Exploring the Viability of Using Eye Tracking to Detect Neurodivergent Learners' Implicit Learning in a Physics Game

[α]Ibrahim Dahlstrom-Hakki, [α]Elizabeth Rowe, [α]Jodi Asbell-Clarke, and [α]Mia Almeda

[α] EdGE at TERC

Corresponding author: idahlstromhakki@terc.edu

Abstract. With the prominence of assessments in education, there is an increasing need to create new forms of assessment that more accurately reflect the needs of the entire student population, particularly neurodivergent learners. To address this challenge, this paper explores the potential for using eye tracking data in a game-based learning environment to assess student's implicit knowledge. Data was collected from a sample of 66 neurodivergent college students playing the physics game Impulse while their eye movements and game play behaviors were recorded. The results indicate that gaze allocation patterns were predictive of students' physics knowledge and aligned with previously identified behavior indicators of learning. These findings provide evidence for further development of eye movement-based assessments in computer-based instruction and demonstrate how these data can be collected, organized, and analyzed.

Dahlstrom-Hakki, I., Rowe, E., Asbell-Clarke, J., Almeda, M. (2024) Exploring the Viability of Using Eye Tracking to Detect Neurodivergent Learners' Implicit Learning in a Physics Game. Computer-Based Learning in Context, 6 (1), 24-40.

Keywords: Game-based learning, assessment, neurodiverse, eye tracking, implicit knowledge.

1 Introduction

For better or worse, assessment has played a prominent role in many conversations related to computer-based learning. Whether one believes that there is too much or not enough assessment in education, most can agree that assessments need to accurately reflect learner knowledge and apply fairly to all students. This sentiment is reflected by the latest edition of The Standards for Educational and Psychological Testing put out jointly by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (2014). The latest edition of the standards highlights the need for fairness in testing and the importance of developing and using assessments that are equally applicable and interpretable across all student populations. Unfortunately, evidence shows that many commonly used assessments may not be fairly assessing some student populations including students with disabilities (Dahlstrom-Hakki & Alstad, 2019; Sideridis, 2016), English Language Learners (ELL) (Abedi, 2006), and students of color (Ford & Helms, 2012).

To address this challenge, researchers have been exploring alternate and potentially more objective means of assessing student knowledge including Educational Data Mining (EDM) & Learning Analytics (Berland et al., 2014), Game-Based Learning Assessment (GBLA) (Ke & Shute, 2015), and Project-Based Learning Assessment (PBLA) (Larmer et al., 2015; Van den Bergh et al., 2006). These approaches primarily focus on assessing knowledge by looking at a student's ability to make use of relevant information to perform a task rather than to explicitly express knowledge using text or symbolic notation. While many of these approaches show real promise in supporting different student populations, they are limited in their ability to reveal the cognitive processes underlying learning.

This paper explores the potential for using eye tracking as a means of more fairly assessing diverse populations of students. Eye tracking offers a robust means of assessing student's visual attention allocation and their cognitive effort thereby providing the potential to not only assess evidence of learning outcomes but to also provide data that can better reveal sources of struggle for students who are not achieving learning goals (Alemdag & Cagiltay, 2018; Dahlstrom-Hakki et al., 2019; Lai et al., 2013). Eye movements have long been used in the cognitive sciences as a means of inferring a variety of cognitive processes related to learning and memory (see Rayner, 2009 for a review). Eye tracking provides information on what an individual is attending to and for how long, thereby helping researchers ascertain the approach players are using to solve a task.

Cognitive researchers use eye movements to infer several elements of cognitive processing including visual attention allocation, cognitive processing difficulty, and short- or long-term memory use (Carter & Luke, 2020;

Dahlstrom-Hakki et al., 2019). Eye movements are generally analyzed by looking at both their temporal and spatial characteristics. These movements generally occur as a series of relatively stable periods where a location in the visual field is foveated, this is known as a fixation, interspersed with fast movements between fixation locations, these are known as saccades. Visual attention allocation is strongly associated with fixation allocation, and often (but not always) co-occur (Corbetta et al., 1998). Cognitive processing difficulty has been analyzed by looking at differences in patterns of fixation durations (Inhoff & Rayner, 1986; Meghanathan et al., 2015) and more recently by looking at changes in pupil dilation (Klingner, 2010).

Eye tracking measures of attention may be particularly important when considering the difference between implicit and explicit knowledge of neurodiverse students. For the purposes of this study, implicit knowledge is defined as knowledge that is evident through performance or actions but which the student may not be aware of or may not be able to express explicitly. Some research has found that students with disabilities may have implicit knowledge even when they are unable to demonstrate it on explicit assessments (McNamara & Wagner, 2001). Polanyi (1966) argued that implicit knowledge (also called tacit knowledge) is foundational and a required element of explicit learning. Implicit understandings are embodied and enacted through our interactions with the world around us but may not yet be formalized or expressed verbally or textually. Vygotsky (1978) used the term preparedness for learning to describe similar abilities and understandings a learner brings to a learning situation that can be scaffolded by a teacher, environment, and tools. More recently, the idea of implicit learning has been expanded to explain the science of successful learning (Brown et al., 2014) and fast and slow thinking (Kahneman, 2011).

Evidence indicates that many assessments of explicit knowledge do not accurately reflect the learning of neurodivergent learners (Nieminen, J. H., 2023). Given this simple fact, measures of implicit knowledge may offer a more accessible and overall valid measure for both neurodivergent learners and students at large. This work explores the viability of using implicit measures of knowledge as a more objective means of assessing the knowledge of neurodivergent learners.

Prior research by this project's team in game-based learning shows that games have unique promise for revealing implicit learning because they can a) encourage players to dwell in the phenomena and b) they leave a digital trail that reveals the patterns the players used in their learning process. Careful alignment of game mechanics with learning and assessment mechanics (Plass et al., 2015) may reveal implicit learning and empower teachers and learners to help bridge game-based knowledge to other forms of learning. The digital nature of these games also allows for the integration of eye tracking data, which can be synchronized with game data logs to allow for a better sense of the cognitive processes underlying students' game behaviors.

This prior work explored the use of automated detectors based on players' gameplay behavior as a means of measuring their implicit knowledge. These detectors were developed for STEM learning games using a six-step process (see Figure 1) aimed at remotely assessing learners' understanding of relevant STEM concepts based on gameplay behavior. These six steps were used to build Game Based Learning Assessments (GBLAs) for several educational games including the physics game Impulse (for more detailed explanations, see Rowe et al., 2014, 2015, 2019). This process starts by identifying gameplay consistent with the game's educational learning goals, in the case of Impulse these were Newton's First and Second Laws. Then, videos are coded in terms of specific strategic moves, noting which moves are consistent with successful achievement of the learning goals. Next, the hand coded video data are merged with the gameplay log data. Following that, educational data mining techniques are used to automate the coding based on the gameplay log data and guided by the hand coded data. Finally, the relationship between play patterns and learner performance is tested using a pre-post assessment of the target concepts (Rowe et al., 2017). This process resulted in the development of several implicit measures of Newton's first and second laws based on players' game behaviors. One measure in particular (termed n-clicks, see Methods for additional details) exhibited strong internal and external validity and was used to explore the viability of using eye tracking to measure implicit knowledge in this paper.

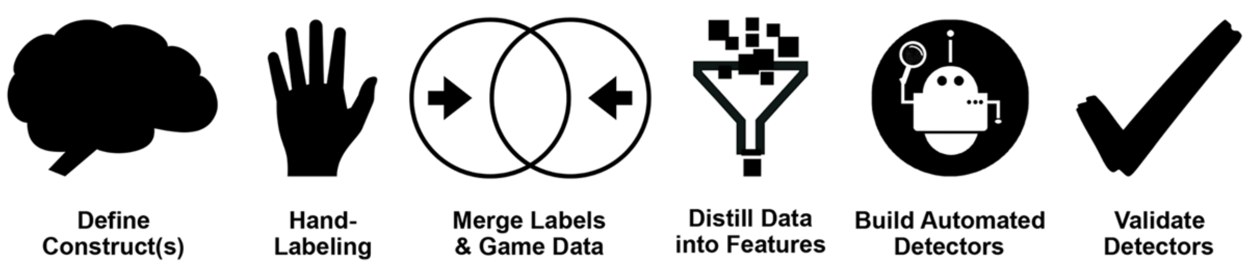

Figure 1. Graphical representation of the six-step emergent approach to GBLA (Source: Rowe et al., 2019)

This paper builds upon these prior efforts by adding an eye tracking component to the data collected during student gameplay. This additional stream of data provides potential insight into the cognitive processes underlying the learning behaviors. Previous reviews by the authors and others in the field (Alemdag & Cagiltay, 2018; Dahlstrom-Hakki et al., 2019; Lai et al., 2013) describe a variety of eye tracking measures for the characterization of learning. These include measures of attention allocation, working memory or cognitive load, and long-term memory formation.

Eye tracking has long been used as a close proxy for visual attention allocation given that eye movements closely follow shifts in both covert and overt attention (Peterson, Kramer, & Irwin, 2004). More recent work pertinent to learning focuses not only on the objects attended to but the temporal allocation of eye movements using scan path analysis techniques (Eraslan et al., 2016; Räihä et al., 2005). Scan path analysis allows for the temporal characterization of information acquisition thereby helping to reveal the cognitive aspects of the learning process in a given setting.

Eye tracking techniques have also been used to better understand the role of working and long-term memory in learning. Impacts on working memory or cognitive load can be inferred from subtle differences in fixation durations (Meghanathan, van Leeuwen, & Nikolaev, 2015) as well as from changes in pupil dilation (Miller & Unsworth, 2020). Measures of long-term memory by contrast tend to be more context dependent. Eye movements can infer the presence of knowledge by looking for anticipatory eye movements or fixations durations that are differently impacted by the presence of relevant knowledge in long-term memory (Gegenfurtner, Lehtinen, & Säljö, 2011). These inferences provide specific insight into implicit learning based on the learner's largely involuntary eye movements as opposed to explicit learning that they would need to consciously articulate.

Eye movements provide both an explicit measure of visual attention allocation using gaze location and an indication of processing difficulty based on fixation durations (Dahlstrom-Hakki et al., 2019; D'Mello, 2016). Eye tracking has been used for over a decade in GBL research to study engagement, user interface design, and general visual attention allocation (Knoepfle et al., 2009; Streicher et al., 2018; Zain et al., 2011). However, that work has generally used temporal accuracy exceeding the 100ms level which is unable to reveal the level of cognitive processing described in this study. The goal of this study therefore was to determine the viability of using a cognitive level analysis of the eye movement record as a means of remotely assessing player's understanding of Newtonian physics in the game Impulse as a first step to building GBLA informed by eye movements for this game.

1.1 **Neurodiversity.** Eye tracking has demonstrated its value in supporting non-invasive measures of cognitive processing and learning. However, the cost of eye tracking equipment and the difficulty of analyzing eye tracking data have largely relegated that work to laboratory settings. Recent advances in eye tracking technology, however, are providing the possibility of affordable and widely available eye tracking that can be widely deployed. Both the temporal and spatial accuracy of early versions of webcam-based eye tracking have been low (Papoutsaki et al., 2016), but more recent work with deep learning models (Rakhmatulin & Duchowski, 2020) and mobile eye tracking (Valliappan et al., 2020) shows a lot more promise. Recent work has shown promise in deploying webcam-based eye tracking to study cognitive processing in neurodivergent learners at scale (Wong et al., 2023). While these technologies are not ready for broad deployment yet, the work described here paves the way to developing assessments using these tools once they are broadly available.

The use of GBLA more generally and the incorporation of multi-modal data streams such as eye tracking to support these assessments more specifically is especially relevant to computer-based assessments. These assessment modalities rely on fully digital data collection with the potential to provide outputs to the learning system in real-time. With emerging eye tracking technology that does not require specialized hardware or setup, the collection and use of this type of data for assessment shows far more promise in a fully digital environment where it can be included unobtrusively than in a traditional classroom setting. Furthermore, output from such assessments in real-time can provide immediate input for adaptive games and intelligent tutoring systems to adapt the educational content in real-time to improve student performance.

1.2 **Eye tracking at Scale.** Eye tracking has demonstrated its value in supporting non-invasive measures of cognitive processing and learning. However, the cost of eye tracking equipment and the difficulty of analyzing eye tracking data have largely relegated that work to laboratory settings. Recent advances in eye tracking technology, however, are providing the possibility of affordable and widely available eye tracking that can be widely deployed. Both the temporal and spatial accuracy of early versions of webcam-based eye tracking have been low (Papoutsaki et al., 2016), but more recent work with deep learning models (Rakhmatulin & Duchowski, 2020) and mobile eye tracking (Valliappan et al., 2020) shows a lot more promise. Recent work has shown promise in deploying webcam-based eye tracking to study cognitive processing in neurodivergent learners at scale (Wong et al., 2023). While these technologies

are not ready for broad deployment yet, the work described here paves the way to developing assessments using these tools once they are broadly available.

The use of GBLA more generally and the incorporation of multi-modal data streams such as eye tracking to support these assessments more specifically is especially relevant to computer-based assessments. These assessment modalities rely on fully digital data collection with the potential to provide outputs to the learning system in real-time. With emerging eye tracking technology that does not require specialized hardware or setup, the collection and use of this type of data for assessment shows far more promise in a fully digital environment where it can be included unobtrusively than in a traditional classroom setting. Furthermore, output from such assessments in real-time can provide immediate input for adaptive games and intelligent tutoring systems to adapt the educational content in real-time to improve student performance.

1.3 **Research Question and Analysis.** The current paper explores the viability of developing eye tracking based assessments in the context of the Impulse game and similar GBL solutions. To that end, this study asks the questions: Can common eye tracking metrics be used to reliably detect implicit learning of Newtonian physics as measured by existing behavioral and paper-based measures. Given the difference in measure scales, the study accomplishes this through the following analysis. Participants are separated into clusters based on their average fixation durations. Clusters are then compared for significant differences on a pre-test of conceptual understanding of Newtonian physics. This analysis is guided by the hypothesis that students with different levels of conceptual understanding of Newtonian physics will exhibit a different overall pattern of gaze allocations. The second analysis focuses on an existing behavioral measure of Newton's second law. Average gaze durations will be used to predict performance on this measure.

2 Methods

This paper reports on a study of game-based learning measured through game detectors along with eye tracking measures of visual attention. The goal of this research was to explore the viability of using eye movements as a measure of implicit physics learning that is more accessible to neurodivergent learners.

2.1 **Sample.** Data was collected from a sample of neurodivergent young adults recruited from a campus that exclusively serves this population. A total of 66 neurodivergent young adults participated in this study. Two technical issues caused some of the data from this study to be lost. For 8 of the participants, the eye tracking data stream was missing from the data record. In addition, the pretest survey in the initial deployment of the software was corrupted causing demographic and pretest information to be missing for another 17 participants. Therefore, full data were available for 41 of the participants, whereas behavioral and eye movement data were available for 58 participants, and demographic and pretest data were available for 49 participants. Based on the demographics of the 49 participants who completed the pretest, the average participant age was 21 years with a range of 18 to 28. In terms of gender, 17 identified as female, 29 as male, and 3 chose not to respond. In terms of diagnosis, 37 participants self-reported having two or more diagnoses, 11 self-reported only one diagnosis, and one participant chose not to respond. Of the 48 participants who shared their diagnosis, 40 self-reported having a diagnosed learning disability, 34 self-reported a diagnosis of ADHD, and 19 reported a diagnosis of autism. Additionally, 36 participants reported playing video games regularly and 23 participants reported never having previously taken a physics course.

2.2 Materials

2.2.1 *Brief description of Impulse*. Data in this study was collected in the physics video game Impulse. Impulse was designed to foster and measure implicit learning about Newton's First and Second Laws of Motion. The game has a simple mechanic (get your particle to the goal without crashing into other particles) embedded in a simulation of gravitationally interacting particles (see Figure 2). Players apply a force (triggered by clicks or touch) that radiates from a single point and impacts particles based on their proximity and position relative to this point. If the player's particle collides with any ambient particle, the level is over and they must start back at the last plateau. A plateau is reached every 5 levels. Each level gets more complex, requiring players to grapple with the increasing gravitational forces of an increasing number of particles (within plateaus) and particles of different mass and thus inertia (across plateaus).

Figure 2. A screenshot from *Impulse*. The player is the green particle.
Red, blue, and white particles have different masses

In prior studies involving Impulse, researchers identified several strategies that were evident and intentional in players' behaviors (Rowe et al., 2014). These patterns were identified and coded, then distillers were used to filter, organize, and export the gameplay data in a format that could be used by a data mined detector to predict the game behaviors that are consistent with implicit understanding. One such detector relevant to this study involves implicit understanding of Newton's Second Law and was based on patterns in players' click activity related to the color (thus the mass) of the closest ball. Players who consistently used more force to accelerate heavier particles than lighter ones were considered to have demonstrated an implicit understanding of Newton's Second Law.

2.2.2 Physics Pretest. This assessment included six animated items, three dealing with Newton's First Law and three dealing with Newton's Second Law. For each topic, there was one question that resembled an animated version of a question from the Force Concept Inventory (Dancy & Beichner, 2006; Hestenes et al., 1992; Savinainen & Scott, 2002; Thornton & Sokoloff, 1998), one question using an example from Impulse, and one using an excerpt from a NASA astronaut video. Figure 3 is a sample item for Newton's First Law and Figure 4 is a sample item for Newton's Second Law (Source: Asbell-Clarke et al., 2019).

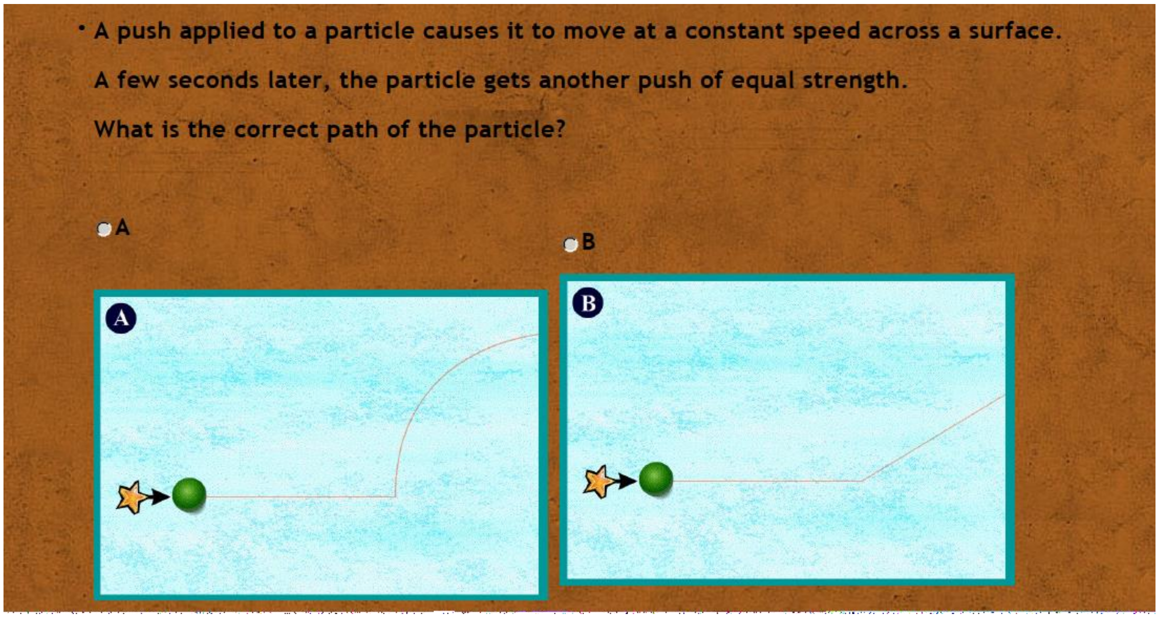

Figure 3. Sample Newton's First Law assessment item from the physics pretest, note actual items were animated. This item was adapted from the Force Concept Inventory.

The physics pretests each had a maximum of 10 points possible, 4 items focused on Newton's First Law (NFL) and 6 items on Newton's Second Law (NSL). One of the Newton's Second Law items was ambiguously worded with more than one potentially correct response and was therefore excluded from analysis. The assessment items had a Cronbach's alpha of 0.44 which was in line with a previously reported Cronbach's alpha of 0.48. This value indicates a fairly low level of internal consistency which is likely driven by the fact that good performance on NFL items was not predictive of good performance on NSL items and vice versa. Analysis therefore looks at performance on individual items rather than a combined score.

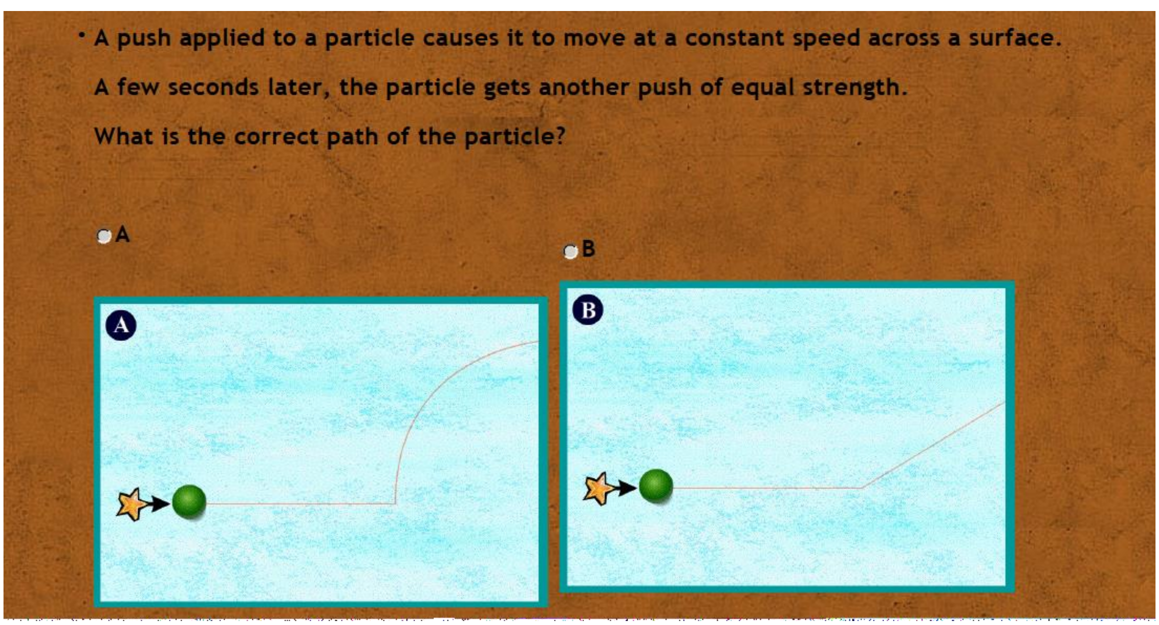

Figure 4. Sample Newton's Second Law assessment item from the physics pretest. This item was adapted from the Force Concept Inventory.

2.2.3 <u>Demographic Survey</u>. This instrument included items asking participants to self-report their age, gender, disability status, familiarity with video games, and prior instruction in physics.

2.2 **Materials.** The Eyelink 1000 eye tracker was used in remote tracking mode to collect data from participants while they played the Impulse game on a monitor with 120 Hertz refresh rate. In remote tracking mode, the Eyelink sits on the desktop and collects images of the eyes using a wide-angle lens on a high-speed infrared camera sampling at the rate of 500 Hertz. No head restraint was used since the Eyelink 1000 system can compensate for head movements. Eye movement data were collected from the right eye. The eye tracker parser automatically classified fixations based on foveated areas of the visual field between saccades. Saccades were detected when an eye movement had a velocity over 30°/s, an acceleration over 8000°/s^2, and moved at least 0.1°.

2.3 **Measures.**

2.3.1 <u>Gameplay Data</u>. All student assessment and game log data were collected through the game data collection architecture, Data Arcade. Data Arcade was designed and built to collect, organize, and visualize data collected from game activity. As part of this architecture, an API is built into the game allowing each player's game activity and every corresponding game event to be logged and associated with a timestamp and a unique (and anonymous) player ID. Over multiple GBLA studies, the authors have designed a suite of tools with the data architecture to enable:

- Registration of players by classes or individuals.
- Synchronization of game data with other sources (e.g., surveys, external pre/post assessments, and multimodal sensory data streams); and
- Visualization tools that allow the "playback" and hand-labeling of gameplay generated from the data logs.

The Playback Tool is a unique innovation in which a "replay" visualization of the gameplay is generated from log activity and is displayed in a window with a series of menus below that researchers use for hand-labeling of the data. Researchers can easily scrub through the video timelines to find events and the Playback Tool can snap to an event to avoid time-consuming and tedious event synchronization tasks. Researchers can customize the labeling tool for different puzzles and different games. The Playback Tool was used for visualization of eye tracking overlaid on game play.

2.3.2 <u>Eye Data</u>. Because of the dynamic nature of the stimuli in this study which elicited frequent instances of smooth pursuit, the eye movement data were based on an analysis of gaze durations and locations rather than fixations. Gaze locations were used as a close proxy for the allocation of visual attention to an object. Given the time sensitive nature of the gameplay, gaze durations were used as an indicator of the processing resources allocated to deal with a foveated object. For the purposes of this study, a gaze was defined as the total time in which a game object's center was closest to the location of the foveated region and within 3 degrees of visual angle. A gaze begins at the start of the first fixation on an object that meets these conditions and lasts until the end of the last fixation that meets these conditions. Therefore, a gaze ends when another object's center was closer to the foveated region or if the object becomes more than 3 degrees of visual angle away from the foveated region. This means that gaze location was based on object location and not absolute spatial location, and gaze duration was based on total time on an object rather than the duration of individual stationary fixations.

2.4 **Procedure.** Data was collected in a lab setting setup on the college campus. Data collection began with the pretest instrument that participants were asked to complete. The eye tracker was then set up and calibrated for the participant, a procedure which typically takes 5-10 minutes. The Impulse game was then started, and the eye tracker, game, and mouse data streams were connected to Data Arcade, the data collection architecture. Participants then independently played Impulse for the remainder of the session with each session lasting a total of 1 hour. Participants received a $25 gift card for their participation in this study.

2.4.1 <u>Analysis</u>. To assess whether the eye movement data collected provided a useful means of assessing students' physics knowledge, two analyses were conducted, one comparing overall gaze patterns to students' knowledge on a physics pretest and one comparing it to previously created detectors based on player behavior. The first analysis involved building a linear mixed-effects model to test the association between participant scores on the physics pretest and their gaze allocation patterns. Players were separated into clusters using a Gaussian Mixture Model (GMM) algorithm based on each individual's average gaze durations on particles of each of the four colors: Blue particles with the lowest mass, Red particles of equal size but double mass, and White and Grey particles with double the mass of Red. In addition, White particles were larger than Red ones, and Grey particles were smaller than Red ones. Resultant clusters were used in a Generalized Linear Mixed Effects Model (GLMEM) to predict individual pretest item scores, with random slopes for each player and pretest item.

The second analysis examined the association between students' eye movement patterns and the sequence length of clicks (termed N-Clicks) on particles of a given color. Sequence length is the number of consecutive times a player clicked to move the same particle in under 4 seconds. Prior work by the authors has supported the hypothesis that learners who have an implicit understanding of Newton's laws exert greater force (and therefore more clicks) to move particles of higher mass (Rowe et al., 2015). Gaze duration data was normalized and centered. This was done because we are interested in how relative rather than absolute gaze durations are predictive of N-Clicks and to facilitate model convergence. A GLMEM Poisson regression modeled sequence lengths on particles of each of the four colors, with random slopes for each player and level. Both sets of analyses were conducted in R Studio 1.1463.

The N-Click measure is intended to provide an implicit measure of understanding of Newton's Second Law. This was done by coding the target of the current click (i.e. force allocation) and whether the target was the same as the previous click (indicating additional force exerted to the same target). The N-Click measure used two detectors (see Table 1) to determine the number of clicks (N-Clicks) on a particle of a given color thereby indicating the amount of force exerted on that particle.

Table 1. Accuracy of detectors used for N-Click measure based on hand coded videos of gameplay.

Code	**Definition**	**Kappa**
Target	Type of particle (player, other, both) the learner intended to move	0.920
Same as Last Target	The learner intended to move the same target as the last action	0.869

Note. Rowe, Baker, Asbell-Clarke, Kasman, & Hawkins (2014).

The N-clicks on particles of different masses were selected for this analysis because they were found to be the strongest evidence of Newton's Second Law in prior research with Impulse (Rowe et al., 2017). Sequence length was based on clicks on the same blue, red, white, or gray particle within a round. For instance, if a student had sequences of clicks for the following particles in round 1: Red 1, Red 1, Blue 1, Blue 1, Blue 1, Blue 2, Blue 2. The maximum N-Click for red would be 2 and the maximum N-Click for blue would be 3 for that round. These maximum N-Clicks for each of the colored particles were used in predicting gaze duration below.

3 Results

GMM clustering using mahalanobis distances as the distance measure was used to create discrete groupings of gaze duration patterns. Participant clusters were created by examining each participant's average gaze durations on particles of each type. Correlations between average gaze durations can be seen in Table 2. To identify an optimal cluster size for use in this analysis based on the GMM clustering of gaze durations on particles of different color, the Akaike Information Criterion (AIC) was used. The AIC is often used to assess the quality of a model by looking at the quality of a model's predictions balanced against the number of predictors. As shown in Figure 5, the first elbow in the graph occurs at cluster 3, indicating a reasonable balance between the number of clusters and the cohesion of gaze duration patterns in each cluster.

Table 2. Correlation between participant average gaze durations across participle types.

	Blue	**Red**	**White**	**Gray**
Blue	1.00			
Red	0.90	1.00		
White	0.89	0.97	1.00	
Gray	0.90	0.94	0.99	1.00

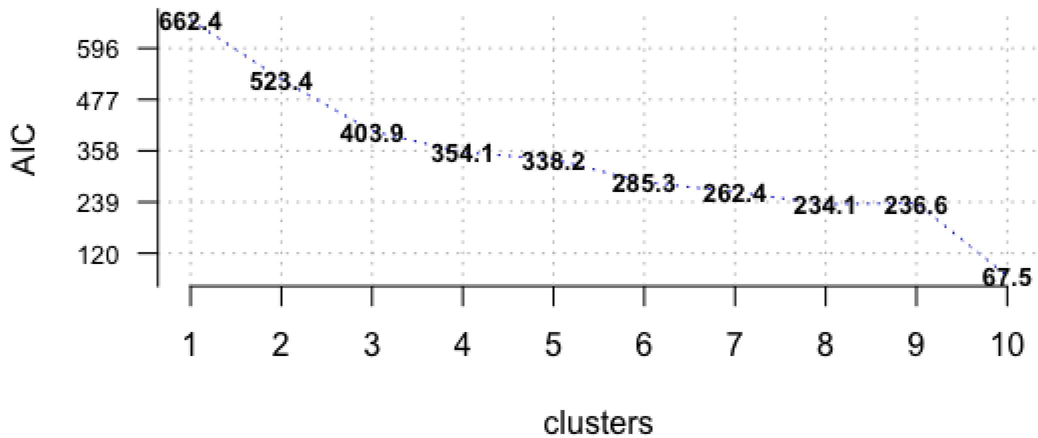

Figure 5. AIC values by number of clusters.

Table 3 provides an overview of descriptive statistics for the three resultant gaze clusters including each cluster's pretest score, percentage of ADHD/autism/learning disability, and maximum N-Clicks for each particle (note that average pretest scores and percentage of ADHD/autism/learning disability are based on the 41 participants for whom full data was available). Figure 6 shows average gaze durations on particles of different color (and therefore mass) for each participant in this study and illustrates how patterns differ across the three clusters. Gaze cluster 1 was the largest and had the lowest pretest performance, gaze cluster 2 had the least number of students and had the highest pretest performance, and gaze cluster 3 was slightly larger than cluster 2 and had performance midway between the two other clusters. Chi-squared analyses indicated no difference across clusters in terms of ADHD, autism, or learning disability. Diagnosis was explored as a potential factor in the below models but did not lead to significant improvements in model fit.

Table 3. Correlation between participant average gaze durations across participle types.

Cluster	**N**	**Avg Pretest Score (SD)**	**ADHD**	**Autism**	**Learning Disability**	**Mean Maximum N-Click (SD)**			
						Blue	*Red*	*White*	*Gray*
1	32	6.59 (1.47)	68%	64%	14%	2.35 (0.80)	2.54 (0.77)	3.18 (0.95)	2.19 (0.60)
2	11	8.00 (1.69)	75%	50%	25%	2.76 (1.03)	2.72 (1.33)	2.96 (1.60)	2.09 (0.30)
3	15	7.46 (1.44)	73%	73%	9%	2.50 (0.56)	2.78 (0.89)	3.06 (0.90)	2.18 (0.68)

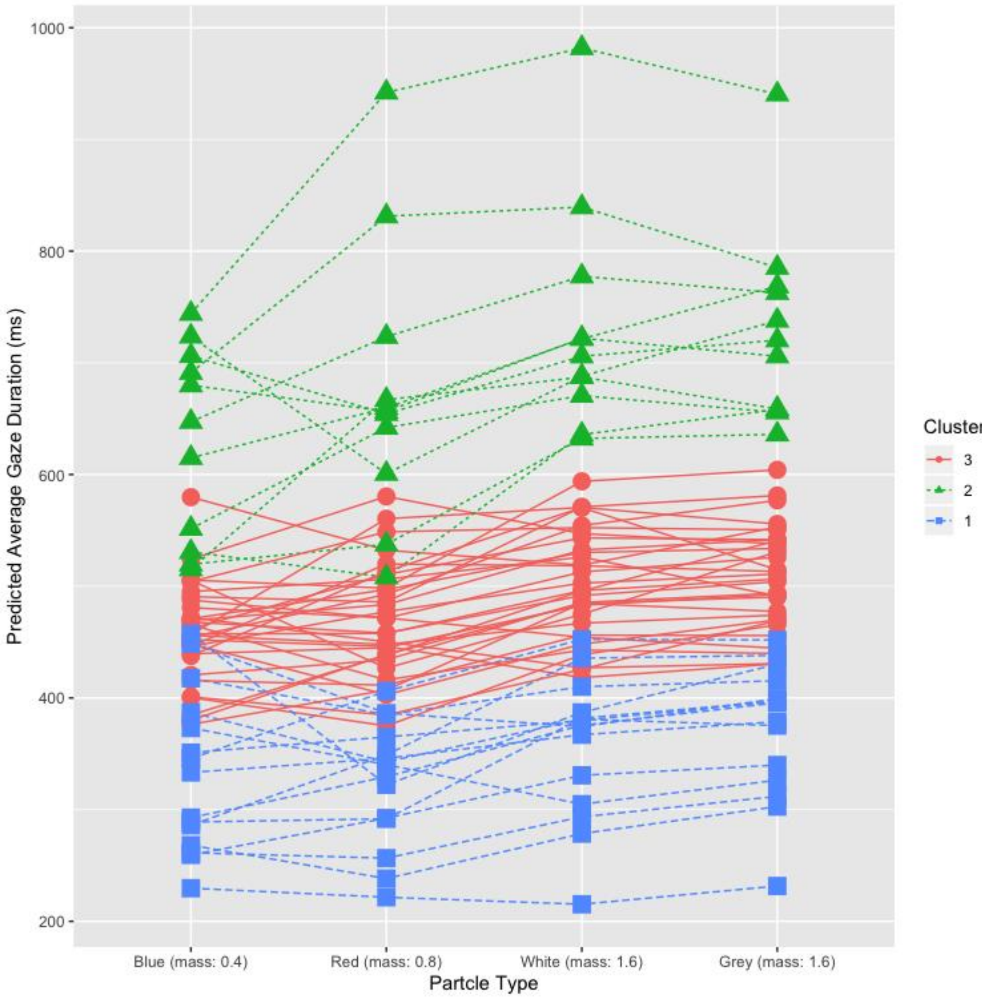

Figure 6. Eye movement patterns across gaze clusters (Cluster 1, n=32; Cluster 2, n=11; Cluster 3, n=15)

To test whether eye movement patterns are predictive of students' physics knowledge, a logit regression was performed using a Generalized Linear Mixed Effects Model (GLMEM). This model predicts each individual participant's likelihood of having answered correctly to each of the physics pre-assessment items while controlling for individual specific and item specific variability. Gaze cluster 1 was used as the baseline cluster and the data was modeled using the following equation:

$$logit(Y_{ij}) = \beta_0 + \mu_{0i} + \nu_{0j} + \beta_1(Cluster\ 2) + \beta_2(Cluster\ 3) + \epsilon_{ij}$$

Y_{ij}= Probability of participant i responding correctly to item j
β= Fixed component
μ= Participant random component
ν= Item random component
ϵ= Residual error

As can be seen in Table 4, the data indicates that participants in gaze cluster 2 were significantly more likely to perform better on the physics pre-assessment items than participants in gaze cluster 1. This is notable given the fact that the clustering was performed purely based on the gaze durations of participants on different color particles and was completely blind to their prior physics knowledge. Gaze cluster 3 had a higher coefficient than gaze cluster 1 but this difference was not statistically significant, however with a p value of 0.11 it is possible that this failed to reach the level of significance due to a lack of power.

Table 4. Results of GLMEM regression analysis of how predictive gaze cluster is of pre-assessed physics performance

Fixed Effect	**Estimate**	**SE**	**z-value**	**p-value**
Intercept (β_0)	1.08	0.58	1.86	0.06 †
Cluster 2 (β_1)	1.07	0.45	2.38	0.02 *
Cluster 3 (β_2)	0.62	0.39	1.61	0.11
Random Effects	**Variance**	**SD**		
Participant (μ_0)	0.37	0.61		
Item (ν_0)	2.84	1.69		

The analysis of the clustering data provides for a fairly coarse-grained test of the association between global gaze patterns in the Impulse game and Newtonian physics knowledge. Despite this, the data indicate that some players with greater Newtonian physics knowledge exhibit distinct gaze allocation patterns during gameplay indicating the gaze durations may be a useful means of implicitly assessing students' physics knowledge. To further corroborate this finding, a more fine-grained analysis was performed looking at the association between gaze allocation patterns and the maximum N-Clicks on particles of different mass, a measure previously found to be strongly associated with strong performance on assessments of Newtonian physics.

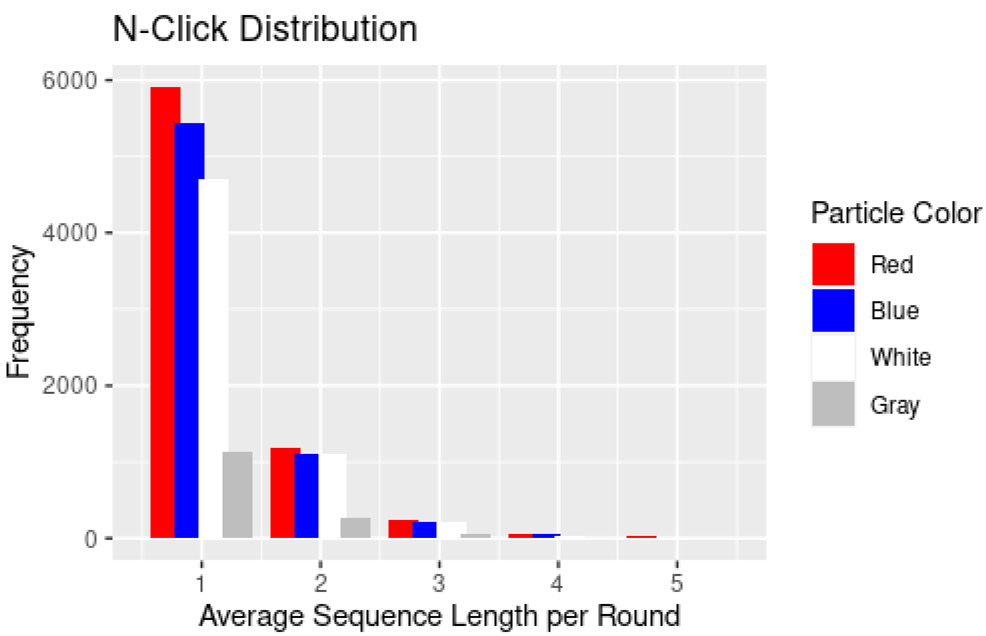

Figure 7. Histogram of average round sequence lengths per particle type.

This measure counts the number of clicks on a specific particle within a short period of time (see Figure 7 for a distribution of average sequence lengths per round). Therefore, to model this count data, a GLMEM Poisson regression was used. The model predicts the maximum number of consecutive clicks on particles of each color on every game round while accounting for participant specific and level specific variability. To facilitate model convergence, the gaze duration data was centered and normalized. The data was modeled as follows:

$$\log(Y_{ij}) = \beta_0 + \mu_{0i} + \nu_{0j} + \beta_1(Red) + \beta_2(White) + \beta_3(Gray) + \beta_4(Normalized\ Average\ Gaze\ Duration) + \epsilon_{ij}$$

Y_{ij}= Expected N-Clicks on particles of a given color
for participant i on game round j
β= Fixed component
μ= Participant random component
ν= Game level random component
ϵ= Residual error

Prior work indicates a strong association between particle color and the N-Click measure with a longer click sequence on higher mass particles being associated with better subsequent performance on assessments of Newtonian physics (Rowe et al., 2017). The regression coefficients for particle color show that participants exhibited this previously identified pattern. The blue particle had the lowest mass and was used as the baseline particle in the regression model. As can be seen in Table 5, red particles with twice the mass of blue particles received significantly higher N-Clicks, and gray and white particles, both twice the mass of red particles (but different sizes), had even higher coefficients. In addition, the normalized average gaze duration had a statistically significant association with N-Clicks indicating that particles that received longer gazes also tended to receive more N-Clicks.

Table 5. Results of GLMEM regression analysis of how predictive normalized average gaze durations are of the N-Click measure.

Fixed Effect	**Estimate**	**SE**	**z-value**	**p-value**
Intercept (β_0)	0.44	0.08	5.31	0.000 ***
Particle: Red (β_1)	0.28	0.05	6.23	0.000 ***
Particle: White (β_2)	0.58	0.06	10.18	0.000 ***
Particle: Grey (β_3)	0.70	0.07	10.15	0.000 ***
Normalized Average Gaze Duration (β_4)	0.04	0.01	2.94	0.003 **
Random Effects	**Variance**	**SD**		
Participant (μ_0)	0.03	0.17		
Game Level (ν_0)	0.24	0.49		

5 Discussion

The goal of this work was to explore whether the eye movement record of participants playing a fast action physics video game could provide evidence of implicit knowledge of Newtonian physics. The results are consistent with this assertion. This paper begins by looking for a broad association between overall eye movement patterns and an explicit measure of physics knowledge. This was accomplished by using a clustering algorithm to segment participants into groups based on their overall gaze patterns on particles of different color and mass. The results indicated a significant difference in performance on the Newtonian physics assessment between these clusters. This provides evidence that there is a strong association between students' knowledge of Newtonian physics and their gaze allocation patterns in the Impulse game. It was unclear a priori whether there were global gaze allocation patterns associated with physics knowledge, but the results of the gaze cluster regression model show that statistically significant differences are present even at this global level.

The first analysis looked at a broad association between overall gaze patterns and assessment scores, whereas the second analysis took a finer grain look at the data using an existing measure of implicit Newtonian physics knowledge. The finer grained analysis provides evidence indicating that the association between gaze patterns and Newtonian physics is not merely a global pattern common to each cluster but is based on gaze differences specific to individual game rounds. This analysis looked for an association between gaze allocation patterns and the N-Click measure, a detector developed in prior work found to be a strong predictor of performance on assessments of Newtonian physics. The analysis showed that longer gazes on particles of a given color (and therefore mass) were associated with more N-Clicks on those particles which in turn was previously found to be associated with implicit knowledge of Newtonian physics (Rowe et al., 2017). Furthermore, while the association with the external assessment scores is an association with an explicit measure, the finer grain analysis provides evidence of an association with the behaviors and cognitive processing consistent with an implicit understanding of Newtonian physics.

It is likely that average gaze durations are driven at least in part by the need for more clicks on particles of higher mass. However, if we take the results of both analyses together, that is unlikely to be the sole driver of differences in gaze duration across participants and rounds. A look at average gaze durations in Figure 6 reveals that participants in cluster 1 have shorter average gaze durations across the board regardless of particle mass, an indication that they may not be exerting a great deal of cognitive processing resources into optimizing their force allocations during the game. Furthermore, while average gaze durations tended to increase for particles of higher mass (therefore requiring more clicks), this was not the case for all participants across all particles in the highest performing cluster (cluster 2). The authors hope to pursue follow-up work in the future aimed at characterizing the dynamics of gaze durations across rounds for individual players.

Critical to this work is that these patterns are evident in a population of neurodiverse learners. As described in the introduction, eye tracking has been used to improve understanding of the deployment of visual attention and cognitive processing for neurodiverse learners. The work here indicates that additional more nuanced analysis of the eye movement record may be helpful in understanding why some neurodivergent learners may struggle to learn in a GBL environment and what supports may be needed to help them address attention or processing difficulties that may impact their performance.

While these analyses are consistent with evidence that the eye movement record may be predictive of implicit knowledge of Newtonian physics, much work remains before such a connection can be established. This work is exploratory and is intended to demonstrate that a connection exists. Follow up work will need to address whether performance can be accurately assessed on individual trials and explore whether this finding can be replicated with emergent more affordable eye tracking solutions. In addition, other forms of eye movement analysis including scan path analyses should be explored to provide a more complex analysis of gaze allocation patterns in this and similar assessment activities.

Key to the successful wide deployment of eye movement-based assessments is the maturation of webcam- and mobile-based eye tracking solutions. The current study used a research grade eye tracker (Eyelink 1000) that retails for thousands of dollars. Indeed, it is unlikely that you would find an eye tracker today for much less than a thousand dollars that would be accurate enough to allow for the analysis proposed here. However, recent advances in deep learning algorithms are making the potential for accurate and affordable remote eye tracking a likely reality in the near future (Rakhmatulin & Duchowski, 2020; Valliappan et al., 2020). Importantly, these solutions require no specialized hardware, setup, or calibration. As these technological solutions become available, it is important to conduct the type of research described in this paper in parallel to have tools ready to support computer-based learning as webcam and mobile-based eye tracking solutions mature.

This assessment model is especially relevant to computer-based learning. Traditional classrooms provide far more opportunities for teachers and other learning support personnel to monitor and intervene during the learning process, whereas there is less opportunity for that with many forms of computer-based learning. The COVID-19 global pandemic has highlighted the need for better tools to monitor the progress of struggling learners in remote learning and the need for approaches to better understand the source of those struggles. The use of real-time eye tracking can provide online formative assessments that can be used by teachers or adaptive models to support learners.

The use of GBLA and other digital assessments informed by multi-modal data streams such as eye tracking show great promise in leveling the playing field for many struggling learners. Using emerging technologies that require no specialized hardware or setup, these tools can seamlessly be deployed and can provide real-time feedback to adaptive tools and intelligent tutoring systems. This can enable more customized computer-based learning to meet the needs of diverse individuals without many of the language and attention barriers faced by neurodivergent learners on more traditional forms of assessment.

This work has several limitations that need to be addressed in future studies. This includes the need to replicate findings with larger samples of students, to provide an explanatory mechanism between specific eye movement patterns and implicit knowledge of Newtonian physics, to gain a better understanding of the eye movement patterns exhibited by different clusters of players, and to explore task independent models with greater predictive power. Prior work in eye tracking supports the finding that when task performance is time sensitive (such as in reading and search tasks) longer fixation or gaze durations are associated with processing difficulty (Rayner, 2009). This finding has been replicated here and serves as the most viable task independent metric for assessing learner performance. However, translating these data into information that supports educators in a learning setting is not trivial because low fixation durations are an indicator of both low processing difficulty and an individual who is not attempting to solve the problem. Work in this area will need to carefully integrate both eye movement and behavioral measures to guide supports for neurodivergent learners. In addition, high-quality eye tracking remains expensive which limits the scalability of this work currently. However, as new and cheaper means of eye tracking become more available, the potential for this work to be taken to scale will become more viable.

While traditional assessments remain far easier to implement and collect than the assessments being explored in this work, there are clear differences between the barriers imposed by these assessments with respect to neurodivergent learners. GBLA that use eye movement and behavioral measures do not impose some of the barriers that can negatively impact the performance of neurodivergent learners. This study has used traditional assessments as a means to provide an initial viability check for the use of eye movements to measure implicit learning. However, given the issues with validity of these items for neurodivergent learners, future work must explore the use of more direct assessments of physics knowledge using such approaches as in-depth structured interviews with a trusted ally.

6 Conclusion

The use of eye movement patterns as a means of providing formative and eventually summative assessment of Newtonian physics holds a lot of promise, particularly for neurodivergent students. Knowledge can be assessed with little impact of the functional limitations experienced by these populations. As understanding of the explanatory mechanisms between particular eye movement patterns and physics knowledge matures, interventions can be customized to better support underserved populations. The work presented in this paper provides guidance to researchers looking to develop eye movement-based measures by illustrating how the data can be collected, segmented into gazes, and analyzed. Further refinement of the patterns associated with student behavior can improve equitable access to assessments not only in physics but in other related fields as well.

7 Acknowledgements & Data Availability Statement

This material is based in part upon work supported by the National Science Foundation under Grant Numbers DRL-1417456 and DRL-1417967. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors have no conflicts of interest to disclose.

The data that support the findings of this study are available on request from the corresponding author, IDH, provided the request does not jeopardize participant anonymity

References

Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. Teachers College Record, 108(11), 2282. https://doi.org/10/cp3sr3

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. Computers & Education, 125, 413–428.

Asbell-Clarke, J., Rowe, E., Bardar, E., & Edwards, T. (2019). The Importance of Teacher Bridging in Game-Based Learning Classrooms. In M. Farber (Ed.), Global Perspectives on Gameful and Playful Teaching and Learning (pp. 211–239). IGI Global.

Association, A. E. R., Association, A. P., & Education, N. C. on M. in. (2014). AERA, APA, & NCME. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association.

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. Technology, Knowledge and Learning, 19(1–2), 205–220.

Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). Make it stick: The Science of successful learning. Harvard University Press.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. International Journal of Psychophysiology, 155, 49–62.

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. Research in Developmental Disabilities, 48, 79–93.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Petersen, S. E., Raichle, M. E., Van Essen, D. C., & others. (1998). A common network of functional areas for attention and eye movements. Neuron, 21(4), 761–773.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78(1), 98.

Dahlstrom-Hakki, I., Asbell-Clarke, J., & Rowe, E. (2019). Showing Is Knowing: The Potential and Challenges of Using Neurocognitive Measures of Implicit Learning in the Classroom. Mind, Brain, and Education, 13(1), 30–40.

Dahlstrom-Hakki, I. H., & Alstad, Z. G. (2019). Challenges Assessing the Conceptual Understanding of Students With Disabilities in Statistics: Learning Disability Quarterly. https://doi.org/10/gf7ddw

Dancy, M. H., & Beichner, R. (2006). Impact of animation on assessment of conceptual understanding in physics. Physical Review Special Topics-Physics Education Research, 2(1), 010104.

D'Mello, S. K. (2016). Giving eyesight to the blind: Towards attention-aware AIED. International Journal of Artificial Intelligence in Education, 26(2), 645–659.

Eraslan, S., Yesilada, Y., & Harper, S. (2016). Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. Journal of Eye Movement Research, 9(1).

Ford, D. Y., & Helms, J. E. (2012). Overview and introduction: Testing and assessing African Americans:"Unbiased" tests are still unfair. The Journal of Negro Education, 81(3), 186–189.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. Educational Psychology Review, 23(4), 523–552.

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. Neuroscience & Biobehavioral Reviews, 42, 279–297.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. The Physics Teacher, 30(3), 141–158. https://doi.org/10.1119/1.2343497

Horowitz, S. H., Rawe, J., & Whittaker, M. C. (2017). The State of Learning Disabilities: Understanding the 1 in 5. New York: National Center for Learning Disabilities.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. Perception & Psychophysics, 40(6), 431–439.

Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

Ke, F., & Shute, V. (2015). Design of game-based stealth assessment and learning support. In Serious games analytics (pp. 301–318). Springer.

Klingner, J. M. (2010). Measuring cognitive load during visual tasks by combining pupillometry and eye tracking [PhD Thesis]. Stanford University Palo Alto, CA.

Knoepfle, D. T., Wang, J. T., & Camerer, C. F. (2009). Studying learning in games using eye-tracking. Journal of the European Economic Association, 7(2–3), 388–398.

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educational Research Review, 10, 90–115.

Larmer, J., Mergendoller, J., & Boss, S. (2015). Setting the standard for project based learning. ASCD.

Lev, A., Braw, Y., Elbaum, T., Wagner, M., & Rassovsky, Y. (2022). Eye tracking during a continuous performance test: Utility for assessing ADHD patients. Journal of Attention Disorders, 26(2), 245–255.

Levantini, V., Muratori, P., Inguaggiato, E., Masi, G., Milone, A., Valente, E., Tonacci, A., & Billeci, L. (2020). Eyes are the window to the mind: Eye-tracking technology as a novel approach to study clinical characteristics of ADHD. Psychiatry Research, 290, 113135.

McNamara, J. K., & Wagner, J. (2001). Implicit and explicit awareness of a phonics rule in the word recognition of students with and without learning disabilities. Alberta Journal of Educational Research, 47(1).

Meghanathan, R. N., van Leeuwen, C., & Nikolaev, A. R. (2015). Fixation duration surpasses pupil size as a measure of memory load in free viewing. Frontiers in Human Neuroscience, 8. https://doi.org/10/gf7dfn

Miller, A. L., & Unsworth, N. (2020). Variation in attention at encoding: Insights from pupillometry and eye gaze fixations. Journal of Experimental Psychology: Learning, Memory, and Cognition, 46(12), 2277–2294. https://doi.org/10.1037/xlm0000797

Nieminen, J. H. (2023). Unveiling ableism and disablism in assessment: A critical analysis of disabled students' experiences of assessment and assessment accommodations. Higher Education, 85(3), 613–636. https://doi.org/10.1007/s10734-022-00857-1

Nilsson Benfatto, M., Öqvist Seimyr, G., Ygge, J., Pansell, T., Rydberg, A., & Jacobson, C. (2016). Screening for dyslexia using eye tracking during reading. PloS One, 11(12), e0165508.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). Webgazer: Scalable webcam eye tracking using user interactions. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence-IJCAI 2016.

Peterson, M. S., Kramer, A. F., & Irwin, D. E. (2004). Covert shifts of attention precede involuntary eye movements. Perception & Psychophysics, 66(3), 398–405.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. Educational Psychologist, 50(4), 258–283.

Polanyi, M. (1966). The Tacit Dimension. Doubleday.

Räihä, K.-J., Aula, A., Majaranta, P., Rantala, H., & Koivunen, K. (2005). Static visualization of temporal eye-tracking data. IFIP Conference on Human-Computer Interaction, 946–949.

Rakhmatulin, I., & Duchowski, A. (2020). Deep Neural Networks for Low-Cost Eye Tracking (SSRN Scholarly Paper ID 3706602). Social Science Research Network. https://papers.ssrn.com/abstract=3706602

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology, 62(8), 1457–1506.

Rello, L., & Ballesteros, M. (2015). Detecting readers with dyslexia using machine learning with eye tracking measures. Proceedings of the 12th International Web for All Conference, 1–8.

Rowe, E., Asbell-Clarke, J., & Baker, R. S. (2015). Serious games analytics to measure implicit science learning. In Serious games analytics (pp. 343–360). Springer.

Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., Brown, R. A., & Edwards, T. (2017). Assessing implicit science learning in digital games. Computers in Human Behavior, 76, 617–630.

Rowe, E., Asbell-Clarke, J., Bardar, E., Almeda, M. V., Baker, R. S., Scruggs, R., & Gasca, S. (2019). Advancing Research in Game-Based Learning Assessment: Tools and Methods for Measuring Implicit Learning. In E. Kennedy & J. Qian (Eds.), Advancing Educational Research With Emerging Technology (pp. 99–123). IGI Global.

Rowe, E., Baker, R., Asbell-Clarke, J., Kasman, E., & Hawkins, W. (2014). Building Automated Detectors of Gameplay Strategies to Measure Implicit Science Learning. Educational Data Mining 2014. http://www.educationaldatamining.org/conferences/index.php/EDM/2014/paper/download/1374/1340

Savinainen, A., & Scott, P. (2002). The Force Concept Inventory: A tool for monitoring student learning. Physics Education, 37(1), 45.

Sideridis, G. D. (2016). Assessing validity of measurement in learning disabilities using hierarchical generalized linear modeling: The roles of anxiety and motivation. Educational and Psychological Measurement, 76(4), 638–661. https://doi.org/10/gf7dfg

Streicher, A., Leidig, S., & Roller, W. (2018). Eye-tracking for user attention evaluation in adaptive serious games. European Conference on Technology Enhanced Learning, 583–586.

Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. American Journal of Physics, 66(4), 338–352. https://doi.org/10.1119/1.18863

Valliappan, N., Dai, N., Steinberg, E., He, J., Rogers, K., Ramachandran, V., Xu, P., Shojaeizadeh, M., Guo, L., Kohlhoff, K., & Navalpakkam, V. (2020). Accelerating eye movement research via accurate and affordable smartphone eye tracking. Nature Communications, 11(1), 4553. https://doi.org/10.1038/s41467-020-18360-5

Van den Bergh, V., Mortelmans, D., Spooren, P., Van Petegem, P., Gijbels, D., & Vanthournout, G. (2006). New assessment modes within project-based education-the stakeholders. Studies in Educational Evaluation, 32(4), 345–368.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard university press.

Wong, A. Y., Bryck, R. L., Baker, R. S., Hutt, S., & Mills, C. (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, 453–463. https://doi.org/10.1145/3576050.3576115

Zain, N. H. M., Razak, F. H. A., Jaafar, A., & Zulkipli, M. F. (2011). Eye tracking in educational games environment: Evaluating user interface design through eye tracking patterns. International Visual Informatics Conference, 64–73.