Pupillometry and Head Distance to the Screen to Predict Skill Acquisition During Information Visualization Tasks

Dereck Toker, Sébastien Lallé, Cristina Conati
Department of Computer Science
University of British Columbia, Vancouver, Canada
{dtoker, lalles, conati}@cs.ubc.ca

ABSTRACT
In this paper we investigate using a variety of behavioral measures collectible with an eye tracker to predict a user's skill acquisition phase while performing various information visualization tasks with bar graphs. Our long term goal is to use this information in real-time to create user-adaptive visualizations that can provide personalized support to facilitate visualization processing based on the user's predicted skill level. We show that leveraging two additional content-independent data sources, namely information on a user's pupil dilation and head distance to the screen, yields a significant improvement for predictive accuracies of skill acquisition compared to predictions made using content-dependent information related to user eye gaze attention patterns, as was done in previous work. We show that including features from both pupil dilation and head distance to the screen improve the ability to predict users' skill acquisition state, beating both the baseline and a model using only content-dependent gaze information.

Author Keywords
User modeling; Information Visualization; Skill Acquisition; Eye tracking; Pupil dilation; Distance to the screen; Classification

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
There is increasing evidence that users' abilities, personality, and preferences influence their performance and satisfaction during information visualization (InfoVis) tasks, e.g., [16,19,20,27]. These findings have prompted researchers to investigate user-adaptive information visualizations, i.e., visualizations that recognize and adapt to each user's specific needs. For instance, work has been done on predicting various human factors for adaptation such as: cognitive measures including perceptual speed, visual working memory, and verbal working memory [17,52]; user knowledge of the content to be visualized [14]; user task performance [26], and user confusion with the visualization interface [39]. This paper focuses on the long-term goal of devising visualizations that provide personalized support to ease a user's learning curve by supporting the transition from unskilled to being skilled at working with visualization-based tasks that are unfamiliar to the user. In order to achieve this goal, in this paper we discuss how to track users as they acquire the set of skills necessary to efficiently perform a new activity, i.e., processing and performing tasks with a target visualization in our specific case. We model skill acquisition based on the presence of a learning curve which is a standard concept in cognitive psychology used to represent the relationship between practice and the associated changes in behavior [51] (i.e., changes in skill, expertise, speed).

While learning curves have been extensively investigated to study and adapt to skill acquisition in educational settings (e.g., [5,40]), their usage for personalization in HCI and visualization has so far been limited. Still, detecting and adapting to skill acquisition is important because customized support could be offered to users if it is inferred that they are in a state of skill acquisition when working with a system, in order to improve both their short-term task performance as well as their acquisition of proficiency. For example, support could be offered by preventing access to more advanced interface features for novice users until the necessary skills are acquired, or specific functionalities that novice users might otherwise overlook could be highlighted [15,30].

In the context of information visualization research, Toker et al. [55] previously showed the presence of a learning curve in a study were users performed basic visualization tasks with bar graphs. That work was the first to explore the feasibility of detecting skill acquisition in real-time from gaze data collected with a non-intrusive eye-tracker. Skill acquisition was modeled into two stages: *during* skill learning vs. *after* skill learning. In their work, Toker et al. [55] reported a gaze-based prediction model, capable of beating a 50% baseline for a binary prediction over these two states during any given study task.

In this paper, we build upon and extend the work in [55] by investigating the benefit of using two additional measurements of user behavior detectable by an eye tracker during visualization processing: pupil dilation and distance

IUI 2017, *March 13-16, 2017, Limassol, Cyprus*

ACM 978-1-4503-4348-0/17/03…$15.00
DOI: http://dx.doi.org/10.1145/3025171.3025187

of the head to the screen (head distance, for short). We make the following hypothesis:

> Adding features related to a user's pupil dilation and head distance during visualization processing will improve prediction accuracy of a user's skill acquisition stage (i.e., *during* vs. *after* learning) as compared to solely relying on gaze data features.

Our hypothesis is based on the fact that these two data sources have been shown to be potential predictors of user states related to learning during interaction with educational software. For instance, pupil dilation has been consistently linked to cognitive load (e.g., [6,29,31]), which in turn has been shown to impact how much users can learn from e-learning environments [35]. Furthermore [41] showed that pupil dilation can be used to detect improvement in performance over time with a visual tool for decision making. Head distance can be seen as an indicator of body postures (i.e., leaning toward or away from the screen) that have been linked to both engagement or boredom [23,33] and to how well users learn with educational systems [4]. Furthermore, [34] has shown that head distance can predict boredom during student interaction with an computer-based tutor for biology. Here we leverage information about a user's head distance and pupil dilation for predicting two different learning states with a visualization system.

The main contribution of this paper is that our hypothesis stands. Using the existing dataset collected from the study in [55], we show that adding pupil and head distance information to previously evaluated gaze features can significantly improve binary prediction accuracy of users' skill acquisition state by as much as 5% in terms of peak accuracy, compared to solely using eye gaze features.

A second contribution relates to the feasibility of a simpler content-independent model, that can predict skill acquisition when information regarding the layout of the visualization is unknown or is potentially too challenging to model, resulting in the impossibility to track many gaze features that are specific to the visualization. We show that a model using only pupil dilation and head distance features (which do not require knowledge of the visualization layout) is still capable of reaching predictive accuracies of 60% in 13 seconds (a bit more than halfway through the duration of a single task), outperforming a majority class baseline. Making predictions using solely content-independent features in this way provides evidence toward the potential generalizability of our approach to other types of visualizations.

In the rest of the paper, we first discuss related work. Next, we describe the visualization and dataset utilized. We then show the presence of a learning curve and how the binary skill acquisition states are defined, similar to [55]. After that, we summarize the new approaches we use to build our predictive models. We conclude with results and a discussion of main findings and work to come.

RELATED WORK

A typical method used in cognitive psychology for modeling how user performance improves with practice is by using a learning curve [51]. Learning curves are also frequently used in HCI for off-line comparison and evaluation of information visualization systems, (e.g., [47,50,57,58]). In contrast, we leverage the concept of a learning curve for building predictive models that can identify in real-time during task interaction two broad stages of a user's skill acquisition while working with an information visualization system.

Similar work has been extensively conducted in the field of Intelligent Tutoring Systems (ITS), using learning curves to track and adapt to a student's evolving domain knowledge (as opposed to level of skill in using the system itself) while working with educational software. Hidden Markov Models [3] or logistic regressions [48] were used to infer students' mastery in a variety of domain skills (e.g., performing one and two digit subtraction for a math tutor) based on students' past performance and interaction logs [3], or based on speech output [7]. In InfoVis, Item Response Theory has been used to assess a user's visualization literacy, i.e., the user's skill in using visualizations to handle information in an effective/efficient manner [13]. In contrast, we use eye tracking data, namely gaze movements, pupil dilation, and head distance to the screen, to dynamically detect a user's evolving proficiency in working with a visualization, in terms of two overall skill acquisition phases (*during* learning and *after* learning).

Eye gaze data has been extensively used to detect different kinds of user's states during interaction with an ITS, such as boredom, curiosity, disengagement [22,34], mind-wandering [11], as well as domain learning [12,36]. In addition, [8] used gaze data to predict users' problem-solving strategies as well as user performance while solving a visual puzzle. In InfoVis, gaze data has previously been used to carry out off-line analysis to understand how users with different expertise or abilities process visualizations. For instance, offline analysis of gaze data was used to explain why performance differences occurred between users while working with bar and radar graph visualizations (e.g., users were having difficulty processing the visualization's legend)[53]. Offline analysis was also used to understand processing differences with highlighting interventions provided on bar graphs [54], and to understand how users with different domain expertise processed visualizations (e.g., [18,46]). Gaze data has also been used online to predict users' problem-solving strategies performance while solving a visual puzzle [8]. In InfoVis, online analysis of gaze data has also been investigated to predict in real time long-term user's

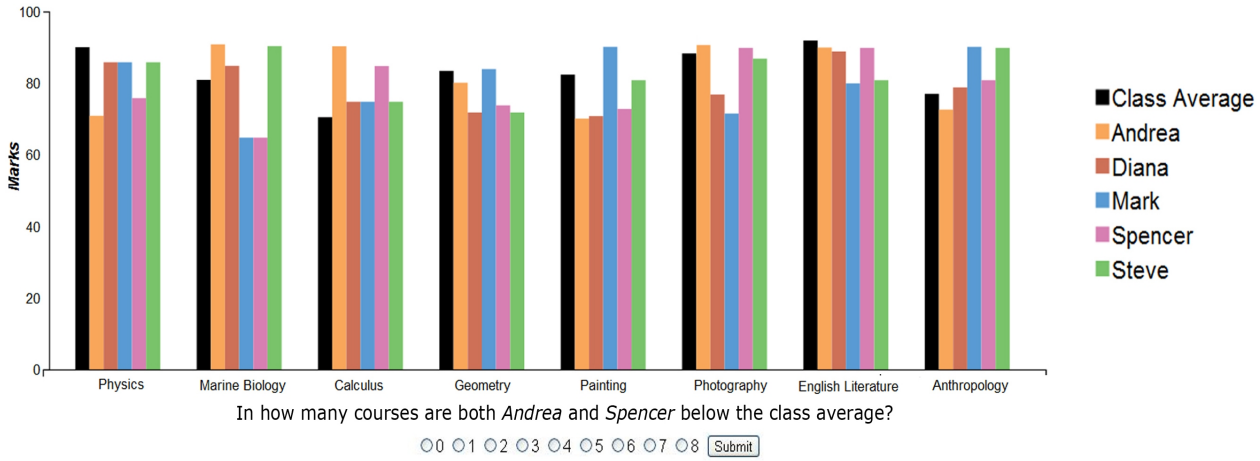

Figure 1: Sample bar graph visualization and task administered to users during the study.

cognitive abilities/traits (e.g., perceptual speed, visual working memory, verbal working memory, locus of control), as well as task type, task completion time, and user confusion [25,39,52].

Pupil dilation has been investigated as a source of information for user-adaptive systems because it has been shown to relate to changes in cognitive load (e.g., [6,29,31]). Iqbal et al. [32] evaluated cognitive workload via pupillary measures during route planning and document editing tasks in order to identify opportune moments for interrupting the user without causing excessive interference with their primary tasks. Prendinger et al. [49] monitored pupil dilation in order to predict user preferences when confronted with a choice of objects presented on the screen. Martinez-Gomez & Aizawa [44] tracked pupil dilation to infer a user's reading comprehension, and consequent topic familiarity. Lallé et al. [39] showed that including pupil dilation measures, in addition to eye gaze measures, improved the capability of predicting user confusion with an interactive visualization to support decision making.

Head distance and body postures have been identified as reliable indicators of users' affective state. For instance, D'Mello et al., [23] found that leaning backward, as tracked by a posture chair fitted with multiple sensors sensitive to pressure, can be a good predictor of boredom or disinterest in an educational context. Jaques [34] found similar results with a simpler indicator of posture, namely the viewing distance of the user's head from the screen, measured by a Tobii T60 eye tracker. Specifically, the results in [34] show that a model solely based on head distance significantly outperforms a majority baseline to predict boredom during interaction with an ITS for biology, and confirmed that a greater head distance was correlated to feeling bored. Since boredom has been related to learning [4], in this paper we investigate whether head distance can also be used as a useful predictor of users' skill acquisition state while working with an information visualization.

An alternative approach to predict skill acquisition in InfoVis is described in [41], which requires gathering information over multiple interface usages for each user. A learning curve was fit for each individual participant using a power law function, which captures the user's initial level of expertise with a given visualization, as well as the rate of learning with the visualization. However, due to how the learning curves are modeled, the approach in [41] requires access to the history of a given user in terms of past exposition to the visualization, as predictions were made across a series of consecutive tasks completed by each user. Therefore, adaptive support could only be provided for subsequent tasks since the prediction of users' learning curves are made either at the end of or between tasks. In our paper, we adopt a within-task oriented approach where user skill is predicted during the task. Specifically, information is collected from the beginning of a task without looking at previous performance data from earlier tasks (if they even exist). This approach is thus more suitable in situations where users interact with a visualization system only once, or when the user history in terms of the amount of practice with a visualization is not available. For instance, these conditions may occur with public kiosks or web-based visual tools which are typically designed for broad general audiences. Our approach can also allow for the swift delivery of adaptive support to users since predictions are possible after only a few seconds of observed interaction with a task.

DATASET, FEATURES, & LABELS

In this paper, we employ an existing corpora of data generated from a prior study. We leverage the data from this study in order to investigate users' skill acquisition while they perform a series of 80 basic visualization tasks using bar graphs. The dataset consists of task performance and eye tracking data for 62 participants. Over the course of 90 minutes, each participant completed 80 randomized tasks, covering several combinations of task type and experimental conditions (Figure 1 shows an example task used in the study). The study tasks involved comparing individuals against a group average (data points in the bar graph) on a set of dimensions (data series in the bar graph). For variety, the task questions were drawn from four

different domains. All tasks involved the same number of data points (six, including the average) and data series (eight). There were two types, chosen from a set of primitive data analysis tasks that [2] identifies as "largely capturing people's activities while employing information visualization". The first task type was Retrieve Value (a relatively simple task), which consisted of retrieving a specific individual in the target domain and comparing it against the group average; (e.g., "Is Christopher's grade in English below the class average for that course?"). The second task type was Compute Derived Value (a more complex task type), which required users to first perform a set of comparisons, and then compute an aggregate of the comparison outcomes; (e.g., "In how many cities is the movie The Lost Explorer above the average revenue and the movie An Unfinished Life below it?"). User gaze was tracked with a Tobii T120 eye-tracker, used as the study main display. Baseline pupil width was collected from each participant at the beginning of the study, with lighting conditions strictly controlled and remaining constant during the study. For a complete description of the study see [17].

Eye Tracking Feature Sets

Here we describe the three different feature sets generated from eye tracking data. All participants were required to have a visual acuity of 20/20, either uncorrected or corrected with glasses.

Gaze Features. Raw gaze data consists of fixations (points of gaze on the screen) and saccades (quick movements between fixations). Raw gaze data is collected from the Tobii T120 eye tracker using the ClearView fixation filter, and is then processed with EMDAT (*www.github.com/ATUAV/EMDAT*) to generate a battery of aggregate gaze-based features. Some of these features capture overall gaze activity on the screen (see 1a. in Table 1) while others do so for specific Areas of Interest (AOI) in the visualization (see 1b. in Table 1). Six areas of interest corresponding to various conceptually distinct regions of the visualization layout are utilized (see Figure 2).

Pupil Dilation Features. The Tobii T120 eye-tracker records the user's pupil diameter (the horizontal width of each pupil) at each sample (120hz). Similar to gaze, we used EMDAT to compute a variety of features that describe the pupil diameter over the span of a task, for a total of 10 features (see 2. in Table 1). The features *mean*, *stddev*, *min* and *max* pupil width are included since other work has used these measures to capture the range of a user's cognitive load during tasks [44]. Additionally, we include the *start* and *end* pupil width, because research has shown that there can be local peaks and troughs of users' cognitive load at boundaries between sub-tasks [32]. As for pupil velocity, we also generated the *mean*, *stdev*, *min* and *max*. Previous work has used pupil velocity to infer users' search intentions in video retrieval tasks [56], as well as reading comprehension [44]. To account for potential physiological differences in pupil size among individual users, measured pupil dilation values for each user are adjusted with respect to their baseline using the percentage change in pupil size (PCPS), reported in μm, which [32] defines as:

$$\frac{\text{measured pupilsize} - \text{baseline pupilsize}}{\text{baseline pupilsize}}$$

Pupil dilation features are generated without any knowledge of the visualization layout, and are thus considered content-independent. While including other more complex features such as Index of Cognitive Activity [42] and maximum pupil power [9] may lead to even better prediction results, we investigate only basic standard pupil features given that our main goal is to determine the general usefulness of including pupil features for predicting users' skill acquisition phase.

1. **GAZE** Features (86 total):
a) AOI-*Independent* Features (14 total):
Sum, Mean, & Stddev of fixation durations
Sum, Mean, & Stddev of saccade distance
Sum, Mean, & Stddev of relative saccade angles
Sum, Mean, & Stddev of absolute saccade angles
Fixation rate
Count of fixations
b) AOI-*Specific* Features (72 total):
Fixation rate on AOI
Longest fixation on AOI
Time of first & last fixation on AOI
Sum of fixation durations on AOI
Count of fixations on AOI
Count of transitions from *this* AOI to *each* AOI
2. **PUPIL** Features (10 total):
Mean, Stddev, Min, & Max of pupil width
Mean, Stddev, Min, & Max of pupil dilation velocity
Pupil width at the first & last fixation in a given task
3. **HEAD DISTANCE** Features (6 total):
Mean, Stddev, Min, & Max of head distance to screen
Head distance at the first & last fixation in a given task

Table 1. Set of features generated using Tobii T-120 eye tracking setup and EMDAT processing.

Head Distance to Screen Features. The Tobii T120 eye tracker measures head distance by recording the viewing distance from both the user's eyes to the screen at each sample (120Hz). In order to estimate head distance to the screen, EMDAT averages the viewing distance of the left and right eye, measured in *cm*. As with pupil width measures, we used EMDAT to compute a similar set of features that describe user head distance to the screen over the span of each task (see 3. in Table 1). Since head distance features are computed independent of the visualization layout, they are also considered content-independent.

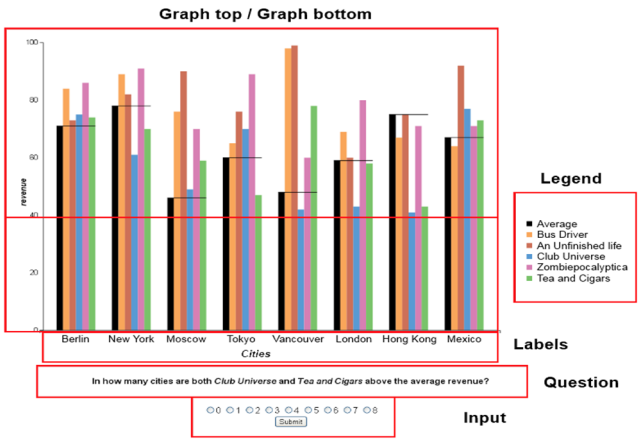

Figure 2: Areas of Interest (AOI) defined over the interface.

Labeling Skill Acquisition

A previous analysis of this dataset detected the presence of a learning curve [55] shown in Figure 3, where the average task completion time across all users is plotted over the 80 study tasks in ascending order of completion. For the first 40 trials task performance continues to improve while users become more practiced as they perform additional tasks (left of blue dashed line in Figure 3). For the successive 40 trials (right of blue dashed line in Figure 3), performance stabilizes as indicated by both reduced variance across trials and a lower bound on performance (dotted green horizontal line). Therefore, the first 40 trials that each user performs are labeled as *during* skill acquisition and the last 40 trials as *after* skill acquisition.

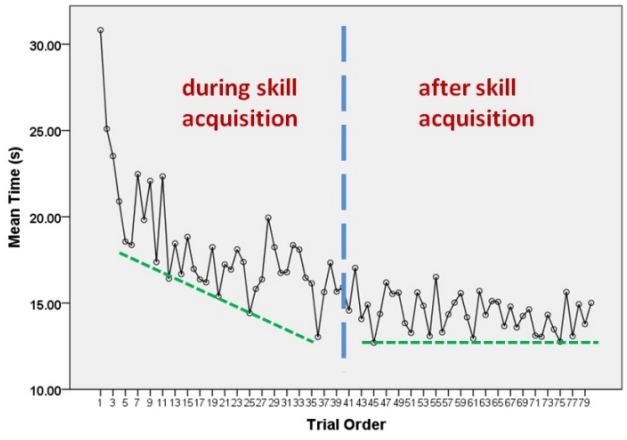

Figure 3: Improvement in average trial completion time across the 80 tasks in the dataset (randomly administered for each user). The blue line separates trials into two general stages of skill acquisition: DURING - skill with the visualization is in the state of being acquired since performance is still improving; and AFTER - skill with the visualization has been acquired since performance change has stabilized.

MACHINE LEARNING SETUP

The aim of this work is to use eye tracking data as input in order to predict the correct skill acquisition label (i.e., *during* vs. *after* skill acquisition) on any given trial without knowing which trial a user is currently doing. In order to simulate real-time predictions of a user's skill level while engaged with a given task, we generate features over consecutively increasing time slices corresponding to partial observations of eye tracking data during a task. These time slices range from 2 to 20 seconds (20 seconds is the mean time to complete a task), over 1 second intervals for each task. For example, features generated at a 6 second time slice would model the real-world scenario where an adaptive visualization has observed only the first 6 seconds of a user's behavior from the beginning of the current task. At each of the 19 time-slices, we evaluate 5 different feature set combinations derived from eye-tracking data (i.e., GAZE, PUPIL, HEADDISTANCE). We also include a baseline model, for a total of 6 models executed at each time slice.

To predict users' skill acquisition phase, we built five different binary machine learning classifiers using the Caret machine learning package in R [38], and reported classifier performance as predictive accuracy, i.e., the total number of correct predictions divided by the total number of correct and incorrect predictions. First we tried linear regression, since it has been used previously for making predictions using similar data (e.g., [52]). Next, we tried four standard machine learning algorithms (Naive Bayes, SVM, Neural net, and Random forest), to see if it was possible to achieve better performance given that our paper includes additional types of attributes (pupil & head distance) compared to previous work. Overall we saw better predictive accuracy from the Random Forest algorithm (which was also found to be the case for data collected from a different study reported in [25]), and we thus opted to report results for Random Forest only.

In order to simulate real-world settings where data regarding a new user is unknown, classifiers were evaluated using 10-fold cross validation over users (i.e., at each fold of the cross validation, users in the test set do not appear in the training set). Then, we repeat this process 5 times (runs) to strengthen the stability and reproducibility of the results, and the performance of each algorithm is averaged over the 10 folds and the 5 runs.

Model baseline

Since classifications are done using consecutively increasing partial observations of eye tracking data within a given task (e.g., 2s, 3s, 4s, ... up to 20s), cases arise where some users complete the task in under 20 seconds, resulting in time slices in which several users are already done with the task. To generate a rigorous baseline, we remove such users from our dataset at those time slices before classifying each new time slice within a task. Retaining these users may bias our eye tracking features since several of them are correlated with time (e.g., sum fixation durations). Thus the majority class baseline is recalculated accordingly as time elapses within a task. In our dataset, not surprisingly, users who finish earlier within a given trial are more likely to be skilled users (i.e., users in the *after* skill acquisition state),

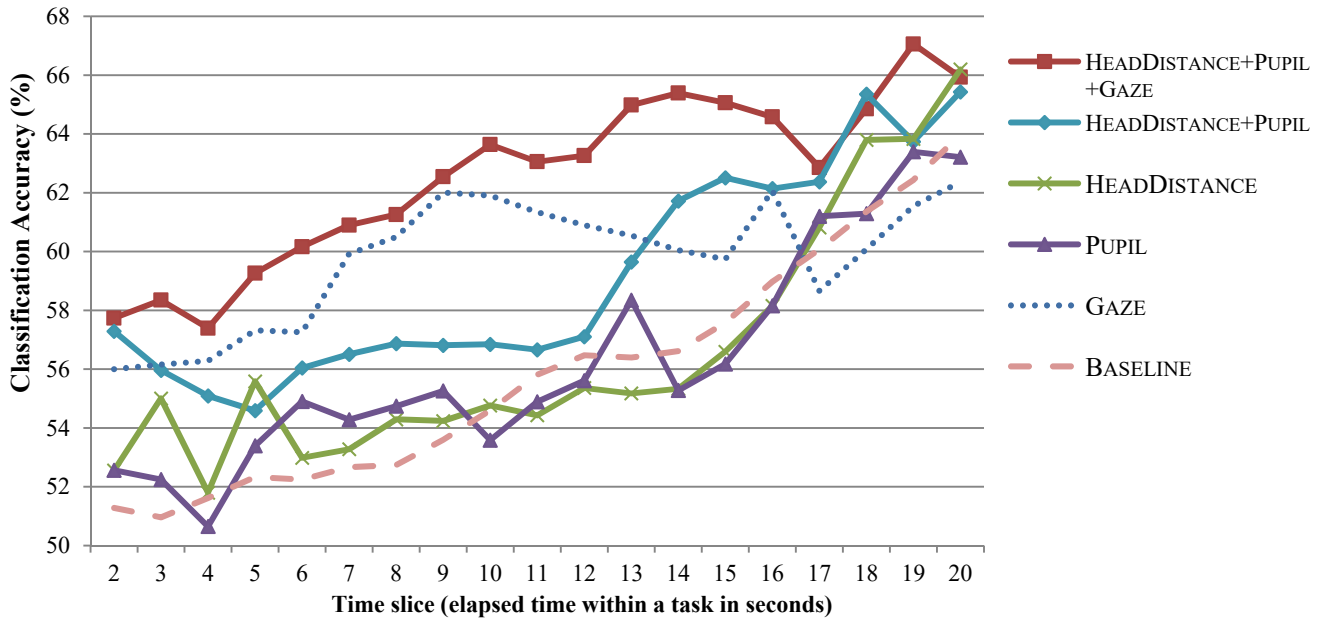

Figure 4: Predictive accuracy across time slices based on feature set combination. GAZE is shown using a dotted blue line and corresponds to the best model previously published in [55].

which results in a rising proportion of unskilled users (i.e., users in the *during* skill acquisition state) as time lapses over a given trial. The dashed red line in Figure 4 shows indeed that this strict baseline becomes more weighted as time unfolds, with a starting baseline accuracy of 51% which rises over time to 64% baseline accuracy.

RESULTS

We first compare the performance accuracies of the various combinations of GAZE, PUPIL and HEADDISTANCE models, with the specific goal of ascertaining the added predictive value when including the PUPIL and HEADDISTANCE features along with GAZE. We then report the most predictive features of the best performing model and discuss how these features relate to skill acquisition in terms of directionality of the underlying features themselves.

Predicting Skill Acquisition

Figure 4 reports the accuracy over consecutive time slices (i.e., over the 19 time windows of increasing length described earlier) of the 5 combinations of tested feature sets, as well as the accuracy over-time of the baseline (dashed red line). Note that the model that previously obtained the highest accuracy in [55] (i.e., GAZE) is represented by dotted blue line[1]. The trends shown in Figure 4 provide an initial assessment of how much interaction data a real-time classifier of skill acquisition would need in order to generate reliable predictions. Ultimately, depending on how early within the task adaptive support is required, Figure 4 illustrates the tradeoff in accuracy when predicting skill acquisition early on versus delaying the prediction as time elapses.

To formally compare the accuracies of the 6 classifiers (i.e., 5 feature set combinations + 1 baseline), we run a linear mixed-effects model [24] with *feature set* (6 levels) and *time-slice* (19 levels) as the two independent variables, and *predictive accuracy* as the dependent measure. The analysis revealed a significant main effect of feature set on classification accuracy ($F_{5,470} = 136.59$, $p < .001$), which indicates that overall significant differences exist between feature sets regardless of the amount of eye-tracking data available for classification as a task unfolds. Follow-up Bonferroni adjusted pairwise comparisons of the *over-time accuracy* for each of the 6 levels, shown in Table 2, revealed that:

- The HEADDISTANCE+PUPIL+GAZE model is better than all other models.
- Each of the PUPIL and HEADDISTANCE models do not beat the baseline.
- The GAZE model and HEADDISTANCE+PUPIL model beat the baseline, but are not significantly different from each other.

We can see in Table 2 that all models using either eye-tracking features or combining pupil and head distance features together outperform the baseline. The GAZE only model (investigated in [55]), outperforms PUPIL only and HEADDISTANCE only, but it is then outperformed by HEADDISTANCE+PUPIL+GAZE, indicating that combining all three feature sets significantly improve prediction accuracy of a user's skill acquisition.

[1] Note that previous work in [55] did not perform user-independent prediction, explaining the slightly higher accuracies reported in [55].

Models	Average over-time accuracy
HEADDISTANCE+PUPIL+GAZE	62.5%
GAZE	59.7%
HEADDISTANCE+PUPIL	59.1%
HEADDISTANCE	56.5%
PUPIL	56.3%
BASELINE	55.9%

Table 2: Effect of feature set combination on overall model performance averaged across all time-slices. Rows are arranged in descending order of classifier accuracy. Dashed lines separate models that are not statistically different from one-another.

In terms of peak accuracies, Figure 4 shows the best accuracy for the GAZE only model at 62.5%, whereas HEADDISTANCE+PUPIL+GAZE has a peak accuracy of 67%. Additionally, in terms of early prediction capabilities, HEADDISTANCE+PUPIL+GAZE achieves 57.5% after having seen only 2 seconds of a user interacting with the system (from a 51% baseline) and gets to 64% halfway through the duration of the interaction. Even though Figure 4 shows that the GAZE only model also performs relatively well during the first 10s, pairwise comparisons of the *over-time accuracy* for only the first 10s indicates that HEADDISTANCE+PUPIL+GAZE is in fact still significantly better than GAZE only ($p < .001$) for early prediction. Interestingly, the upward trend of the GAZE model ceases after 10 seconds. Although we don't have a clear explanation for this finding, it is worth considering that many GAZE features are sensitive to accumulation (e.g., *sum*, *count*, *total time spent, etc.*), and thus might become less informative as time elapses.

Also worth noting is the fact that the model combining HEADDISTANCE +PUPIL still beats the baseline, with an over-time accuracy of 59.1%, which is only 3% behind the model using HEADDISTANCE+PUPIL+GAZE. This result is important because it illustrates the potential of utilizing only a few eye tracker features (in this case 16 features, as opposed to 102 features when including GAZE information) and leaner feature sets are generally known to be less likely to overfit unseen data [1]. Furthermore, HEADDISTANCE and PUPIL do not require knowledge of what is displayed on the screen, namely, they are content-independent. Although HEADDISTANCE+PUPIL reaches 60% accuracy in about 13 sec. (around two thirds of the interaction), Figure 4 shows that the accuracy of this model is not as good as GAZE during the first 12 seconds, and increases considerably afterward. Interestingly, HEADDISTANCE+PUPIL exhibits similar accuracy as the best model at the beginning 2 seconds of the task. Although we can't clearly explain these findings, investigation of the most important features (see next Subsection) can provide more details about these trends. Overall, from a practical point of view, these results suggest that content-independent features only (pupil and head distance) are promising toward generalization, but may require a slightly delayed adaptation or customization offered to the users.

In terms of the added value of HEADDISTANCE and PUPIL feature sets as predictive sources, the fact that the combined HEADDISTANCE+PUPIL significantly outperforms either PUPIL only or HEADDISTANCE alone indicates that these two features set do not capture overlapping information and thus both feature sets ought to be utilized if possible.

Most Predictive Features

We report the top features from the best performing classifier identified in the previous subsection, namely: HEADDISTANCE+PUPIL+GAZE. The purpose of reporting the features with the highest impact on classification accuracy is to shed light on which specific features within GAZE, PUPIL, and HEADDISTANCE contribute to the model and thus to what extent these features may relate to skill acquisition. Once trained, the random forest algorithm we used provides importance scores based on how much each feature contributes to making successful predictions. Since classifiers are constructed at each time-slice from 2s to 20s, we determine the features with the highest importance by averaging their scores across all time slices. The resulting averages are normalized so that the most important feature has a score of 100. Features with the 10 highest scores are shown in Table 3.

Next, to gain insight into the underlying directionality of the features, we compute a difference in values of each feature between the two states of skill acquisition (last column in Table 3), by subtracting a feature's mean value for all *after* tasks from the mean value of *during* tasks. For instance, since the difference in mean values for starting head distance is negative, -17.06 cm, it indicates that the values for this feature are typically lower in the *during* state of skill acquisition (i.e., closer to the screen).

Feature	Importance	Unit	*during - after**
pupil_width · max	100	*µm*	27.84
head_distance · start	91	*cm*	-17.06
pupil_width · mean	84	*µm*	32.02
question AOI · duration	82	*ms*	51.32
head_distance· min	79	*cm*	-17.17
pupil_width · start	64	*µm*	29.14
pupil_width · end	59	*µm*	29.44
pupil_velocity · max	56	*µm/ms*	27.02
question AOI · longest fix.	52	*ms*	97.17
pupil_velocity · stddev	51	*µm/ms*	111.2

Table 3: Top 10 most predictive features across all time slices, along with directionality of the feature. *Negative values indicate the feature is lower DURING skill acquisition.

Head Distance: As Table 3 shows, 2 head distance features are in the top 10 (*start* and *min*), with head distance at the start of a task being the more important feature. *Start* head

distance captures how close a user is to the screen at the very beginning of the tasks. In terms of directionality, starting head distance to the screen is closer *during* skill acquisition, meaning that users lean in more at the beginning of the task while they are still learning the system. It is worth mentioning that the value for the starting head distance feature does not change as a task unfolds. Thus it makes sense that if starting head distance is the second most predictive feature, then it would offer similar predictive value whether 2 seconds or 20 seconds have elapsed in the task. This is a very promising feature in terms of early predictions because it can be obtained at the very beginning of a task with little knowledge about the user. As users become more accustomed to the study tasks/visualization in the latter half of the study, they are leaning back more at the beginning of each task and are likely more relaxed and confident with the system. *Minimum* head distance is the next most important head distance feature. In particular, this feature tracks the closest recorded head distance to the screen as a task unfolds. Unlike starting distance, this value could change during the course of a task (e.g., a user may lean in close partway in the task as opposed to at the start). Thus, minimum head distance likely captures engagement in the same way as starting distance, but this measure is sensitive to engagement/difficulty that occurs at later moments in the task.

Pupil Dilation: Six pupil dilation features are among the top 10 in Table 3: *max*, *mean*, *start*, and *end* pupil width, along with *max* and *stddev* pupil velocity. *Max* pupil width is also the most important feature overall. For all of these pupil features, they are larger *during* skill acquisition. Larger pupil width [6,29,31] and faster repeated changes in pupil dilation [42] have reliably be linked to higher cognitive load. Thus these results suggest that cognitive load was both greater and less consistent *during* skill acquisition. Or conversely, for *after* tasks, users required less cognitive load (and consistently so) once the necessary skills to work with the tasks/visualization were obtained. Similar to *start* head distance, the *start* pupil is obtained at the very beginning of a task, and thus is promising in terms of early predictions as well.

Gaze: Two AOI (Area of Interest) features are among the 10 most important in Table 3, and are both related to the question AOI, which covers the region of the visualization where the study tasks were displayed to the user (see Figure 3). These two features track the total fixation duration (*sum_fixation_durations*) and the duration of the *longest fixation* in the question AOI. The directionality indicates that users spent more time fixated within the question AOI of the visualization *during* skill acquisition, and also had larger maximum fixation durations. This finding indicates that features relating to the question AOI (as opposed to the other AOIs) are most useful for predicting skill acquisition, likely due to the fact that as time passes, users become more familiar and proficient with how the task questions are posed and structured, and thus come to need less time to read/process them.

DISCUSSION AND CONCLUSIONS

In this paper, we presented work on classifying skill acquisition using various eye tracking data sources, with the long-term goal of using this research to design user-adaptive visualizations that can personalize the interaction to a user's current skill state. Specifically, we investigated if and how using added feature sets based on pupil dilation and head distance to the screen can improve prediction of skill acquisition compared to solely using gaze movements, as was done in [55].

We show that when using pupil dilation and head distance feature sets together we can beat the baseline compared to using either feature set alone. Furthermore, combining pupil, head distance, and gaze features not only performs significantly better than using gaze only, but also achieve accuracies that are promising toward guiding real-time interventions. This better performing classifier achieves an overtime accuracy of 62.5% on unseen users, compared to 59.7% using solely gaze behavior. Even after seeing only 10 seconds of observed data, this classifier can predict a new user's skill acquisition phase with 64% accuracy halfway through the duration of the interaction (from a 55% baseline), providing encouraging evidence on the feasibility of early prediction of users' skill acquisition phase based on the various information sources available through an eye-tracker. Early prediction is of prime importance for our long-term goal of adapting a visualization to the current skill acquisition phase of the users. We also show that when using only content-independent eye tracking features together (pupil and head distance), skill acquisition can be predicted with an overtime accuracy of 60% after having seen about two thirds of the duration of the interaction. Although this result indicates that adaptation or customization driven by only pupil and head distance features may require a slightly delayed prediction, our findings are still promising for the possible generalization to other visualizations or interfaces since pupil and head distance features are computed independent of the visualization/interface layout.

By investigating the most predictive features in our best performing classifier, we identified both increased pupil dilation related measures and leaning closer to the screen as key behaviors present while users are becoming familiar with the visualization system. Increased pupil dilation measures are most likely an indication that users have a higher cognitive load while learning the skills necessary to work with the visual tool. For head distance, leaning forward to the screen might indicate that users pay more attention to the components of visualization or are trying to concentrate more while they are less familiar with the tasks and visualization. Conversely, leaning back from the screen might reveal that users feel more at ease *after* skill acquisition has occurred.

One caveat of our findings is that it can be difficult to reliably track pupil dilation in real-world settings, because of its well-known sensitivity to changes in environment lighting (e.g., [29]). Nevertheless there is already work showing that changes in lighting can be mitigated using advanced techniques based on wavelet decomposition [43], thus as part of our future work we plan to conduct studies to evaluate the effectiveness of these techniques for our user-modeling purposes.

In sum, our work has provided initial evidence on the added value of using pupil dilation and head distance to predict skill acquisition during InfoVis interactions with bar graphs, with the long-term goal of creating visualizations that can support users detected to be in the skill acquisition phase. Having such visualizations is especially useful in single-serving or walk-up-and-go contexts, where users need to interact with a visualization for a limited time and would benefit from having support that helps them accomplish their desired tasks if they are not familiar with the interface.

To illustrate with a real-world example, *multi-modal documents* containing text that describe different aspects of accompanying graphs are extensively used in publications directed toward a broad audience (e.g., articles from the *Economist*) [21]. Typically, documents of this type are viewed only once, thus detecting skill acquisition quickly within a single-serving scenario could be of great value. There is already work on generating corpora of multimodal documents with explicit links between elements in the visualization and related sentences [37]. We are planning to leverage these corpora to generate an adaptive system that can track which reference to the visualization a user is reading in the text, and whether the user is unskilled with the visualization. Users' attention can then be adaptively cued to relevant elements of the visualization using techniques such as highlighting (see [17] for examples of visual prompts evaluated on bar graphs).

A second example of where we envision user-adaptive visualizations based on user skill is with MetroQuest [28]. MetroQuest is a commercialized decision-support tool deployed to engage and educate communities about urban plans, as well as to collect informed input to help policy makers understand the expectations of their target audiences. This tool aims to increase community awareness by providing users with several visualizations like deviation charts and interactive maps. Designing MetroQuest interfaces is challenging as this tool is often used in public kiosks by users with very heterogeneous backgrounds. For instance, while complex visualizations conveying rich information would satisfy some users, they may overwhelm others who abandon their task as a result. The challenge is exacerbated since MetroQuest is typically used as a walk-up-and-use system (e.g., in public kiosks) that, in order to avoid attrition, must be self-explanatory and engaging to first-time users. Having the ability to provide adaptive support or customization based on a user's skill acquisition phase would allow MetroQuest to potentially increase user engagement, and reduce attrition. Adaptive support could involve, for instance, displaying only one visualization for which the system detects that the user has sufficient skill for comprehension. Alternatively, the system could provide visual cues to facilitate the processing of the available visualizations, as discussed above.

As future work, we plan to run studies to establish if/how the results we have presented on predicting skill acquisition generalize to other visualizations beyond bar graphs, especially in settings relating to the two real-world applications described above. We will also investigate further improvements to our classifiers for skill acquisition. For example, we plan to expand our set of eye tracking features to include more complex pupil measures such as maximum pupil power [9], as well as features based on the rate of change of our eye tracking measures (e.g., pupil and saccade acceleration) given that evidence has shown that kinematic features have the potential to further improve prediction accuracies of other user states [10,44]. We will also explore integrating eye-tracking data with complementary input features such as mouse movements [45], or interface actions when available as suggested by [36], that could also serve to improve prediction accuracies of skill acquisition.

REFERENCES

1. Akaike, H. Factor analysis and AIC. *Psychometrika 52*, 3 (1987), 317–332.
2. Amar, R., Eagan, J., and Stasko, J. Low-Level Components of Analytic Activity in Information Visualization. *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, IEEE Computer Society (2005), 15–21.
3. Baker, R., Corbett, A., and Aleven, V. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. *Proceedings of the 9th International Conference Intelligent Tutoring Systems*, (2008), 406–415.
4. Baker, R.Sj., D'Mello, S.K., Rodrigo, M.M.T., and Graesser, A.C. 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 (2010), 223–241.
5. Barnes, T. and Stamper, J. Toward automatic hint generation for logic proof tutoring using historical student data. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, Springer (2008), 373–382.
6. Beatty, J. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin 91*, 2 (1982), 276–292.
7. Beck, J.E. and Sison, J. Using knowledge tracing in a noisy environment to measure student reading proficiencies. *International Journal of Artificial Intelligence in Education 16*, 2 (2006), 129–143.
8. Bednarik, R., Eivazi, S., and Vrzakova, H. A Computational Approach for Prediction of Problem-Solving Behavior Using Support Vector Machines and Eye-Tracking Data. In Y.I.

Nakano, C. Conati and T. Bader, eds., *Eye Gaze in Intelligent User Interfaces*. Springer London, London, 2013, 111–134.
9. Biswas, P., Dutt, V., and Langdon, P. Comparing Ocular Parameters for Cognitive Load Measurement in Eye-Gaze-Controlled Interfaces for Automotive and Desktop Computing Environments. *International Journal of Human-Computer Interaction 32*, 1 (2016), 23–38.
10. Bixler, R. and D'Mello, S. Automatic Gaze-Based Detection of Mind Wandering with Metacognitive Awareness. In F. Ricci, K. Bontcheva, O. Conlan and S. Lawless, eds., *User Modeling, Adaptation and Personalization*. Springer, Cham, 2015, 31–43.
11. Bixler, R., Kopp, K., and D'Mello, S. Evaluation of a Personalized Method for Proactive Mind Wandering Reduction. *Proceedings of the 4th Workshop on Personalization Approaches for Learning Environments, 22nd conference on User Modeling, Adaptation, and Personalization*, Springer (2014), 33–41.
12. Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J.M., Azevedo, R., and Bouchet, F. Inferring Learning from Gaze Data during Interaction with an Environment to Support Self-Regulated Learning. In H.C. Lane, K. Yacef, J. Mostow and P. Pavlik, eds., *Artificial Intelligence in Education*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, 229–238.
13. Boy, J., Rensink, R.A., Bertini, E., and Fekete, J.-D. A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1963–1972.
14. Brusilovsky, P., Ahn, J., Dumitriu, T., and Yudelson, M. Adaptive Knowledge-Based Visualization for Accessing Educational Examples. *Proc. 10th International Conf. on Information Visualization*, IEEE (2006), 142–150.
15. Bunt, A., Conati, C., and McGrenere, J. Supporting interface customization using a mixed-initiative approach. ACM Press (2007), 92.
16. Carenini, G., Conati, C., Hoque, E., Steichen, B., Toker, D., and Enns, J.T. Highlighting Interventions and User Differences: Informing Adaptive Information Visualization Support. *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, (2014).
17. Carenini, G., Conati, C., Hoque, E., Steichen, B., Toker, D., and Enns, J.T. Highlighting Interventions and User Differences: Informing Adaptive Information Visualization Support. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2014), 1835–1844.
18. Çöltekin, A., Fabrikant, S.I., and Lacayo, M. Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Science 24*, 10 (2010), 1559–1575.
19. Conati, C., Carenini, G., Hoque, E., Steichen, B., and Toker, D. Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. *Computer Graphics Forum*, Wiley Online Library (2014), 371–380.
20. Conati, C., Carenini, G., Toker, D., and Lallé, S. Towards User-Adaptive Information Visualization. *AAAI Conf. on Artificial Intelligence*, (2015).
21. Conati, C., Hoque, E., Toker, Dereck, and Steichen, Ben. When to Adapt: Detecting User's Confusion During Visualization Processing. *Proc. of the 1st International Workshop on User-Adaptive Information Visualization (WUAV), in conjunction with the 21st Conference on User Modeling, Adaptation and Personalization (UMAP)*, (2013).
22. D'Mello, S., Olney, A., Williams, C., and Hays, P. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies 70*, 5 (2012), 377–398.
23. D'Mello, S.K., Craig, S.D., Sullins, J., and Graesser, A.C. Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education 16*, 1 (2006), 3–28.
24. Field, A. *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications Ltd, London, 2012.
25. Gingerich, M.J. and Conati, C. Constructing Models of User and Task Characteristics from Eye Gaze Data for User-Adaptive Information Highlighting. *Twenty-Ninth AAAI Conference on Artificial Intelligence*, (2015).
26. Gotz, D. and Wen, Z. Behavior-driven visualization recommendation. *Proc. of the 14th international conf. on Intelligent user interfaces*, ACM (2009), 315–324.
27. Green, T.M. and Fisher, B. Impact of personality factors on interface interaction and the development of user profiles: Next steps in the personal equation of interaction. *Information Visualization 11*, 3 (2012), 205–221.
28. Haas Lyons, S., Walsh, M., Aleman, E., and Robinson, J. Exploring regional futures: Lessons from Metropolitan Chicago's online MetroQuest. *Technological Forecasting and Social Change 82*, (2014), 23–33.
29. Hess, E.H. and Polt, J.M. Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science 143*, 3611 (1964), 1190–1192.
30. Hurst, A., Hudson, S.E., and Mankoff, J. Dynamic Detection of Novice vs. Skilled Use Without a Task Model. *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2007), 271–280.
31. Hyönä, J., Tommola, J., and Alaja, A.-M. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology 48*, 3 (1995), 598–612.
32. Iqbal, S.T., Adamczyk, P.D., Zheng, X.S., and Bailey, B.P. Towards an index of opportunity: understanding changes in mental workload during task execution. ACM Press (2005), 311.
33. Jacobs, A.M., Fransen, B., McCurry, J.M., Heckel, F.W.P., Wagner, A.R., and Trafton, J.G. A Preliminary System for Recognizing Boredom. *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, ACM (2009), 299–300.
34. Jaques, N., Conati, C., Harley, J.M., and Azevedo, R. Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System. *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*, Springer (2014), 29–38.
35. Kalyuga, S. Enhancing Instructional Efficiency of Interactive E-learning Environments: A Cognitive Load Perspective. *Educational Psychology Review 19*, 3 (2007), 387–399.
36. Kardan, S. and Conati, C. Comparing and Combining Eye Gaze and Interface Actions for Determining User Learning with an Interactive Simulation. *In: Proc. of UMAP, 21st Int. Conf. on User Modeling, Adaptation and Personalization*, (2013).
37. Kong, N., Hearst, M.A., and Agrawala, M. Extracting references between text and charts via crowdsourcing.

Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM (2014), 31–40.
38. Kuhn, M. Building predictive models in R using the caret package. *Journal of Statistical Software 28*, 5 (2008), 1–26.
39. Lallé, S., Conati, C., and Carenini, G. Predicting confusion in information visualization from eye tracking and interaction data. *Proceedings on the 25th International Joint Conference on Artificial Intelligence*, (2016), 2529–2535.
40. Lallé, S., Mostow, J., Luengo, V., and Guin, N. Comparing Student Models in Different Formalisms by Predicting their Impact on Help Success. *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, (2013), 161–170.
41. Lallé, S., Toker, D., Conati, C., and Carenini, G. Prediction of Users' Learning Curves for Adaptation while Using an Information Visualization. *Proceedings of the 20th International Conference on Intelligent User Interfaces*, ACM (2015), 357–368.
42. Marshall, S.P. The index of cognitive activity: Measuring cognitive workload. *Proceedings of the 7th IEEE Human Factors Meeting*, IEEE (2002), 5–9.
43. Marshall, S.P. Identifying cognitive state from eye metrics. *Aviation, space, and environmental medicine 78*, Supplement 1 (2007), B165–B175.
44. Martínez-Gómez, P. and Aizawa, A. Recognition of understanding level and language skill using measurements of reading behavior. ACM Press (2014), 95–104.
45. Mueller, F. and Lockerd, A. Cheese: tracking mouse movement activity on websites, a tool for user modeling. *CHI'01 Extended Abstracts on Human Factors in Computing Systems*, ACM Press (2001), 279–280.
46. Ooms, K., De Maeyer, P., and Fack, V. Study of the attentive behavior of novice and expert map users using eye tracking. *Cartography and Geographic Information Science 41*, 1 (2014), 37–54.
47. Pascual-Cid, V., Vigentini, L., and Quixal, M. Visualising Virtual Learning Environments: Case Studies of the Website Exploration Tool. IEEE (2010), 149–155.
48. Pavlik, P.I., Cen, H., and Koedinger, K.R. Performance Factors Analysis–A New Alternative to Knowledge Tracing. *Proceedings of the International Conference on AI in Education*, IOS Press (2009), 531–538.
49. Prendinger, H., Hyrskykari, A., Nakayama, M., Istance, H., Bee, N., and Takahasi, Y. Attentive interfaces for users with disabilities: eye gaze for intention and uncertainty estimation. *Universal Access in the Information Society 8*, 4 (2009), 339–354.
50. Saraiya, P., North, C., and Duca, K. An Insight-Based Methodology for Evaluating Bioinformatics Visualizations. *IEEE Transactions on Visualization and Computer Graphics 11*, 4 (2005), 443–456.
51. Speelman, C. and Kirsner, K. *Beyond the Learning Curve*. Oxford University Press, 2005.
52. Steichen, B., Conati, C., and Carenini, G. Inferring Visualization Task Properties, User Performance, and User Cognitive Abilities from Eye Gaze Data. *ACM Transactions on Interactive Intelligent Systems*, (2014).
53. Toker, D., Conati, C., Steichen, B., and Carenini, G. Individual user characteristics and information visualization: connecting the dots through eye tracking. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 295–304.
54. Toker, D. and Conati, C. Eye Tracking to Understand User Differences in Visualization Processing with Highlighting Interventions. In V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog and G.-J. Houben, eds., *User Modeling, Adaptation, and Personalization*. Springer International Publishing, 2014, 219–230.
55. Toker, D., Steichen, B., Gingerich, M., Conati, C., and Carenini, G. Towards Facilitating User Skill Acquisition - Identifying Untrained Visualization Users through Eye Tracking. *Proceedings of the 2014 international conference on Intelligent user interfaces*, ACM (2014).
56. Vrochidis, S., Patras, I., and Kompatsiaris, I. An Eye-tracking-based Approach to Facilitate Interactive Video Search. *Proc. of the 1st ACM International Conference on Multimedia Retrieval*, ACM (2011), 43:1–43:8.
57. Zhu, Y. Measuring Effective Data Visualization. *Proc. of the 3rd International Symposium on Advances in Visual Computing*, Springer (2007), 652–661.
58. Ziemkiewicz, C. and Kosara, R. The Shaping of Information by Visual Metaphors. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1269–1276.