One Size Does Not Fit All: Considerations when using Webcam-Based Eye Tracking to Models of Neurodivergent Learners' Attention and Comprehension

Grace D. Jaiyeola
University of Denver
Denver, Colorado, USA
grace.jaiyeola@du.edu

Aaron Y. Wong
University of Minnesota
Minneapolis, Minnesota, USA
wonga@umn.edu

Richard L. Bryck
Landmark College
Putney, Vermont, USA
rickbryck@landmark.edu

Caitlin Mills
University of Minnesota
Minneapolis, Minnesota, USA
cmills@umn.edu

Stephen Hutt
University of Denver
Denver, Colorado, USA
stephen.hutt@du.edu

Abstract

This study investigates the use of webcam-based eye tracking to model attention and comprehension in both neurotypical and neurodivergent learners. Leveraging the WebGazer, a previously used online data collection tool, we collected gaze and interaction data (N=354) during online reading tasks to explore task unrelated thought (TUT) and comprehension in an ecologically valid setting. Our findings challenge the "one size fits all" approach to learner modeling by demonstrating distinct differences in indicators of both constructs between neurotypical and neurodivergent learners. We compared general models trained on the entire population with tailored models specific to neurodivergent and neurotypical groups. Results indicate that diagnosis-specific models provide more accurate predictions (AUROC's .59-.70 vs. .57 for the general model), and through SHAPley analysis, we note that the strongest indicators of each construct vary as the training population is refined, highlighting the limitations of generalized approaches. This work supports the scalability of webcam-based cognitive modeling and underscores the potential for personalized learning analytics and modeling to better support diverse learning needs.

CCS Concepts

• **Applied Computing – Education; E-learning**;

Keywords

Eye tracking, Neurodivergent learners, Task Unrelated Thought (TUT), Comprehension, Learner modeling

ACM Reference Format:

Grace D. Jaiyeola, Aaron Y. Wong, Richard L. Bryck, Caitlin Mills, and Stephen Hutt. 2025. One Size Does Not Fit All: Considerations when using Webcam-Based Eye Tracking to Models of Neurodivergent Learners' Attention and Comprehension. In *LAK25: The 15th International Learning Analytics and Knowledge Conference (LAK 2025), March 03–07, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3706468.3706472

LAK 2025, Dublin, Ireland

ACM ISBN 979-8-4007-0701-8/25/03
https://doi.org/10.1145/3706468.3706472

1 Introduction

In educational settings, maintaining attention is critical for effective learning; however, we all regularly experience task unrelated thoughts (TUT) (when our internal thoughts transition from task-related content to task-unrelated content; [3, 25]) during learning tasks. Such attentional shifts, often referred to as mind wandering [47], can significantly impede information processing and negatively impact learning outcomes [20, 59]. This is particularly relevant for neurodivergent learners [39, 50]. Despite representing 20% of the US population [28], neurodivergent learners still experience significantly worse educational outcomes than their neurotypical peers [40, 56] for many reasons, including insufficient support. One particularly relevant finding is that neurodivergent learners report significantly more TUT [3] and make less "accurate" judgments about what they do and do not know [12], all of which are likely to benefit from in-the-moment support to help redirect their thoughts and encourage deeper reflection. The challenge, however, is to provide real time analytics that can inform us when to provide that support.

Methods of tracking attention in real-time, such as the use of research-grade eye trackers [34, 68], are costly and typically require controlled laboratory environments. As a result, their applicability in broader and more naturalistic settings, particularly in educational contexts, remains limited. Recent advancements in webcam-based eye tracking technology [34, 35, 65, 68] present a promising, scalable alternative. By utilizing readily available webcam technology, these systems allow for real-time monitoring of eye movements in diverse, uncontrolled environments, providing an accessible means of studying attention and comprehension. This technology is particularly valuable in online learning contexts, in which understanding and addressing attention lapses can lead to improved educational outcomes.

This type of technology may be particularly beneficial for supporting neurodivergent learners, who may face unique challenges with attention and focus compared to their neurotypical peers [64], but who have not been traditionally represented in the LAK community (as noted by [68]). The ability to use webcam-based eye tracking to investigate their gaze patterns and attention shifts provides valuable insights into how these learners process information as well as how to provide real-time support. For example, this tracking technology can help educators and researchers better understand the cognitive and visual behaviors of neurodivergent learners, encouraging more tailored and supportive learning environments that cater to their specific needs [68].

However, much of the existing work in TUT detection has followed a "one size fits all" approach [41], which does not lend itself well to "personalized learning" that both supports the strengths as well as meets the needs that certain learner populations bring to the table. This paper, therefore, explores the use of webcam-based eye tracking to model attention in both neurotypical and neurodivergent learners together and separately, investigating the generalizability of models. Using supervised machine learning, we provide insights into how attentional patterns differ across these populations and how these differences can inform more tailored educational approaches.

1.1 Related Work

Neurodivergence. Let us first be clear about what we mean by "neurodivergent" learners. *Neurodiversity*, in general, is a concept that states that we all have different ways of processing information—much like we all have diversity in our physical differences. In turn, individuals who do not fall into the traditional or "typical" styles of processing are referred to as *neurodivergent* (i.e., diverging from the "norm") [6]. Of course, there is a range of lived experiences that are not homogenous, and it is very likely that such divergence operates on a multi-dimensional continuum. For purposes of this paper, however, we nevertheless use the conventional term to describe a group of learners who have self-identified as having one of a few different diagnoses: ADHD, autism, or a learning disability (or some combination of those). Further, we believe such diversity in our cognitive processes is useful for society and should not be viewed as a deficit or a way to "other" people in systemic and harmful ways. In fact, these differences can be a strength in processing, offering new insights and ways of solving problems.

Although different styles of thinking can be a strength, neurodivergence is not often celebrated in many existing societies (though things are moving in a positive direction in many ways), and in fact, is often explicitly ignored in the way we design our learning experiences to fit the norm [38, 61]. This is particularly noticeable when you consider that a majority of learners who identify as neurodivergent are not enrolled in special education and instead are in traditional classrooms [18, 27].

The question is then, how can we best support these individuals? One way is to provide real-time support that addresses, in the moment, key aspects of meta-cognition and attention—two well-known areas of "divergence" in students who have diagnoses that align with a neurodivergent identity. In earlier research, educational tools and interventions have often relied on a "one size fits all" approach [4], assuming that strategies effective for neurotypical learners will work equally well for neurodivergent students, or that a single strategy can effectively serve all neurodivergent learners, regardless of their specific diagnoses. However, this assumption can disadvantage neurodivergent learners by failing to account for their individual differences and needs [42]. Recent research highlights the need for inclusive educational approaches that better accommodate and provide specifically for, neurodiversity in higher education [26, 43]. To better support these students, adaptive models must be developed that not only distinguish between neurotypical and neurodivergent learners but also address the specific needs of different neurodivergent diagnoses. Therefore, another effective way to support these individuals is through adaptive, tailored learning models that respond to the distinct cognitive patterns and attentional challenges of neurodivergent learners.

Webcam-based eye tracking. One solution, which we explore in the current paper, is using webcam-based eye tracking. The theoretical foundation for using eye tracking to study attention stems from the concept of the "eye-mind link", which suggests a strong connection between where a person looks and where their attention is focused [19, 34, 55]. This link suggests that eye movements can serve as a window into cognitive processes, reflecting what information individuals are attending to and how they are processing it. The shift in attention can have detrimental effects on learning, as it disrupts the processing of information and reduces task performance [24, 34].

Traditional eye trackers typically rely on specialized infrared cameras and software, which can be expensive and require controlled laboratory settings for optimal accuracy. This limits their applicability in educational research and practice. Webcam-based eye tracking offers a more scalable solution, using the webcams readily available in most personal computers and laptops to track eye movements [34, 52, 68].

One webcam-based eye tracking software, WebGazer, has been used by researchers to record gaze [34, 52, 68] in online settings. This software uses computer vision techniques to estimate gaze location from webcam images. Although less accurate than traditional infrared-based eye tracking methods, such as PCCR (~4 degrees of error vs. ~0.05 degrees, respectively), WebGazer has been successfully used to collect gaze data suitable for cognitive modeling and effectively detect task-unrelated thoughts (TUT) [34, 68].While the lower sampling rate (typically less than 30Hz vs. 500Hz+ in research grade tracking [35]) can limit the features that are able to be calculated from this data [35], research has shown that the data is still suitable for modeling and analytics and can still provide valuable insights that can benefit the learning process [34, 35]. This scalable, cost-effective solution provides the opportunity for both research and applications that leverage the eye-mind link without the need for expensive traditional eye-tracking devices [34, 68].

Webcam-based eye tracking to model attention and comprehension. Prior work has leveraged this technology to identify specific gaze patterns associated with TUT. This includes increased off-screen gaze and reduced fixations within the text, both implying decreased engagement with the content. Further, students who are experiencing TUT have had longer saccades, potentially reflecting difficulty maintaining focus on the text [19, 34, 41]. Similarly, work has used webcam-based eye tracking to study comprehension. Analysis of gaze patterns during reading, has been able to replicate lab findings with features such as fixation duration and regressions (returning to text previously read) indicating how well a reader understands and integrates information from the text [1, 34, 49].

Neurodivergent learners, including individuals with ADHD, autism, and other learning disabilities, may exhibit distinct attentional and comprehension profiles compared to neurotypical individuals. Webcam-based eye tracking offers a valuable tool for investigating these differences and tailoring educational interventions to better support their needs. Understanding the unique gaze patterns of neurodivergent learners can inform not only our understanding of learning, but also the development of personalized learning technologies that adapt to their individual needs. Webcam based eye tracking can provide valuable feedback to both learners and educators, fostering greater self-awareness and facilitating more effective teaching strategies. For learners, it can offer insights into their own attentional patterns and comprehension processes. For educators, it can provide a deeper understanding of their students' individual learning styles and needs, enabling them to tailor their instruction and interventions accordingly [68]. Prior studies [31, 45] show that gaze-based TUT detection systems can enhance learning, but this has not yet been applied to neurodivergent learners, despite the clear connection to their symptomatology.

Webcam-based eye tracking to study neurodivergent learners. Of particular relevance to the current work is a prior study that used a webcam-based eye tracker with neurodivergent learners [68], focusing on students with ADHD, autism, or learning disabilities. This work used webcam-based eye tracking to identify different trends in gaze patterns, specifically for neurodivergent learners. While this work did not compare these patterns to neurotypical learners directly, there is extensive work considering the gaze patterns of neurotypical (or assumed/majority neurotypical) populations. This work highlights the potential for real-time TUT detection specifically for neurodivergent learners, which could, in turn, then lead to more adaptive learning systems that specifically support neurodivergent learners.

1.2 Current Study and Novelty

The current study leverages webcam-based eye tracking to model attentional changes (TUT, probes) and comprehension in both neurotypical and neurodivergent learners. By employing WebGazer, a scalable and accessible eye-tracking tool, we capture real-time gaze data during online reading tasks, allowing us to study participants in ecologically valid environments, such as their homes. This naturalistic setting provides more authentic insights into learner behavior compared to traditional lab-based studies.

A key contribution of this work is providing further evidence that webcam-based eye tracking can be used effectively for cognitive modeling. While earlier work identified gaze trends in neurodivergent populations, in this work, we use supervised machine learning to train detectors that can be used in real-time. We then consider more deeply if "one size fits all," i.e., can one model work for both populations? We investigate how models trained on the entire population differ from those trained separately on neurotypical and neurodivergent learners, comparing both model performance and predictive features. Our analysis probes deeper into how separating learners by diagnosis influences model behavior, offering insights into when a more tailored, diagnosis-specific model may outperform a broader approach.

Ultimately, this study contributes to the growing body of literature on webcam-based cognitive modeling and highlights the potential for scalable, personalized learning analytics. The findings support the argument that bespoke models—those tailored to specific learner profiles—provide more accurate and actionable insights than generalized approaches, particularly when applied to diverse populations.

2 Data Collection

2.1 Participants

A total of 354 learners were recruited for an online study delivered via Prolific [51], a high-quality online data collection platform. Prolific has gained popularity in recent years, with research indicating that it (and other similar platforms) provides data quality comparable to lab-based studies across various tasks [62].

Learners self-reported their neurodivergent and neurotypical status as well as the specific diagnoses within neurodivergence with the option to select multiple diagnoses. The data collected included 176 neurodivergent learners and 178 neurotypical learners. Of the neurodivergent learners, 75 reported being diagnosed with ADHD (Attention Deficit Hyperactivity Disorder) or ADD (Attention Deficit Disorder), 67 reported an Autism, Asperger's or Autism Spectrum Disorder (ASD) diagnosis, 14 reported Dyslexia, Dyspraxia, Dyscalculia, or Dysgraphia, grouped here as Specific Learning Disability (see further discussion in limitations) 13 reported another (undefined) language, reading, math, or non-verbal learning disorders, 90 reported Generalized Anxiety Disorder, 38 reported that they had other diagnoses not listed in the options. Furthermore, 34 neurodivergent learners didn't disclose specific diagnoses.

The learners' gender distribution was 182 males, 150 females, 20 identifying as Other/Non-Binary, and 2 who preferred not to disclose their gender. Additionally, learners provided information about their race with the distribution as follows: 1 person identified as Black or African, 40 as Hispanic or Latina/Latino/Latinx, 5 as Native American or Alaskan Native or First Nations, 1 as Native Hawaiian or other Pacific Islander, 4 as Middle Eastern or North African, 19 as East Asian, 9 as Southeast Asian, 4 as South Asian, 3 as Other Asian, 251 as White or Caucasian, 27 as Another race or ethnicity, and 2 who preferred not to respond. Some learners identified as multiple races. They also provided their ages, which ranged from 18 to 84, with an average age of 37 years.

2.2 WebGazer

WebGazer is a real-time eye tracking library built entirely in JavaScript, designed for integration into any web browser that supports webcam access [23]. The library utilizes various facial feature detection tools, including clmtrackr, js-objectdetect, and tracking.js [52], to identify facial and eye regions within the webcam feed. After detecting the eyes, WebGazer identifies the pupil by its darker contrast relative to the iris and extracts additional eye features by converting the eye region into a grayscale image patch. This image is then processed into a 120-dimensional feature vector, which is mapped to screen coordinates using ridge regression models that are continuously updated as the user interacts with the webpage. WebGazer's sampling rate varies based on the performance of the browser and webcam [22]. Importantly, WebGazer only reports x/y coordinates and does not record or store any video data, ensuring privacy. Previous research has shown that it can adapt through user interactions and offers sufficient accuracy for approximating a user's gaze [52].

2.3 Study Design and Procedure

Prior to any reading, calibration (using WebGazer native methods) was conducted for each learner to ensure that the eye-tracking data accurately aligned with on-screen content. Learners were instructed to ensure that their study environment was well-lit to facilitate accurate eye-tracking data collection. Learners then read a text about the psychological mechanisms influencing consumer behavior. The text consisted of 40 paragraphs, averaging 46 words in length, and was presented one paragraph at a time. Reading was self-paced with each learner spending an average of 11.55 seconds on each paragraph. As learners read, their eye movements were recorded using WebGazer [52].

Throughout the reading task, learners were intermittently prompted with thought probes designed to assess their current attentional state, with a total of seven prompts presented. When learners reported that their thoughts were unrelated to the task, this was labeled as TUT, data collected from other prompts was not used in this study. Thought probes are the gold standard for assessing TUT due to their highly internal nature, and have been used in a number of previous studies on these topics [48, 67, 68].

Upon completing the reading task, learners responded to 10 multiple-choice comprehension questions. These questions required learners to actively engage with and comprehend the content, as many of the correct answers were not explicitly stated in the text; instead, they required deeper comprehension.

2.4 Feature Engineering

For each learner, gaze data was mapped to appropriate paragraphs of text. We used the fixation approximation described by Hutt et al. [35] to compute fixation approximations from the gaze data, also relative to the text. In parallel, we applied natural language processing (NLP) to extract key linguistic features, using TextBlob for sentiment detection, classifying text as positive, negative, or neutral, textstat for syllable count, the Flesch Reading Ease score for readability, and Natural Language Toolkit (NLTK) for part-of-speech tagging [15, 37, 57].

Both eye gaze and text features were calculated for each paragraph. Gaze features included eye movements for both onscreen and offscreen gaze. We also leveraged eye-tracking data to calculate fixation features, including number of fixations and statistics on fixation duration (mean, standard deviation, and skew), which have been previously shown to relate to attention within a variety of tasks [9, 19, 30, 33, 46]. NLP features included word count and syllable count per paragraph, counts for different parts of speech, and sentiment analysis to determine whether the text was positive, negative, or neutral (each a binary feature).

2.5 Data processing

In this research, we focused on two primary objectives: developing models to predict (1) task unrelated thought (TUT) and (2) comprehension in learners.

Following initial feature development, we then examined the data quality and began a data cleaning process, to ensure that errors were not included in the machine learning process. We excluded all paragraphs with a low number of fixations, as it was unclear whether this was caused by tracker errors or represented genuine data—this threshold was set relative to the average time spent on the paragraph, as well as the number of words in the paragraph, and existing literature surrounding fixations in reading [45, 47, 54]. This resulted in 16 learners being completely removed from the data (ostensibly due to tracker failure) and 2 learners having a portion of their paragraphs removed from the data (36 paragraphs removed for one learner, 1 paragraph removed for the other). The final dataset contained 338 learners. We also standardized the data by instance level using z-score normalization to transform the values into the same scale [36].

3 Predicting Task Unrelated Thought (TUT)

3.1 Method

Our goal was to develop predictive models that could identify moments of TUT that could ultimately be used to inform targeted interventions aimed at helping these learners maintain focus in educational settings. We utilized three supervised machine learning algorithms—Random Forest (RF), Support Vector Machine (SVM) and XGBoost (XGB).

For the TUT data, we filtered the instances to only contain data from the paragraphs just before the probe for each learner, reducing the instances to 2,359 while the learner number was reduced to 337 (one participant had no paragraphs with a probe remaining after data cleaning). We then further removed 381 instances with missing values, resulting in 317 learners and 1,978 instances. We utilized features from the three paragraphs preceding each probe resulting in a total of 90 features.

We used learner-level k-fold cross-validation (k=5) to support future generalization and reduce the risk of overfitting. This approach ensured that all data from each learner stayed within the same fold, preventing overfitting to individual learners. Each model was fine-tuned to optimize performance using a grid search method (on the training set only) and we applied principal component analysis (again on the training set only) as a form of feature selection retaining components that explains at least 95% of the variance in the data. The dataset used for modeling was imbalanced (see Table 2 and Table 3). To address this, we applied the Synthetic Minority Oversampling Technique (SMOTE) [14] to generate synthetic instances and balance the training set, the test set was left unaltered.

The dataset was split into three feature combinations: gaze + fixation, NLP, and gaze + fixation + NLP features. These combinations were used to analyze how different aspects of learners' interactions with the text influenced their tendency to have TUT. Table 1 outlines the different feature combinations, their description and number of features in each. Additional information about

Table 1: Feature combinations implemented for TUT with descriptions and number of features (full feature lists available at the repository)

Feature combination	Description	Number of features
Gaze + Fixation	Features pertaining to learners' eye movements, including gaze patterns and fixation points	27
NLP	Natural language processing features derived from text read such as parts of speech, word and syllable frequency, readability, and sentiment analysis.	63
Gaze + Fixation + NLP	Combination of gaze, fixation and NLP features	90

the specific features used, as well as further details about the study, can be found in the online repository[1].

In evaluating our models, we focused on two metrics: Kappa, and Area Under the Receiver Operating Characteristic Curve (AUROC). These metrics both correct for chance, which was critical given the varying base rates in our data. Kappa, ranging from -1 to 1, measures the agreement between predicted and actual outcomes accounting for imbalanced datasets and correcting for chance, where 0 indicates chance performance. AUROC, a combination of precision and recall, also indicates how well the model distinguishes between classes, with a range from 0 to 1, where 0.5 represents chance performance [41]. These metrics are also common in the literature and allow easier comparison [41].

After running each model on the datasets, we selected the best feature combination for each group by assigning weights to each evaluation metric, with AUROC receiving the highest weight. Composite scores for each model were calculated by multiplying each metric score by its respective weight and summing the results. The model with the highest composite score was selected. Below is a breakdown of the model selection process in algorithmic form:

Algorithm 1 Algorithm for best model selection

Weights = {'Weighted F1': 0.23, 'Precision_1': 0.23, 'Recall_1': 0.23, 'AUROC': 0.31}
Composite score = $\sum$ (metrics * weights) for each model
Return model that has the highest composite score

The weights were selected to balance different aspects of model performance with AUROC receiving the highest weight due to it being comparable to chance with varying base rates. Assigning these weights ensured a comprehensive evaluation of the model's performance.

To understand the key features influencing model predictions, we calculated SHapley Additive exPlanations (SHAP) [44]. This granular analysis was crucial for understanding which aspects of the data—such as gaze patterns, fixation points, or text characteristics—most influenced the detector predictions of TUT, and if prediction profiles varied by diagnosis.

[1]https://github.com/HumanCenteredAI/One-size-does-not-fit-all-supplementary-material.git

3.2 Results

The performance of the machine learning models predicting TUT was assessed across three learner groups: all learners (noted below as "All"), neurotypical learners, and neurodivergent learners. The base rate for TUT varied across learner groups. In the dataset containing all learners, the base rate was 0.33, with neurotypical learners showing a lower base rate of 0.26 compared to neurodivergent learners at 0.39. Further segmentation within the neurodivergent group revealed even more variability, with some subgroups exhibiting base rates ranging from 0.27 to 0.56.

Across the different learner groups—all, neurotypical, and neurodivergent—the models showed varying levels of performance (Table 2). The neurodivergent group had better prediction of TUT (kappa of .1 vs .08 and .04). This finding is consistent with studies that have shown neurodivergent learners often exhibit distinct cognitive patterns, showing strength in areas like verbal reasoning [63, 66]. In contrast, the neurotypical group had lower prediction accuracy, indicating the presence of greater variability in indicators of TUT, which is consistent with research highlighting the broader range of cognitive behaviors within neurotypical populations [23].

To further explore variability within the neurodivergent population, models were trained and tested for each self-reported diagnosis. The results show high variability in model performance across the neurodivergent groups. Critically, all models outperformed the model that grouped all neurodivergent learners together (kappa = .1). Learners with Other Learning Disabilities exhibited the highest predictive accuracy, suggesting more consistent cognitive patterns within this group. Similarly, learners with specific learning disorders also showed strong model performance, reinforcing the idea that specific learning disabilities may present more predictable cognitive profiles. This performance aligns with prior studies using research-grade eye trackers, which have reported kappa values ranging from 0.15-0.22 for TUT detection [7, 11]. These findings highlight the feasibility of using webcam-based eye tracking for tasks including prediction TUT despite its inherent limitations.

Models trained and tested for learners with ADD/ADHD demonstrated moderate predictive performance, while those in the "Other" group showed lower accuracy—perhaps due to greater variability within this group. The "No Response/Never Diagnosed" group had the poorest model performance, underscoring the challenges of applying generalized models to populations lacking clear diagnostic categories.

These results highlight the varying effectiveness of TUT prediction models across different diagnoses. The findings suggest

Table 2: Comparison of TUT results across all learners, neurotypical and neurodivergent groups

Learner Group	#	Feature combination	Model	Base rate	Kappa	AUROC
All	317	NLP	RF	0.33	0.08	0.57
Neurotypical	161	NLP	XGB	0.26	0.04	0.52
Neurodivergent	156	NLP	XGB	0.39	0.10	0.57

Table 3: Comparison of TUT results across neurodivergent diagnoses

Group	#	Feature combination	Model	Base rate	Kappa	AUROC
ADD/ADHD	68	Gaze + fixation + NLP	SVM	.50	.15	.61
Autism/Asperger's/ASD	60	NLP	RF	.45	.12	.59
Specific Learning Disabilities	13	Gaze + Fixation + NLP	SVM	.56	.30	.65
Other language/reading/math/non-verbal learning disorders	13	NLP	RF	.43	.35	.70
Generalized Anxiety Disorder	81	NLP	XGB	.43	.11	.59
Other	32	NLP	XGB	.43	.12	.62
No response/ never diagnosed	30	NLP	XGB	.27	.04	.48

TUT prediction is better when considering individual diagnoses, rather than a "one size fits all" model, or even a model that just considers neurodivergent learners. Conversely, groups without clear or consistent diagnoses pose a challenge for accurate prediction, emphasizing the need for more tailored modeling approaches in these cases.

The SHAP analysis (Figure 1, shows a subset of this analysis) indicated that the skew of fixation duration was a significant predictor of task unrelated thoughts (TUT) in both the all learner group and neurodivergent group, indicating a skew in fixation duration distribution towards shorter fixations to be predictive of TUT. Offscreen behavior, including both offscreen gaze count and proportion of gaze offscreen, also played a crucial role across all learner groups, though its impact varied. In the general and neurodivergent groups, both low and high values of these features were important in predicting TUT, indicating that even minimal or frequent shifts in attention away from the screen can contribute to TUT. For the Specific Learning disabilities group, higher values of offscreen behavior were more impactful, suggesting that this group is particularly affected by significant shifts in attention.

We note that number of gaze points was another critical factor, especially in the neurodivergent and subset with specific learning disabilities. While the exact effects varied between the groups, this finding reinforces the eye-mind link, and the variability further reinforces earlier findings that a more personalized modeling approach is required.

The plots shown in Figure 1 each represent a subset of the previous, All learners -> Neurodivergent learners -> learners with Specific learning disabilities. We note that by reducing the variability in the groups, the order of important variables and their effects fluctuates. This effect was consistent across other subsets of the data and supports the model prediction results in that not only is there a stronger signal by subsetting, but the variables that contain that signal are also changing. It is possible that some of the indicators for a subset are being "washed out" when included in the broader dataset.

3.3 Discussion

This study modelled attention in both neurotypical and neurodivergent learners by predicting task unrelated thought (TUT) using gaze, fixation, and Natural Language Processing (NLP) features. The models showed varying performance across different learner groups, with the neurodivergent group demonstrating more accurate predictions of TUT, though this difference was minimal. This reinforces earlier findings that webcam based gaze tracking can be used for TUT detection [34] but that the lower data quality may impact the predictive power. We also build upon work by Wong et al. [68] and show that we can train a detector of TUT specifically for a neurodivergent population.

Despite the slightly better performance for neurodivergent learners, we note that we can improve performance further by considering further subsetting data by diagnosis. This suggests that even the blanket term of "Neurodivergent" may be overlooking individual differences. These findings indicate that bespoke models, specifically tailored to capture the varied patterns of attention and cognition, were more effective in accounting for individual differences and improving prediction accuracy within the neurodivergent group. The findings underscore the importance of considering individual identity in the development of predictive models.

4 Predicting Comprehension

In addition to modeling learners' attention, we also investigated comprehension. In contrast to TUT, comprehension presents a more robust ground truth measurement that is typically less noisy [2, 34, 35].

4.1 Method

Comprehension questions were graded as either correct or incorrect. These answers were then mapped to data from the relevant

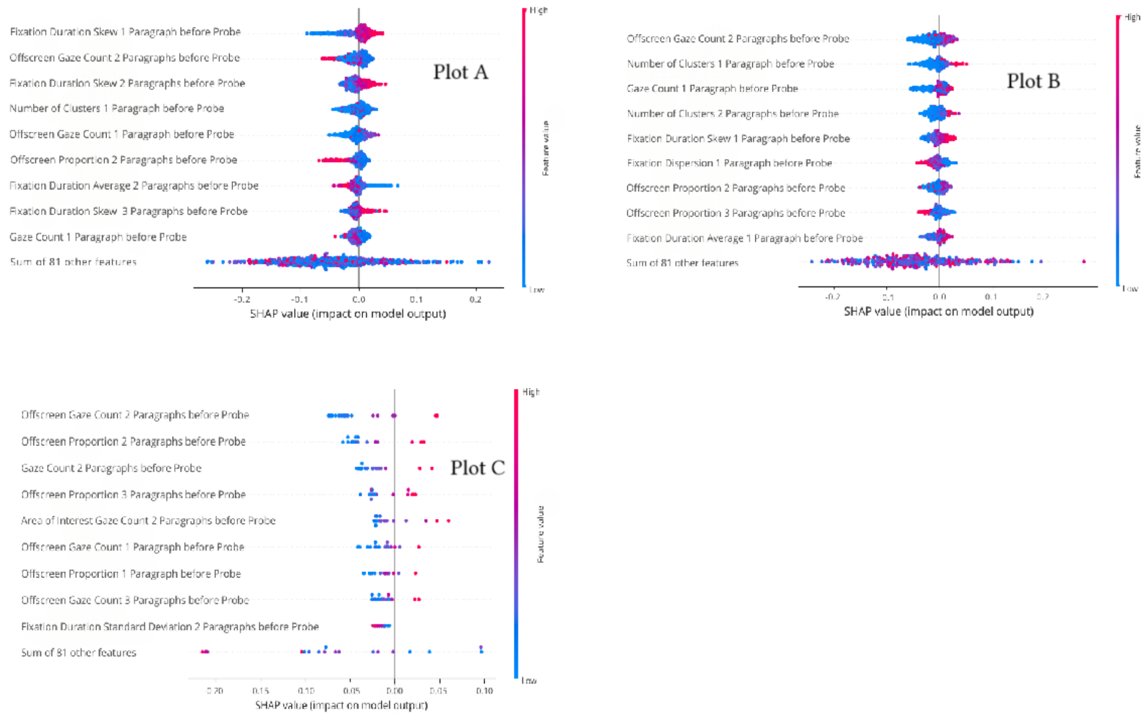

Figure 1: SHAP Plots for TUT prediction using gaze, fixation, and NLP features across all learners (Plot A), neurodivergent learners (Plot B), and learners with specific learning disabilities (Plot C). Each point represents an individual prediction, with the position on the x-axis showing the SHAP value and the color indicating the feature value (e.g., low to high). Features are listed on the y-axis in order of their importance. Feature values increase the likelihood of TUT occurring for red points, while for blue points, feature values decrease that likelihood.

paragraphs in the text that contained the corresponding information. In many cases, multiple paragraphs were mapped to the same question, as the answers were spread across more than one paragraph. The machine learning procedures was identical to those used in section 3 with a few exceptions, documented below.

For the comprehension data, we initially had data from 338 learners, resulting in 6,746 paragraphs (once allowing for multiple paragraphs per question). We removed 437 instances with missing values, resulting in 6,309 instances remaining, with the number of learners still at 338.

In contrast to TUT modeling, we only considered features from the paragraph in question (ignoring the preceding paragraphs). We also included a new feature related to the question, not the reading, which was question response time. This was included to align with prior work [5, 34] and provide an alternate measurement that helps to identify what the predictive power of gaze/NLP features is, when compared to a simple question measurement. Table 4 describes each feature combination in more detail. The dataset containing all learners had a base rate of 0.79 (i.e., 79% of questions answered correctly), consistent across both neurotypical and neurodivergent groups. However, variability was observed within neurodivergent subgroups, with base rates ranging from 0.73 to 0.84.

4.2 Results

The results, as outlined in Table 5, demonstrate that the model performed best when trained and tested on neurotypical learners, showing slightly better accuracy within this group. The model trained on all learners also showed a reasonable level of performance, though slightly lower than the neurotypical group. In contrast, the model trained on neurodivergent learners exhibited reduced accuracy. The neurotypical group's comprehension patterns were more effectively captured by the model compared to those of neurodivergent group.

The results in Table 6 show variability in model performance across different neurodivergent diagnoses. The Specific Learning disabilities group performed the best, as indicated by higher Kappa and AUROC values. Other groups, including those with ADD/ADHD, Autism/Asperger's/ASD, and Generalized Anxiety Disorder, exhibited lower accuracy, with lower Kappa and AUROC scores. The "Other" and "No response/never diagnosed" groups showed intermediate model performance, with Kappa and AUROC values that were between the highest and lowest performing groups. While all results are above chance, these results indicate differences in how well the model captured patterns across the various neurodivergent groups. We note that in the majority of cases, separating by diagnosis resulted in equal or better performance than the model trained and tested on the entire neurodivergent group.

Table 4: Feature combinations implemented for comprehension with descriptions and number of features (full feature lists available at the online repository)

Feature combination	Description	Number of features
Gaze + Fixation	Features pertaining to learners' eye movements, including gaze patterns and fixation points.	9
NLP	Natural language processing features derived from text read such as Parts of Speech, word and syllable frequency, readability, and sentiment analysis.	25
Gaze + Fixation + NLP	Combination of gaze, fixation and NLP features.	34
Gaze + Fixation + Question Response Time	Combination of gaze, fixation and the time spent on each question per learner.	10
Gaze + Fixation + Question Response Time + NLP	Combination of gaze, fixation, question response time and NLP.	35

Table 5: Comparison of comprehension results across all learners, neurotypical, and neurodivergent groups

Learner Group	#	Feature combination	Model	Base rate	Kappa	AUROC
All	338	NLP	XGB	0.79	0.27	0.71
Neurotypical	171	NLP	XGB	0.79	0.29	0.71
Neurodivergent	167	Gaze+Fixation+RT+NLP	RF	0.79	0.19	0.68

Table 6: Comparison of comprehension results across neurodivergent diagnoses

Group	#	Feature combination	Model	Base rate	Kappa	AUROC
ADD/ADHD	70	Gaze+Fixation+RT+NLP	RF	0.79	0.19	0.69
Autism/Asperger's/ASD	61	Gaze+Fixation+RT+NLP	RF	0.80	0.19	0.68
Specific Learning Disabilities	13	Gaze + Fixation + NLP	RF	0.84	0.35	0.79
Other language/reading/math/non-verbal learning disabilities	13	NLP	RF	0.73	0.17	0.68
Generalized Anxiety Disorder	84	Gaze+Fixation+RT+NLP	RF	0.81	0.18	0.69
Other	37	Gaze+Fixation+RT+NLP	RF	0.80	0.25	0.74
No response/ never diagnosed	33	NLP	RF	0.76	0.25	0.68

We again performed SHAPley analysis, investigating how the subsetting of data impacts the most predictive features Figure 2 shows an example, using the same sets of data as used in section 3 – All learners -> Neurodivergent learners -> learners with Specific learning disabilities. Simpler sentence structures and ease of reading were both consistently associated with better comprehension, highlighting the universal importance of text simplicity. Question Response Time was a critical predictor in all learner and neurodivergent groups, in which quicker responses suggest better comprehension, but this feature was less emphasized in the Specific learning disabilities group. Gaze-related metrics were crucial across all groups, with focused and consistent visual attention linked to better comprehension, though the specific features that were most influential vary. In all learner and neurodivergent groups, a combination of focused gazes and lower variability was beneficial to comprehension, whereas in the Specific Learning disabilities group, reducing offscreen distractions are important to comprehension. Overall, while there were commonalities in predictors of comprehension, including the importance of simple text and focused attention, the specific impacts of these features differed as the dataset becomes more refined, reflecting the unique identity of each group and again showing that where possible, bespoke models pick up on signatures of comprehension that may be unique to each group.

4.3 Discussion

In this analysis, we developed predictive models for comprehension using gaze, fixation, NLP and Question response time features. The results showed that the model trained with neurotypical learners performed better than the one trained on neurodivergent learners. However, again noting that neurodivergent is a very broad term, we examined modeling performance by diagnosis. There was notable variability in model performance across different neurodivergent diagnoses, with learners diagnosed with Specific Learning Disabilities having the highest prediction score. This supports findings from

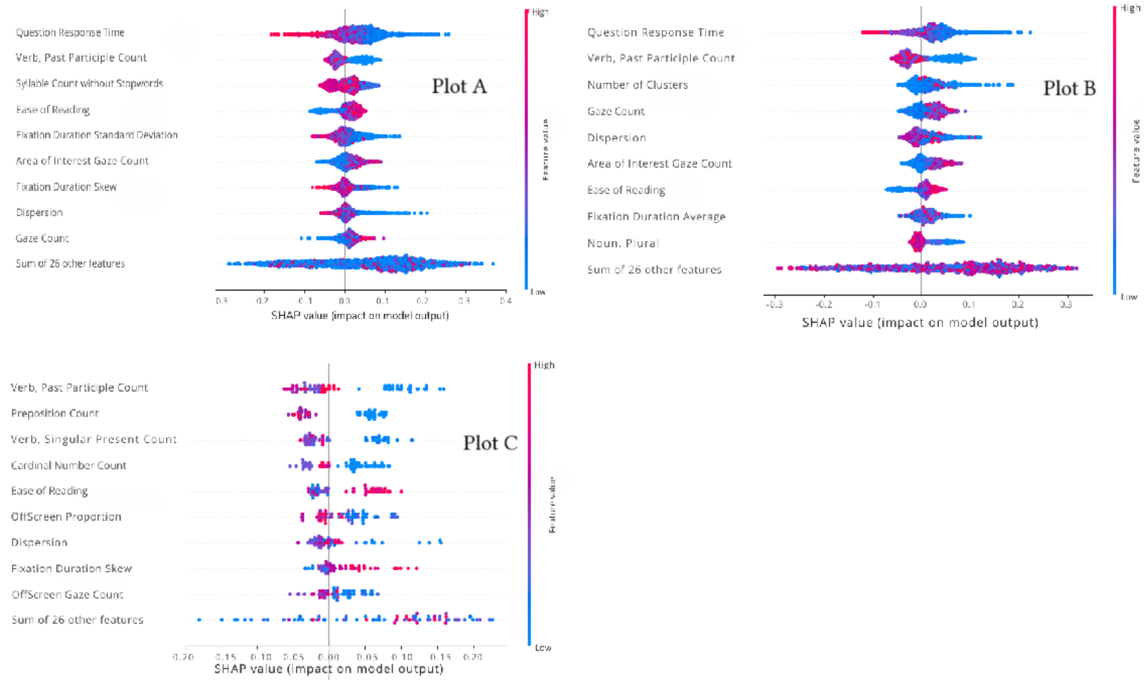

Figure 2: SHAP plot for comprehension prediction using gaze, fixation, and NLP features across all learners (Plot A), neurodivergent learners (Plot B), and learners with specific learning disabilities (Plot C)

Study 1 that indicate a one size fits all approach is not appropriate for neurodivergent learners.

The findings indicate that neurotypical learners' comprehension can be predicted more consistently. In contrast, we note variability across neurodivergent learners. For instance, learners with Specific learning disabilities showed higher prediction scores than when in the larger group, suggesting that there may be indicators of comprehension that are valid for that group, but not valid for other diagnoses.

These results indicate that for future personalized interventions to be effective, truly personalized detection must first be implemented so that when an intervention is delivered, we can be confident in the detector's accuracy. This study adds to the body of research on how individual differences influence comprehension and highlights that general detection approaches may have a negative impact on neurodivergent learners.

5 General Discussion

Eye gaze has consistently been a research tool in psychology that can be harnessed to better understand how people think and learn [29, 46, 53]. This paper has investigated the potential of scalable webcam-based eye tracking for cognitive modeling, specifically in detecting task unrelated thought and comprehension across neurotypical and neurodivergent learners. By conducting the study in an ecologically valid, online setting, we gathered data that more accurately reflects real-world learning environments, such as learners' homes. Through comparison across individual differences, we examine the variability in what is predictive of both constructs and consider if a general model is suitable for all learners.

5.1 Main Findings

In this research, we developed predictive models for TUT and comprehension, utilizing gaze, fixation, NLP, and question response time features, captured using a webcam-based eye tracker, WebGazer. Our work supports earlier findings that webcam-based tracking is "good enough" for cognitive modeling at above-chance levels and with accuracy that may be suitable to drive future interventions [9, 10, 46].

Our models showed varying levels of success across neurotypical and neurodivergent groups, as well as across different neurodivergent diagnoses. All models consistently performed above chance, and our TUT detection results are comparable to earlier work using webcam-based eye tracking, for example [34], which had the highest reported kappa value of 0.15 for gaze-based TUT detection. Similarly, our findings are roughly comparable to others using research-grade eye-tracking [8] or EEG signals [21]. Comprehension detector findings show greater variability than previously observed in the literature [34]. Where prior work had investigated gaze patterns for neurodivergent learners, we specifically trained detectors and made comparisons, with the view to examining how a general model performs, relative to more bespoke models, trained on data subsets. Our findings further challenge the efficacy of a "one size fits all" model common in this work [41]. Considering both predictive performance and feature analysis, we see distinct differences in how neurotypical and neurodivergent learners interact with text. These results underscore the need for tailored cognitive models and subsequent analytics that can accommodate diverse learning profiles, paving the way for more personalized educational interventions that go beyond generalized approaches.

5.2 Applications

It is first critical to state that any kind of application of this work should celebrate neurodivergent learners. The separation of bespoke models runs the risk of working against a movement for more inclusive education environments, which we support [27, 42]. Indeed, an adaptive perspective of neurodivergence is supported by the Theory of Complementary Cognition [60], which highlights the value of diverse cognitive approaches for human societies. Similarly, other arguments highlight how neurocognitive variations can act as unique tools within human groups, enabling communities to persist and adapt to challenges [13], as well as highlight how unique, neurodivergent thinking can provide innovation in engineering [17] and other areas [58]. With this in mind, it is critical that bespoke modelling be leveraged to *support* neurodivergent learners in a way that aids this inclusive goal, and not for further segregation.

One benefit of real time analytics (and thus potential application of this work) is real-time intervention [32, 46]. By highlighting the differences in indicators of both TUT and comprehension prediction across different groups, the findings underscore the need for tailored approaches to support learners. While intervention techniques themselves may or may not vary, by providing more bespoke detection, we are able to ensure that students are more likely to receive an intervention when they need it, and not when they don't. Any such application of this work must, of course, be very clear with users as to what data is being collected and how it is being used. This is especially critical if usage moves out of controlled environments and into spaces such as homes, and study spaces. Effective communication that is appropriate for the target population will be critical to the success of any such application.

5.3 Limitations and Future Work

A significant limitation was the reliance on self-reported diagnoses. While we have no reason to believe that participants would lie, we cannot ignore this as a limitation. That said, access to medical diagnosis information does not seem feasible for a study of this kind. A strength of this work was the ability to collect a significant amount of data from neurodivergent populations, a population that is typically challenging to study. Future work may wish to validate these findings with medical diagnosis data, that will likely have to be in the lab, with a smaller sample.

Further, due to subsetting, we inevitably had smaller sample sizes for some diagnoses which may impact generalizability. Future work should consider a larger online study or a study that only considers neurodivergent learners, now that we have performed the initial comparison. While webcam-based eye gaze data performed well in predicting TUT and comprehension, there remain limitations of webcam-based eye tracking compared research-grade hardware. These limitations include reduced sampling rate [34] and increased vulnerability to environmental factors such as lighting [16]. Despite these constraints, webcam-based tracking remains a cost-effective and scalable option [34].

The results presented here follow the reporting convention of almost all TUT detection papers to date [41]. In this research (kappa and AUROC). For a more detailed analysis, future studies can combine this test with performance metrics. Future work will consider the gaze-based TUT detection systems' potential to enhance learning amongst neurodivergent learners, exploring how such tools can be adapted to address their unique challenges and strengths. Future research should also continue to explore the application of these predictive models in real-time educational environments to assess their effectiveness in reducing TUT and enhancing comprehension. By validating these models for real-time use, we open the door to future interventions and real-time analytics, providing support for learners who have typically been underrepresented in learning analytics and AI in education.

5.4 Conclusion

This study advances our understanding of modeling and analytics for diverse learner populations. By leveraging webcam-based eye tracking in an ecologically valid, online context. We provided evidence that this scalable technology can effectively capture and differentiate cognitive processes such as task unrelated thought and comprehension across neurotypical and neurodivergent learners. Our findings highlight the limitations of a "one size fits all" approach, demonstrating that models tailored to specific learner profiles yield more accurate predictions and deeper insights into individual learning patterns. By showing the impact of diagnosis on gaze-based cognitive modeling, we pave the way for more personalized and inclusive learning analytics.

References

[1] Ahn, S. *et al.* 2020. Towards Predicting Reading Comprehension From Gaze Behavior. *ACM Symposium on Eye Tracking Research and Applications* (New York, NY, USA, Jun. 2020), 1–5.
[2] Ahn, S. *et al.* 2020. Towards predicting reading comprehension from gaze behavior. *ACM Symposium on Eye Tracking Research and Applications* (2020), 1–5.
[3] Alperin, B.R. *et al.* 2021. More than off-task: Increased freely-moving thought in ADHD. *Consciousness and Cognition.* 93, (2021), 103156. DOI:https://doi.org/10.1016/j.concog.2021.103156.
[4] Bayetto, A. 2002. All my students are reading the same book and they're successful: an inclusive teaching approach. *The Australian Journal of Language and Literacy.* (Oct. 2002).
[5] Beck, J.E. 2004. Using response times to model student disengagement. *Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments.* 20, (2004), 88–95.
[6] Bielecki, K. and Varvarides, L. 2019. What is the Neurodivergent Aesthetic? (Jul. 2019).
[7] Bixler, R. *et al.* 2015. Automatic Detection of Mind Wandering During Reading Using Gaze and Physiology. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (New York, NY, USA, Nov. 2015), 299–306.
[8] Bixler, R. *et al.* 2015. Automatic Detection of Mind Wandering During Reading Using Gaze and Physiology. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (New York, 2015), 299–306.
[9] Bixler, R. and D'Mello, S.K. 2015. Automatic Gaze-Based Detection of Mind Wandering with Metacognitive Awareness. *User Modeling, Adaptation and Personalization* (Dublin, Ireland, 2015), 31–43.
[10] Blanchard, N. *et al.* 2014. Automated Physiological Based Detection of Mind Wandering during Learning. *Intelligent Tutoring Systems* (Switzerland, 2014), 55–60.
[11] Blanchard, N. *et al.* 2014. Automated Physiological-Based Detection of Mind Wandering during Learning. *Intelligent Tutoring Systems* (Cham, 2014), 55–60.
[12] Brosnan, M. *et al.* 2016. Deficits in metacognitive monitoring in mathematics assessments in learners with autism spectrum disorder. *Autism: the international journal of research and practice.* 20, 4 (2016), 463–472. DOI:https://doi.org/10.1177/

1362361315589477.
[13] Chapman, R. 2021. Neurodiversity and the social ecology of mental functions. *Perspectives on Psychological Science*. 16, 6 (2021), 1360–1372. DOI:https://doi.org/10.1177/1745691620959833.
[14] Chawla, N.V. *et al.* 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. 16, 1 (Jun. 2002), 321–357. DOI:https://doi.org/10.1613/jair.953.
[15] Chiche, A. and Yitagesu, B. 2022. Part of speech tagging: a systematic review of deep learning and machine learning approaches. *Journal of Big Data*. 9, 1 (Jan. 2022), 10. DOI:https://doi.org/10.1186/s40537-022-00561-y.
[16] Ch'ng, Y.H. *et al.* 2021. An Innovation-Driven Approach to Specific Language Impairment Diagnosis. *The Malaysian Journal of Medical Sciences: MJMS*. 28, 2 (Apr. 2021), 161–170. DOI:https://doi.org/10.21315/mjms2021.28.2.15.
[17] Chrysochoou, M. *et al.* 2022. Reframing neurodiversity in engineering education. *Frontiers in Education*. 7, (Nov. 2022). DOI:https://doi.org/10.3389/feduc.2022.995865.
[18] Cook, S.C. *et al.* 2017. Co-teaching for students with disabilities: A critical and updated analysis of the empirical literature. *Handbook of special education*. Routledge. 233–248.
[19] D'Mello, S. *et al.* 2016. Attending to Attention: Detecting and Combating Mind Wandering during Computerized Reading. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, May 2016), 1661–1669.
[20] D'Mello, S.K. and Mills, C.S. 2021. Mind wandering during reading: An interdisciplinary and integrative review of psychological, computing, and intervention research and theory. *Language and Linguistics Compass*. 15, 4 (2021), e12412. DOI:https://doi.org/10.1111/lnc3.12412.
[21] Dong, H.W. *et al.* 2021. Detection of mind wandering using EEG: Within and across individuals. *Plos one*. 16, 5 (2021). DOI:https://doi.org/10.1371/journal.pone.0251490.
[22] Eye Tracking - jsPsych: https://www.jspsych.org/6.3/overview/eye-tracking/. Accessed: 2024-08-20.
[23] Fischer-Baum, S. *et al.* 2018. Individual Differences in the Neural and Cognitive Mechanisms of Single Word Reading. *Frontiers in Human Neuroscience*. 12, (Jul. 2018), 271. DOI:https://doi.org/10.3389/fnhum.2018.00271.
[24] Fishman, D.J. and Schooler, J. 2006. Mind-wandering and Educational Performance Psychonomic Bulletin and Review. (2006).
[25] Franklin, M.S. *et al.* 2017. Tracking distraction: The relationship between mind-wandering, meta-awareness, and ADHD symptomatology. *Journal of Attention Disorders*. 21, 6 (Apr. 2017), 475–486. DOI:https://doi.org/10.1177/1087054714543494.
[26] Hamilton, L.G. and Petty, S. 2023. Compassionate pedagogy for neurodiversity in higher education: A conceptual analysis. *Frontiers in Psychology*. 14, (Feb. 2023). DOI:https://doi.org/10.3389/fpsyg.2023.1093290.
[27] Hehir, T. *et al.* 2016. A summary of the evidence on inclusive education. *Abt Associates*. (2016).
[28] Horowitz, S.H. *et al.* 2017. *The state of learning disabilities: Understanding the 1 in 5*. Technical Report #3. National Center for Learning Disabilities.
[29] Huey, E.B. 1898. Preliminary Experiments in the Physiology and Psychology of Reading. *The American Journal of Psychology*. 9, 4 (1898), 575. DOI:https://doi.org/10.2307/1412192.
[30] Hutt, S. *et al.* 2019. Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction*. 29, 4 (Sep. 2019), 821–867. DOI:https://doi.org/10.1007/s11257-019-09228-5.
[31] Hutt, S. *et al.* 2021. Breaking out of the Lab: Mitigating Mind Wandering with Gaze-Based Attention-Aware Technology in Classrooms. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2021), 1–14.
[32] Hutt, S. *et al.* 2021. Breaking out of the Lab: Mitigating Mind Wandering with Gaze-Based Attention-Aware Technology in Classrooms. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery.
[33] Hutt, S. *et al.* 2017. Gaze-based Detection of Mind Wandering during Lecture Viewing. *Proceedings of the 10th International Conference on Educational Data Mining (EDM 2017)* (2017), 226–231.
[34] Hutt, S. *et al.* 2024. Webcam-based eye tracking to detect mind wandering and comprehension errors. *Behavior Research Methods*. 56, 1 (Jan. 2024), 1–17. DOI:https://doi.org/10.3758/s13428-022-02040-x.
[35] Hutt, S. and D'Mello, S.K. 2022. Evaluating Calibration-Free Webcam-Based Eye Tracking for Gaze-Based User Modeling. *Proceedings of the 2022 International Conference on Multimodal Interaction* (New York, NY, USA, 2022), 224–235.
[36] Javaheri, S.H. *et al.* 2014. Response modeling in direct marketing: A data mining-based approach for target selection. *Data Mining Applications with R*. Y. Zhao and Y. Cen, eds. Academic Press. 153–180.
[37] Kannappan, S. 2023. Sentiment Analysis using NLP and Machine Learning Techniques on Social Media Data. *Shu Ju Cai Ji Yu Chu Li/Journal of Data Acquisition and Processing*. 38, (Apr. 2023), 520–526. DOI:https://doi.org/10.5281/zenodo.7766376.
[38] Kelly, C. *et al.* 2024. Recognising and responding to physical and mental health issues in neurodivergent girls and women. *British Journal of Hospital Medicine*. 85, 4 (Apr. 2024), 1–12. DOI:https://doi.org/10.12968/hmed.2023.0337.
[39] Kletenik, D. *et al.* 2024. From Awareness to Action: Teaching Software Accessibility for Neurodiverse Users. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (New York, NY, USA, Mar. 2024), 687–693.
[40] Kuriyan, A.B. *et al.* 2013. Young Adult Educational and Vocational Outcomes of Children Diagnosed with ADHD. *Journal of abnormal child psychology*. 41, 1 (Jan. 2013), 27–41. DOI:https://doi.org/10.1007/s10802-012-9658-z.
[41] Kuvar, V. *et al.* 2023. Detecting When the Mind Wanders Off Task in Real-time: An Overview and Systematic Review. *Proceedings of the 25th International Conference on Multimodal Interaction* (New York, NY, USA, Oct. 2023), 163–173.
[42] Laiser, M.L. *et al.* 2024. Differentiated Assessments for Learners with Neurodevelopmental Disabilities in Inclusive Elementary Schools in Tanzania. *Journal of Issues and Practice in Education*. 16, (Jan. 2024). DOI:https://doi.org/10.61538/jipe.v16i.1426.
[43] Liebel, G. and Sigurðardóttir, S.G. 2023. Economical Accommodations for Neurodivergent Students in Software Engineering Education: Experiences from an Intervention in Four Undergraduate Courses. (2023). DOI:https://doi.org/10.48550/ARXIV.2306.07643.
[44] Lundberg, S. and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. arXiv.
[45] Mills, C. *et al.* 2021. Eye-Mind reader: an intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering. *Human–Computer Interaction*. 36, 4 (Jul. 2021), 306–332. DOI:https://doi.org/10.1080/07370024.2020.1716762.
[46] Mills, C. *et al.* 2020. Eye-Mind reader: an intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering. *Human-Computer Interaction*. 00, 00 (2020), 1–27. DOI:https://doi.org/10.1080/07370024.2020.1716762.
[47] Mills, C. *et al.* 2018. Is an off-task mind a freely-moving mind? Examining the relationship between different dimensions of thought. *Consciousness and Cognition*. 58, (Feb. 2018), 20–33. DOI:https://doi.org/10.1016/j.concog.2017.10.003.
[48] Mills, C. and D'Mello, S. 2015. *Toward a Real-Time (Day) Dreamcatcher: Sensor-Free Detection of Mind Wandering during Online Reading*. International Educational Data Mining Society.
[49] Navalpakkam, V. *et al.* 2011. Using gaze patterns to study and predict reading struggles due to distraction. *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, May 2011), 1705–1710.
[50] Neurodiversity - NCI: 2022. https://dceg.cancer.gov/about/diversity-inclusion/inclusivity-minute/2022/neurodiversity. Accessed: 2024-09-10.
[51] Palan, S. and Schitter, C. 2018. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*. 17, (Mar. 2018), 22–27. DOI:https://doi.org/10.1016/j.jbef.2017.12.004.
[52] Papoutsaki, A. *et al.* 2016. WebGazer: Scalable Webcam Eye Tracking Using User Interactions. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA, Jul. 2016), 3839–3845.
[53] Rayner, K. 2009. Eye movements and attention in reading, scene perception, and visual search.
[54] Rayner, K. *et al.* 2006. Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading - SCI STUD READ*. 10, (Jul. 2006). DOI:https://doi.org/10.1207/s1532799xssr1003_3.
[55] Reichle, E.D. *et al.* 2012. Using E-Z Reader to simulate eye movements in nonreading tasks: a unified framework for understanding the eye-mind link. *Psychological Review*. 119, 1 (Jan. 2012), 155–185. DOI:https://doi.org/10.1037/a0026473.
[56] Robb, J.A. *et al.* 2011. The Estimated Annual Cost of ADHD to the U.S. Education System. *School mental health*. 3, 3 (Sep. 2011), 169–177. DOI:https://doi.org/10.1007/s12310-011-9057-6.
[57] Santucci, V. *et al.* 2020. Automatic Classification of Text Complexity. *Applied Sciences*. 10, 20 (Jan. 2020), 7285. DOI:https://doi.org/10.3390/app10207285.
[58] Schippers, L.M. *et al.* 2022. A qualitative and quantitative study of self-reported positive characteristics of individuals with ADHD. *Frontiers in Psychiatry*. 13, (Oct. 2022). DOI:https://doi.org/10.3389/fpsyt.2022.922788.
[59] Smallwood, J. *et al.* 2007. Counting the cost of an absent mind: mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review*. 14, 2 (Apr. 2007), 230–236. DOI:https://doi.org/10.3758/BF03194057.
[60] Taylor, H. *et al.* 2022. The evolution of complementary cognition: humans cooperatively adapt and evolve through a system of collective cognitive search. *Cambridge Archaeological Journal*. 32, 1 (2022), 61–77. DOI:https://doi.org/10.1017/s0959774321000329.
[61] Tumlin, Z. 2019. "This Is a Quiet Library, Except When It's Not:" On the Lack of Neurodiversity Awareness in Librarianship. *Music Reference Services Quarterly*. 22, 1–2 (Apr. 2019), 3–17. DOI:https://doi.org/10.1080/10588167.2019.1575017.
[62] Uittenhove, K. *et al.* 2023. From Lab-Testing to Web-Testing in Cognitive Research: Who You Test is More Important than how You Test. *Journal of Cognition*. 6, (Jan. 2023). DOI:https://doi.org/10.5334/joc.259.

[63] Van Lambalgen, M. 2010. Logical Form as a Determinant of Cognitive Processes. (Berlin, Heidelberg, 2010), 59–83.
[64] Wang, K.D. *et al.* 2024. Exploring the learning experiences of neurodivergent college students in STEM courses. *Journal of Research in Special Educational Needs*. 24, 3 (Jul. 2024), 505–518. DOI:https://doi.org/10.1111/1471-3802.12650.
[65] Wang, R. *et al.* 2024. Usability of Responsible Gambling Information on Gambling Operator' Websites: A Webcam-Based Eye Tracking Study. *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*. (Jun. 2024), 1–3. DOI:https://doi.org/10.1145/3649902.3655662.
[66] Wilson, A.C. 2024. Cognitive Profile in Autism and ADHD: A Meta-Analysis of Performance on the WAIS-IV and WISC-V. *Archives of Clinical Neuropsychology*. 39, 4 (May 2024), 498–515. DOI:https://doi.org/10.1093/arclin/acad073.
[67] Wong, A.Y. *et al.* 2022. Task-unrelated thought during educational activities: A meta-analysis of its occurrence and relationship with learning. *Contemporary Educational Psychology*. 71, (Oct. 2022), 102098. DOI:https://doi.org/10.1016/j.cedpsych.2022.102098.
[68] Wong, A.Y. *et al.* 2023. Using a Webcam Based Eye-tracker to Understand Students' Thought Patterns and Reading Behaviors in Neurodivergent Classrooms. *LAK23: 13th International Learning Analytics and Knowledge Conference* (New York, NY, USA, Mar. 2023), 453–463.