerWM: Figure 9 shows that the XGB classifier yields its best accuracy (.72) after 10 tasks have been completed, indicating that accumulating data over multiple tasks is useful for VerWM. Figure 9 also shows that XGB reaches .67 accuracy at the second task, suggesting that observing only 2 tasks may be sufficient to make a useful prediction for adaptation, depending on the adaptation mechanism. Overall, the results for VerWM indicate

Table 5. Comparisons of the classifiers across tasks.

Cog. Ability	Ranking of classifiers	Acc. at Best Step		
		Step	Acc.	Base
ReadP	**RF** > SVM > XGB > LR > Base	Task 4	.67	.51
VisLit	**RF = SVM** > XGB = Base > LR	Task 3	.66	.53
VerWM	**XGB** > RF > LR > SVM > Base	Task 10	.72	.51

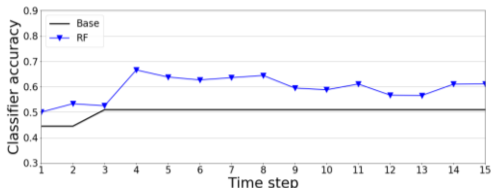

Figure 7. Results for ReadP across tasks

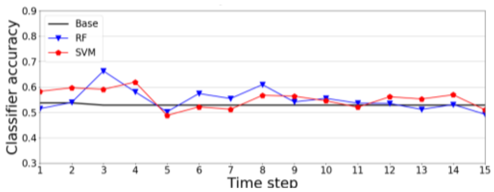

Figure 8. Results for VisLit across tasks

that it is needed to accumulate data over at least two tasks to make any good prediction as no classifier outperformed the baseline in the within task. This contradicts findings from Steichen et al. [56], who were able to predict VerWM within task involving stand-alone bar and radar charts. A possible explanation for this is that the varied verbal information in MSNVs (text, title, labels) might generate more challenging gaze patterns for predicting VerWM than the simple stand-alone charts studied in [56].

While LR was found to be the best classifier within task, it performs poorly across-tasks, as shown in Table 5. On the other hand, tree-based classifiers (RF and XGB) perform well here, possibly because accumulating several minutes of gaze data captures more varied behaviors that are better explored by tree-based ensemble models, albeit further analysis is neeed to ascertain this point and the generalizability of this finding.

5.4 Further Insights into the Best Classifiers

We scrutinize here the best classifiers by discussing their class accuracies (Section 5.4.1) and top features (Section 5.4.2).

5.4.1 Class accuracies. Table 6 shows the class accuracies of the best performing classifiers, i.e., the ability to correctly identify the "Low" (Acc Low) and "High" (Acc High) users for each predicted variable. Overall class accuracies for the "Low" groups are substantially higher than for the "High" groups, for all target variables. This finding is promising because users with low levels of cognitive abilities and task performance are the ones who need adaptive support the most. Our results show that our classifiers can correctly identify a large majority of those users who need

Table 6. Class accuracy of the best classifiers and step.

Target	Clas-sifier	Time-step	Majority class	Acc. Low	Acc. High
VisLit	RF	3 tasks	low (.53)	.75	.57
ReadP	LR	26 secs	low (.51)	.69	.65
VerWM	XGB	10 tasks	high (.51)	.74	.69
TaskComp	RF	24 secs	low (.58)	.84	.42
TaskSpeed	RF	10 secs	high (.51)	.72	.58

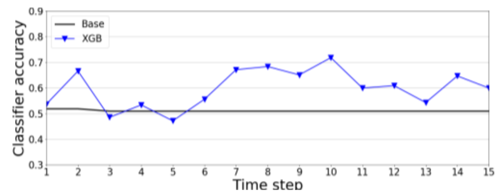

Figure 9. Results for VerWM across tasks

help with accuracy from .69 (low ReadP) to .84 (low TaskComp). While the class accuracies for the "High" groups are generally lower than the "Low" groups (see Table 6, column "High Acc"), the classifiers can still correctly identify a substantial proportion of the "High" users for VerWM (.69 class accuracy) and ReadP (.65). However, class accuracies are lower for users with "High" VisLit, TaskComp and TaskSpeed, such that a large portion of these users would be misclassified as belonging to the "Low" groups. This may indicate that users with "Low" VisLit, "Low" TaskSpeed, and "Low" TaskComp exhibit consistent gaze behaviors, whereas their "High" counterparts may have been leveraging different, more individualized strategies to comprehend the MSNVs. While there is room for improvement in the "High" group, in terms of providing adaptations, the benefits of helping the "Low" groups might outweigh the possible distraction caused by unwarranted adaptation being delivered to the misclassified "High" users, depending on the nature and intrusiveness of the adaptations.

As mentioned in Section 5.1, it would be most valuable to accurately identify participants that are both slow and inaccurate, which are mostl likely to benefit from adaptations. These results are encouraging as our classifiers lead to an average accuracy of .78 for the "Low" users in the two task performance measures.

5.4.2 Feature importance. Table 7 reports the top 3 features of the best performing classifier and their directionality for each target variable, as generated by the Caret package [32]. Higher mean values for each feature in the "High" or "Low" groups are indicated as (>) and (<) respectively in Table 7.

TaskSpeed: As shown in Table 7, TaskSpeed is better predicted by fixations and saccade information (both overall and in terms of fixations over the text), possibly because these features can reveal the user's overall levels of activity. In fact, the top features indicate that users with shorter completion times perform faster saccades and shorter fixations, which is related to how quickly the user reads and process information in the MSNV.

TaskComp: The top 2 features for TaskComp in Table 7 pertain to AOIs over the text, possibly because the way users read the text can indicate whether they comprehend the story. Indeed, the top two features indicate that higher activity in the text and longer time to reach the references are associated with higher comprehension. The third feature relates to the transitions from the labels AOI to the rest of the visualization, which indicates that the labels are key elements of the visualizations to comprehend the datapoints and map them to the text.

ReadP: We focus only on the within-task classifiers as accumulating tasks does not help for ReadP. Table 7 shows that the top features within-task and across-tasks are different.

Table 7. Top features selected by the best classifiers.

Target	Top 3 Features
TaskSpeed (Within)	1. Fixation to saccade time ratio (<) 2. Mean saccade speed (>) 3. Text AOI: st. dev. fixation duration (<)
TaskComp (Within)	1. Text AOI: #Transitions within self (>) 2. Refs AOI: Time to first fixation (>) 3. Vis AOI: Prop. Trans. from Labels (>)
ReadP (Within)	1. Mean saccade duration (>) 2. Mean fixation duration (<) 3. Reference AOI: fixation rate (>)
VisLit (Across)	1. Non-relevant bars AOI: st. dev. head distance (>) 2. End pupil size (<) 3. Relevant bars AOI: min head distance (<)
VerWM (Across)	1. Vis AOI: Prop. Trans. from Non Relevant bars (>) 2. Legend AOI: fixation rate (>) 3. Legend AOI: max fixation duration (>)

Specifically, the top features within-task include increased *fixation rate* over the reference AOI (i.e., the relevant sentences in the text) for users in the "High" group, which is not surprising as ReadP relates to reading behaviors. The other two top features within-task indicate longer saccades and shorter fixations for users in the "High" group, which are related to users' reading speed [50,61], an important component of ReadP.

VisLit: The first and third most important features for VisLit in Table 7 relate to head distance when processing the bar AOIs. This can explain the higher accuracy obtained across-tasks for VisLit compared to within-task, as the bars generate fewer fixations than the other AOIs due to their small size, so accumulating several tasks is required to capture relevant behaviors when looking at them. The last feature, *end pupil size*, may be capturing differences in the users' cognitive load, depending on their levels of VisLit. Specifically, users in the "Low" group show higher pupil sizes at the end of the task, which has been linked to higher cognitive load [5]. Overall, the fact that all top predictors for VisLit relate to the pupil and head distance may have interesting practical implications as pupil size and head distance to the screen can be tracked with ubiquitous HD webcams, albeit future work should ascertain this point [14].

VerWM: Table 7 shows that all top features for VerWM pertain to AOIs over the visualization. Two of these features pertain to the Legend AOI, showing that processing of the verbal information in the legend can reveal the user's VerWM, specifically in terms of higher fixation rate and longer fixations for users in the "High" group. This is in line with previous work showing that VerWM impacts visual processing of the legend [58].

6 Discussion

We discuss the implications of our results both for user modelling and personalization in Infovis, along with avenues for future work.

Implications for user modelling. Our work provides a first successful attempt at performing user modeling from eye-tracking data during the processing of narrative visualizations by predicting two task performance measures and three cognitive abilities. In particular, we extend this previous work by showing the feasibility to predict a new measure of task performance (TaskComp), which is important to provide timely adaptations to improve users' performance. Our results indicate the need to collect eye-tracking data for ~20s to reach the best performance, which leaves substantial time to provide adaptive support even in our rather short MSNV excerpts. Such adaptation would be even more valuable with longer MSNVs. Furthermore, two of the 3 most important features to predict *TaskComp* are related to information gathered while users read the text. Thus, in longer MSNVs, users may spend the first ~20s reading the text side of the MSNV which would allow providing adaptive interventions on the visualization side in a timely manner, without being disruptive.

We showed the feasibility to predict three important cognitive abilities (VerWM, ReadP and VisLit) known to impact MSNV processing. We found that peak predictions are reached after observing a few seconds of data for ReadP, a few tasks for VisLit, and several tasks for VerWM. This exemplifies the lack of one-size-fits-all approach in terms of the amount of eye tracking data required to predict cognitive abilities.

We showed that our user models are especially accurate at identifying users with low levels of cognitive abilities and task performance, i.e., those users who may benefit from adaptive support the most. This is encouraging, given that our results should be interpreted as lower bounds in terms of classification performance from eye-tracking data. In future work we will focus on improving the accuracy of our user models, in particular, by leveraging other machine learning algorithms and more elaborated feature selection mechanisms. We will also consider identifying more generalizable, absolute thresholds to define the low and high groups, than the median split used in this paper.

Implications for personalization. Our user models will make it possible to evaluate the value of personalized narrative visualizations, which we plan to do as follows.

In our previous work [36] we found that gaze-driven highlighting cues, which were solely triggered by reading behaviors, improved comprehension of users with low VisLit only but did not help users with medium to high levels of VisLit who found them distractive. Thus, our VisLit predictor could be used to provide these cues only to low VisLit users, avoiding unnecessary interventions for medium/high VisLit users.

Previous work found that low ReadP users have difficulties processing chart legends in MSNVs, e.g., they spend more time to process the legend and transition more often to it [59], thus based on our classifiers, these users could receive adaptations designed to reduce these difficulties, e.g., by displaying a simpler legend with less items, or by highlighting the relevant information in the legend that are described in the MSNV text.

In addition, not all users with a high VisLit or a low ReadP encounter the aforementioned difficulties. Our predictors for task performance can be leveraged to trigger the adaptations discussed above specifically when the user is predicted to be inaccurate and slow so as to help these users solely when they need it.

ACKNOWLEDGMENTS

This research was funded by the Natural Sciences and Engineering Research Council of Canada, NSERC grants RGPIN-2016-04611 and RTI-2019-00711.

REFERENCES

[1] Stephen Akuma, Chrisina Jayne, Rahat Iqbal, and Faiyaz Doctor. 2015. Inferring Users' Interest on Web Documents Through Their Implicit Behaviour. In *Proceedings of the International Conference on Engineering Applications of Neural Networks*. Springer, Rhodes, Greece, 315–324. DOI:https://doi.org/10.1007/978-3-319-23983-5_29

[2] Florian Alt, Alireza Sahami Shirazi, Albrecht Schmidt, and Julian Mennenöh. 2012. Increasing the user's attention on the web: using implicit interaction based on gaze behavior to tailor content. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. ACM, 544–553.

[3] Tobias Appel, Natalia Sevcenko, Franz Wortha, Katerina Tsarava, Korbinian Moeller, Manuel Ninaus, Enkelejda Kasneci, and Peter Gerjets. 2019. Predicting Cognitive Load in an Emergency Simulation Based on Behavioral and Physiological Measures. In *2019 International Conference on Multimodal Interaction* (ICMI '19). Association for Computing Machinery, Suzhou, China, 154–163. DOI:https://doi.org/10.1145/3340555.3353735

[4] Paul Ayres and Gabriele Cierniak. 2012. Split-attention effect. In *Encyclopedia of the Sciences of Learning*. Springer, 3172–3175.

[5] Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* 91, 2 (1982), 276–292. DOI:https://doi.org/10.1037/0033-2909.91.2.276

[6] Kenan Bektas, Arzu Cöltekin, Jens Krüger, and Andrew T. Duchowski. 2015. A testbed combining visual perception models for geographic gaze contingent displays. In *Eurographics Conference on Visualization (EuroVis)-Short Papers*. .

[7] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 1 (1995), 289–300.

[8] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. 2019. Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). Association for Computing Machinery, Glasgow, Scotland Uk, 1–12. DOI:https://doi.org/10.1145/3290605.3300451

[9] Daria Bondareva, Cristina Conati, Reza Feyzi-Behnagh, Jason M. Harley, Roger Azevedo, and François Bouchet. 2013. Inferring Learning from Gaze Data during Interaction with an Environment to Support Self-Regulated Learning. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education*. Springer, Memphis, TN, USA, 229–238.

[10] J. Boy, R.A. Rensink, E. Bertini, and J.-D. Fekete. 2014. A Principled Way of Assessing Visualization Literacy. *IEEE Trans. Vis. Comput. Graph.* 20, 12 (December 2014), 1963–1972. DOI:https://doi.org/10.1109/TVCG.2014.2346984

[11] Eli T. Brown, Alvitta Ottley, Hang Zhao, Quan Lin, Richard Souvenir, Alex Endert, and Ronald Chang. 2014. Finding waldo: Learning about users from their interactions. *IEEE Trans. Vis. Comput. Graph.* 20, 12 (2014), 1663–1672. DOI:https://doi.org/10.1109/tvcg.2014.2346575

[12] Georg Buscher, Andreas Dengel, and Ludger van Elst. 2008. Eye movements as implicit relevance feedback. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '08). Association for Computing Machinery, Florence, Italy, 2991–2996. DOI:https://doi.org/10.1145/1358628.1358796

[13] Thomas H. Carr. 1981. Building theories of reading ability: On the relation between individual differences in cognitive skills and reading comprehension. *Cognition* 9, 1 (January 1981), 73–114. DOI:https://doi.org/10.1016/0010-0277(81)90015-9

[14] Siyuan Chen and Julien Epps. 2014. Efficient and Robust Pupil Size and Blink Estimation From Near-Field Video Sequences for Human–Machine Interaction. *IEEE Trans. Cybern.* 44, 12 (December 2014), 2356–2367. DOI:https://doi.org/10.1109/TCYB.2014.2306916

[15] Cristina Conati, Sébastien Lallé, Md. Abed Rahman, and Dereck Toker. 2017. Further Results on Predicting Cognitive Abilities for Adaptive Visualizations. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, Melbourne, Australia, 1568–1574. DOI:https://doi.org/10.24963/ijcai.2017/217

[16] Leana Copeland, Tom Gedeon, and Sabrina Caldwell. 2015. Effects of text difficulty and readers on predicting reading comprehension from eye movements. In *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. 407–412. DOI:https://doi.org/10.1109/CogInfoCom.2015.7390628

[17] Leana Copeland, Tom Gedeon, and B. Sumudu U. Mendis. 2014. Predicting reading comprehension scores from eye movements using artificial neural networks and fuzzy output error. *Artif Intell Res.* 3, 3 (2014), 35–48.

[18] Scott A. Crossley, Stephen Skalicky, Mihai Dascalu, Danielle S. McNamara, and Kristopher Kyle. 2017. Predicting Text Comprehension, Processing, and Familiarity in Adult Readers: New Approaches to Readability Formulas. *Discourse Process.* 54, 5–6 (July 2017), 340–359. DOI:https://doi.org/10.1080/0163853X.2017.1296264

[19] Sidney D'Mello, Jonathan Cobian, and Matthew Hunter. 2013. Automatic Gaze-Based Detection of Mind Wandering during Reading. In *Proceedings of the 6th International Conference on Educational Data Mining*. Memphis, TN, USA, 364–365.

[20] Sidney D'Mello, Catlin Mills, Robert Bixler, and Nigel Bosch. 2017. Zone out no more: Mitigating mind wandering during computerized reading. In *Proceedings of the 10th International Conference on Educational Data Mining*. Wuhan, China, 8–15.

[21] Sidney D'Mello, Andrew Olney, Claire Williams, and Patrick Hays. 2012. Gaze tutor: A gaze-reactive intelligent tutoring system. *Int. J. Hum.-Comput. Stud.* 70, 5 (2012), 377–398. DOI:https://doi.org/10.1016/j.ijhcs.2012.01.004

[22] Shahram Eivazi and Roman Bednarik. 2011. Predicting Problem-Solving Behavior and Performance Levels from Visual Attention Data. In *Proceedings of the 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction, in conjunction with IUI 2011*. ACM, Palo Alto, CA, USA, 9–16.

[23] Nir Friedman, Tomer Fekete, Kobi Gal, and Oren Shriki. 2019. EEG-Based Prediction of Cognitive Load in

Intelligence Tests. *Front. Hum. Neurosci.* 13, (2019). DOI:https://doi.org/10.3389/fnhum.2019.00191
[24] Matthew Gingerich and Cristina Conati. 2015. Constructing Models of User and Task Characteristics from Eye Gaze Data for User-Adaptive Information Highlighting. In *Proceedings of the 29th Conference on Artificial Intelligence*. AAAI Press, Austin, TX, USA, 1728–1734.
[25] Fabian Göbel, Peter Kiefer, Ioannis Giannopoulos, Andrew T. Duchowski, and Martin Raubal. 2018. Improving Map Reading with Gaze-adaptive Legends. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (ETRA '18). ACM, New York, NY, USA, 29:1–29:9. DOI:https://doi.org/10.1145/3204493.3204544
[26] Shamsi T. Iqbal, Piotr D. Adamczyk, Xianjun Sam Zheng, and Brian P. Bailey. 2005. Towards an index of opportunity: Understanding changes in mental workload during task execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Portland, OR, USA, 311–320. DOI:https://doi.org/10.1145/1054972.1055016
[27] Young-Min Jang, Rammohan Mallipeddi, and Minho Lee. 2014. Identification of human implicit visual search intention based on eye movement and pupillary analysis. *User Model. User-Adapt. Interact.* 24, 4 (October 2014), 315–344. DOI:https://doi.org/10.1007/s11257-013-9142-7
[28] Natasha Jaques, Cristina Conati, Jason M. Harley, and Roger Azevedo. 2014. Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System. In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*. Springer, Honolulu, HI, USA, 29–38. DOI:https://doi.org/10.1007/978-3-319-07221-0_4
[29] Halszka Jarodzka, Hans Gruber, and Kenneth Holmqvist. 2017. Eye tracking in Educational Science: theoretical frameworks and research agendas. (2017). DOI:http://dx.doi.org/10.16910/jemr.10.1.3
[30] Marcel Adam Just and Patricia Ann Carpenter. 1987. *The psychology of reading and language comprehension.* Allyn & Bacon.
[31] Samad Kardan and Cristina Conati. 2012. Exploring Gaze Data for Determining User Learning with an Interactive Simulation. In *Proceedings on the International Conference on User Modeling, Adaptation, and Personalization* (Lecture Notes in Computer Science). Springer Berlin Heidelberg, 126–138.
[32] M. Kuhn. 2008. Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 5 (2008), 1–26. DOI:http://dx.doi.org/10.18637/jss.v028.i05
[33] Sébastien Lallé, Cristina Conati, and Roger Azevedo. 2018. Prediction of Student Achievement Goals and Emotion Valence during Interaction with Pedagogical Agents. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*. IFAAMAS, Stockholm, Sweden, 1222–1231.
[34] Sébastien Lallé, Cristina Conati, and Giuseppe Carenini. 2016. Predicting confusion in information visualization from eye tracking and interaction data. In *Proceedings on the 25th International Joint Conference on Artificial Intelligence*. AAAI Press, New York, NY, USA, 2529–2535.
[35] Sébastien Lallé, Cristina Conati, and Giuseppe Carenini. 2016. Prediction of individual learning curves across information visualizations. *User Model. User-Adapt. Interact.* 26, 4 (2016), 307–345. DOI:https://doi.org/10.1007/s11257-016-9179-5
[36] Sébastien Lallé, Dereck Toker, and Cristina Conati. 2019. Gaze-Driven Adaptive Interventions for Magazine-Style Narrative Visualizations. *IEEE Trans. Vis. Comput. Graph.* (2019), 1–12. DOI:https://doi.org/10.1109/TVCG.2019.2958540
[37] Jason Lankow, Josh Ritchie, and Ross Crooks. 2012. *Infographics: the power of visual storytelling*. John Wiley & Sons, Inc, Hoboken, N.J.
[38] Päivi Majaranta and Andreas Bulling. 2014. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*. Springer, 39–65.
[39] Lucia Mason, Patrik Pluchino, Maria Caterina Tornatora, and Nicola Ariasi. 2013. An Eye-Tracking Study of Learning From Science Text With Concrete and Abstract Illustrations. *J. Exp. Educ.* 81, 3 (July 2013), 356–384. DOI:https://doi.org/10.1080/00220973.2012.727885
[40] Lucia Mason, Maria Caterina Tornatora, and Patrik Pluchino. 2013. Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Comput. Educ.* 60, 1 (January 2013), 95–109. DOI:https://doi.org/10.1016/j.compedu.2012.07.011
[41] Paul Meara. 2010. *EFL Vocabulary Tests* (second edition ed.). Lognostics, Swansea: Wales.
[42] Paul Meara and Glyn Jones. 1990. *Eurocentres Vocabulary Size Test 10KA.* Eurocentres Learning Service, Zurich.
[43] Caitlin Mills, Robert Bixler, Xinyi Wang, and Sidney K. D'Mello. 2016. *Automatic Gaze-Based Detection of Mind Wandering during Narrative Film Comprehension.* International Educational Data Mining Society.
[44] Alex Mott, Joshua Job, Jean-Roch Vlimant, Daniel Lidar, and Maria Spiropulu. 2017. Solving a Higgs optimization problem with quantum annealing for machine learning. *Nature* 550, 7676 (October 2017), 375–379. DOI:https://doi.org/10.1038/nature24047
[45] Tamara Munzner. 2014. *Visualization analysis and design.* CRC press.
[46] Alvitta Ottley, Huahai Yang, and Remco Chang. 2015. Personality as a Predictor of User Strategy: How Locus of Control Affects Search Strategies on Tree Visualizations. In *Proceedings of the 33rd Annual Conference on Human Factors in Computing Systems.* ACM, Seoul, Korea, 3251–3254. DOI:https://doi.org/10.1145/2702123.2702590
[47] Luc Paquette, Jonathan Rowe, Ryan Baker, Bradford Mott, James Lester, Jeanine DeFalco, Keith Brawner, Robert Sottilare, and Vasiliki Georgoulas. 2016. Sensor-Free or Sensor-Full: A Comparison of Data Modalities in Multi-Channel Affect Detection. *Int. Educ. Data Min. Soc.* (2016).
[48] Ramkumar Rajendran, Anurag Kumar, Kelly E. Carter, Daniel T. Levin, and Gautam Biswas. 2018. *Predicting Learning by Analyzing Eye-Gaze Data of Reading Behavior.* International Educational Data Mining Society.
[49] George E. Raptis, Christina Katsini, Marios Belk, Christos Fidas, George Samaras, and Nikolaos Avouris. 2017. Using Eye Gaze Data and Visual Activities to Infer Human Cognitive Styles: Method and Feasibility Studies. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization* (UMAP '17). Association for

Computing Machinery, New York, NY, USA, 164–173. DOI:https://doi.org/10.1145/3079628.3079690

[50] Keith Rayner, Timothy J. Slattery, and Nathalie N. Bélanger. 2010. Eye movements, the perceptual span, and reading speed. *Psychon. Bull. Rev.* 17, 6 (December 2010), 834–839. DOI:https://doi.org/10.3758/PBR.17.6.834

[51] E Segel and J Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Trans. Vis. Comput. Graph.* 16, 6 (2010), 1139–1148. DOI:https://doi.org/10.1109/TVCG.2010.179

[52] Nelson Silva, Tobias Schreck, Eduardo Veas, Vedran Sabol, Eva Eggeling, and Dieter W. Fellner. 2018. Leveraging eye-gaze and time-series features to predict user interests and build a recommendation model for visual analysis. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications.* ACM, Warsaw, Poland, 13:1-13:9.

[53] Ronal Singh, Tim Miller, Joshua Newn, Eduardo Velloso, Frank Vetere, and Liz Sonenberg. 2020. Combining gaze and AI planning for online human intention recognition. *Artif. Intell.* 284, (July 2020), 103275. DOI:https://doi.org/10.1016/j.artint.2020.103275

[54] Heyjin Song and Nammee Moon. 2018. A Preference Based Recommendation System Design Through Eye-Tracking and Social Behavior Analysis. In *Advances in Computer Science and Ubiquitous Computing* (Lecture Notes in Electrical Engineering). Springer, Singapore, 1014–1019. DOI:https://doi.org/10.1007/978-981-10-7605-3_162

[55] Ben Steichen, Giuseppe Carenini, and Cristina Conati. 2013. User-adaptive information visualization: using eye gaze data to infer visualization tasks and user cognitive abilities. In *Proceedings of the 2013 international conference on Intelligent user interfaces* (IUI '13). ACM, New York, NY, USA, 317–328. DOI:https://doi.org/10.1145/2449396.2449439

[56] Ben Steichen, Cristina Conati, and Giuseppe Carenini. 2014. Inferring Visualization Task Properties, User Performance, and User Cognitive Abilities from Eye Gaze Data. *ACM Trans. Interact. Intell. Syst.* 4, 2 (2014), Article 11. DOI:https://doi.org/10.1145/2633043

[57] Ben Steichen, Bo Fu, and Tho Nguyen. 2020. Inferring Cognitive Style from Eye Gaze Behavior During Information Visualization Usage. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (UMAP '20). Association for Computing Machinery, New York, NY, USA, 348–352. DOI:https://doi.org/10.1145/3340631.3394881

[58] Dereck Toker and Cristina Conati. 2014. Eye tracking to understand user differences in visualization processing with highlighting interventions. In *Proceedings of the 22nd International Conference on User Modeling, Adaptation, and Personalization.* Springer, Aalborg, Denmark, 219–230. DOI:https://doi.org/10.1007/978-3-319-08786-3_19

[59] Dereck Toker, Cristina Conati, and Giuseppe Carenini. 2019. Gaze analysis of user characteristics in magazine style narrative visualizations. *User Model. User-Adapt. Interact.* to appear, (2019).

[60] Marilyn L Turner and Randall W Engle. 1989. Is working memory capacity task dependent? *J. Mem. Lang.* 28, 2 (April 1989), 127–154. DOI:https://doi.org/10.1016/0749-596X(89)90040-5

[61] Geoffrey Underwood, Alison Hubbard, and Howard Wilkinson. 1990. Eye Fixations Predict Reading Comprehension: The Relationships between Reading Skill, Reading Speed, and Visual Inspection. *Lang. Speech* 33, 1 (January 1990), 69–81. DOI:https://doi.org/10.1177/002383099003300105

[62] M.C. Velez, D. Silver, and M. Tremaine. 2005. Understanding visualization through spatial ability differences. In *Proceedings of the IEEE Conference on Visualization.* IEEE, Minneapolis, MN, USA, 511–518. DOI:https://doi.org/10.1109/VISUAL.2005.1532836

[63] Suowei Wu, Zhengyin Du, Weixin Li, Di Huang, and Yunhong Wang. 2019. Continuous Emotion Recognition in Videos by Fusing Facial Expression, Head Pose and Eye Gaze. In *2019 International Conference on Multimodal Interaction* (ICMI '19). Association for Computing Machinery, Suzhou, China, 40–48. DOI:https://doi.org/10.1145/3340555.3353739

[64] Zehui Zhan, Lei Zhang, Hu Mei, and Patrick S. W. Fong. 2016. Online Learners' Reading Ability Detection Based on Eye-Tracking Sensors. *Sensors* 16, 9 (September 2016), 1457. DOI:https://doi.org/10.3390/s16091457

[65] W. Zheng, B. Dong, and B. Lu. 2014. Multimodal emotion recognition using EEG and eye tracking data. In *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, Chicago, IL, USA, 5040–5043. DOI:https://doi.org/10.1109/EMBC.2014.6944757