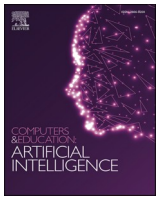
Computers and Education: Artificial Intelligence 3 (2022) 100078

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence



Real-time prediction of science student learning outcomes using machine learning classification of hemodynamics during virtual reality and online learning sessions

Richard Lamb *, Knut Neumann, Kayleigh A. Linder

East Carolina University, College of Education, Neurocognition Science Laboratory, 128 Rivers Building, Greenville, NC, 27858, USA

ARTICLE INFO

Keywords:
Science student learning
Online learning systems
Functional near infrared spectrometry (fNIRS)
Student prediction
User experience

ABSTRACT

Current data sources used for the prediction of student outcomes average about 55% accuracy and require a significant amount of input data and time for researchers and educators to produce predictive models of student outcomes. The aim of this study is to examine how neurocognitive data collected via functional near infrared spectroscopy (fNIRS) may be used to create predictive models of student outcomes with greater speed and accuracy when using a synthetic adaptive learning environment (SALEs). Specifically, this study examines the utility of using neurocognitive data to develop student response prediction on a science content test. Participants were recruited from schools located in the United States (n = 40). Participants in the study engaged in three conditions: no content, video and virtual reality. The lesson video and virtual reality lesson provides an explanation of deoxyribonucleic acid replication. Observed neurocognitive responses were collected during each condition and used to predict the success of student responses on an assessment. On average the predicative accuracy of this approach is 85% and occur within 300 ms. Predictive error rates are less than 15%. Results of this study provides evidence to support the use of neurocognitive data for adaption of digitally presented content and how machine learning approaches and artificial intelligence may be used to classify student data in real-time as students engage with content. Results also illustrate good accuracy and capture of moment-to- moment fluctuations of cognition in real-time. These findings may help the development of artificially intelligent tutors and improve student-based learning analytics.

1. Introduction

There has been an increased awareness of the benefits of synthetic adaptive learning environments (SALEs) in a variety of sectors from the military to K-12 education (Alexander et al., 2019, pp. 3–41). SALEs are digital learning environments which respond and change due to user preferences and inputs over time. To accomplish this the SALEs must be able to predict student outcomes and responses quickly and accurately which is currently a barrier to more successful implementation in education. To underscore this point, the Department of Defense (DOD) and other agencies such as the National Science Foundation (NSF) and Department of Education (DOE) have increased investments year-over-year to identify and develop technologies and methods to increase the accuracy and speed of prediction within adaptive synthetic learning environments under the Synthetic Learning Environments Group. DOD, DOE, and NSF are federal divisions within the United States government responsible for supervising military forces, national educational policy, and supporting the progress of science, respectively. The large-scale spending by DOD, DOE, and NSF is driven by the understanding that computerized adaptive learning environments, if given appropriate inputs can be used to create a more ideal learning environment, individualize learning, and increase the efficacy and efficiency of learning for all students. SALEs can be used at scale, increasing access to learning opportunities for students who may not be able to attend class or are a part of an online learning program, particularly in a post COVID-19 educational environment (Simamora, 2020). A keystone of SALEs is the use of student response data to build an individualized experience and adapt content presentations. Using student data, the individualization of content and experiences can create levels of engagement and help to strengthen connections between concepts increasing learning outcomes. However, individualization comes at a cost, the cost is the requirement for substantial amounts of data in short

* Corresponding author. Richard Lamb Neurocognition Science Laboratory, East Carolina University, Greenville, NC, 27858, USA.
E-mail address: lambr19@ecu.edu (R. Lamb).

https://doi.org/10.1016/j.caeai.2022.100078
Received 2 February 2022; Received in revised form 4 May 2022; Accepted 6 May 2022
Available online 20 May 2022

periods (McMahon, Wright, Cihak, Moore, & Lamb, 2016; Umlauft & Hirche, 2019). This data collection often takes the form of questionaries, surveys, and other assessments which do not generate particularly high levels (~55%) of model accuracy related to model prediction (Gardner & Brooks, 2018). Other concerns about these forms of data collection can lead to student fatigue, reduced motivation, and other maladaptive behaviors within SALEs (Mädamürk, Tuominen, Hietajärvi, & Salmela-Aro, 2021).

SALEs are environments that encompass virtual reality (VR), augmented reality, flat screen video, and other digital reality mixes which are specifically designed with pedagogy in mind. VR is a computer-generated simulation using fully immersive three-dimensional environments that are interactive in a realistic way. SALEs, learning management systems, and associated technologies have been shown to develop student learning related to the content and practice of science (Sablić, Mirosavljević, & Škugor, 2020; Lamb, Annetta, Hoston, Shapiro, & Matthews, 2018; Lamb et al., 2018). As a part of the learning individualization process, SALEs and other systems work by collecting assessment response data, individual preference data, and student learning system interaction data to create a student driven on-demand learning system (Lamb & Annetta, 2009, 2012; Lamb et al., 2014). Meaning as a student progresses through the content, the system collects information about their selections related to videos, lecture notes, and other content. This data is used to help design content and increase student engagement and attention. To accomplish this, large amounts of data must be collected in a short time and processed quickly to allow the digital system to adaptively respond (Khan & Khojah, 2022). The underlying assumption within these systems is that the student is not only capable, but well versed enough in their understanding of content and practices associated with learning to effectively select the correct "next step" as they learn. In this case, student "mistakes" in selection reduce the accuracy and speed of SALEs adaption. Even in cases where selection and adaption of content occurs because of student assessments, the underlying assumption is that the assessment data accurately reflects student learning and understanding. Misalignment of the assessment to the content will result in reduced accuracy and speed of adaption. A means to address these shortcomings -the need for substantial amounts of data, rapid processing, and an accurate understanding of student cognitive state-when developing predictive models in SALEs is through the integration of neurocognitive data.

Within SALEs and other digital learning environments, adaption refers to several automatic modifications of the digital environment which facilitate and support a student's use of the environment and improved learning outcomes (Al-Samarraie & Saeed, 2018; Lamb, Cavagnetto, & Akmal, 2016; Lamb, 2017). Adaption of content and the environment can occur at multiple points through interactions with the content and environment (Dumford & Miller, 2018). Each of the adaptions is predicated on the ability of the SALEs to identify the student's cognitive state from data, select the appropriate change in time for learning, and accurately predict outcomes related to the student. Adaption of content during the process of learning is difficult because of the complexity of assessing student cognitive states using content data and results from testing (Thees et al., 2020). It is even more complicated by the need for processing of the data in real-time. Real-time refers to a system in which data is processed and available for use within milliseconds for feedback. In many cases, content data does not accurately reflect levels of demand, confusion related to their learning, or other aspects of cognitive state that influence outcomes. In many cases the results of current predictive models exhibit accuracy rates in the 50%–60% range and require minutes to days to be properly analyzed for decisions related to adaption of contnet (Thees et al., 2020). One means to address the lack of accuracy and the need for real-time data analysis, is through real-time examination of a student's cognitive state using neurocognitive data measured through a device such as fNIRS while the student interacts with content. For example, functional near infrared spectrometer (fNIRS) data when combined with a machine learning classifier such as an artificial neural network (ANN) are available for use by the adaptive systems within 300 ms (Abdalmalak et al., 2020). fNIRS is a device which measures blood flow around the brain to a depth of 10 mm due to neuronal activity related to a specific task or tasks. Understanding cognitive states in real-time with only a 300 ms latency allows semi-instantaneous (e.g., within milliseconds as opposed to second or minutes) adaption of the learning system and appropriate content to be presented to the student. Using non-invasive functional neurological measurement such as fNIRS and machine learning classifiers can allow educators to use SALEs to adapt content more effectively and, in less time, than SALEs using other forms of data.

While research around learning outcome prediction has been on going, recent technological advances within the last ten years have adjusted the focus to making use of intelligent systems for the prediction of student performance and identification of individual differences in student outcomes (Brooker, Corrin, De Barba, Lodge, & Kennedy, 2018; Musso, Hernández, & Cascallar, 2020). This is particularly true when combining tools such as fNIRS which, produce the needed volume of data, with data intensive prediction tools such as machine learning, artificial intelligence, data mining, and learning analytics. Table 1 illustrates a summary of studies from the last decade examining specific data sources, prediction accuracy, and other key findings related to each of these prediction model development tools. A recent systematic analysis by Namoun and Alshanqiti (2020) examined several studies ($k = 67$) to identify the primary prediction method and data sources. A key finding of this synthesis is that models averaged a prediction accuracy across all methods of 55.4% with adaption occurring slowly over hours

Table 1
Summary of studies.

Prediction Type Method	Number of Studies	Prediction Accuracy	Major Data Types	Major Identified Weakness for this Class of Studies
Data Mining	72	Moderate (~60%)	Homework assignments, classroom assessments, and end of course tests.	(1) Singler type of learning system; (2) Did not forecast student outcomes; (3) Did not compare models.
Algorithm	88	Low (~40%)	Learning system data, student responses, and student survey results	(1) Did not compare models; (2) Data quality.
Machine Learning	240	High (~70%)	Student classroom assignments, Learning management system data, student behaviors	(1) Did not focus on academic outcomes; (2) Most papers were not focused on student level outcomes only factors; (3) Did not triangulate findings.
Traditional Student Performance Metrics	366	Low (~30%)	Academic outcomes, non-academic outcomes, homework assignments, classroom assignments, and assessments.	(1) Data quality was not consistent; (2) Did not assess multiple models; (3) Small sample sizes.

Note: Synthesized from Namoun & Alshanqiti, 2020, Polyzou & Karypis, 2019, and Shahiri & Husain, 2015

of use. These finding were also consistent with findings by Polyzou & Karypis, 2019 and Tatar & Düştegör, 2020. This indicated that while the methods may be robust there is a significant need to increase feedback (Cavalcanti et al., 2021), accuracy, and speed (Elbadrawy et al., 2016). Research conducted by Chen, Xie, Zou, and Hwang (2020) illustrated several gaps within the study of artificial intelligence (AI) for educational purposes. This work shows-that there is an increasing interest in the impact of Artificial Intelligence in Education (AIEd)- that little research in AIEd has been conducted, advanced AI and measurement technologies are rarely adopted, and AIEd technologies are rarely accounted for in educational theories. With these gaps clearly identified by the Chen et al. (2020) study, a series of recommendations to the field have been made. These recommendations among others include: (1) exploration of AI use in classroom environments; (2) the need for closer alignment of the relationship between student responses and responses within intelligent systems; and (3) the need to make use of neurocognitive technologies for measurement and data inputs such as fNIRS, EEG, or other forms of psychophysiomeasurement; and (4) identification of how to apply data from these measurements during the learning process. In considering the use of neurocognitive technologies in combination with AI, Carvalho, Martinez-Maldonado, Tsai, Markauskaite, and De Laat (2022) cautions that it is necessary to protect student agency and social wellbeing and ensure teachers and educational professionals should be empowered to develop pedagogical practices. When using AI for educational decisions and framing it is necessary to emphasize humanistic approaches in the development of AI based education around learning (Carvalho et al., 2022). Further examination of the synthesized papers illustrates that data sources were typically retrospective in nature, product data, and did not capture in-process fluctuations as the student's completed tasks or activities (Tatar & Düştegör, 2020; Magalhães, Ferreira, Cunha, & Rosário, 2020; Moreno-Marcos, Pong, Munoz-Merino, & Kloos, 2020; Kabudi, Pappas, & Olsen, 2021). The lack of process data when using prediction techniques associated with learning analytics and educational data mining, does not allow for the identification of critical changes as they happen (Hasan, Palaniappan, Raziff, Mahmood, & Sarker, 2018; Lemay, Baek, & Doleck, 2021). For example, a student who is working on understanding a graphic presented on a digital platform may have several fluctuations in their cognition as they process the graphic and attempt to respond to questions on an assessment. Since the assessment only captures the student's overall ability to process the graphic and is retrospective, it is difficult to identify at which points the student had trouble processing the graphic. In addition, many cognitive processes used in learning are automatic and are not available for introspection so the student may not be able to identify them even when asked. In this light, educators miss substantial amounts of data and identification of points at which adaption can occur. The lack of ability to capture in-process data, data with sufficient volume, and quality are three of the principal reasons that models identified in these study results showed reduced prediction accuracy. Despite the promise of predictive models in digital environments their potential has not been fully realized. One mode of data that has not been explored and has potential to address some of the identified shortcomings of current modeling approaches is neurocognitive data. Neurocognitive data provides means to increase model prediction accuracy, capture in-process task completion, and reduce the time between data capture and adaption.

The aim of this study is to examine how neurocognitive data collected via functional near infrared spectroscopy (fNIRS) may be used to create predictive models of student outcomes with greater speed and accuracy when using a synthetic adaptive learning environment (SALEs). Specifically, this study examines the utility of using neurocognitive data to develop student response prediction on a science content test. Research Question 1 for this study is, does neurocognitive data taken via fNIRS while a student engages with science content lend itself to machine learning classifications? Research Question 2 for this study is, does analysis of hemodynamic response data predict student outcomes on a multiple-choice embedded content test in a SALE? Consideration of the research questions suggests the following hypotheses. Hypothesis 1 the quantified and continuous nature of the neurocognitive data collection during content presentation will be of sufficient quantity and quality to allow classification via machine learning algorithms. Hypothesis 2, hemodynamic data collected from the fNIRS will be predictive of correct and incorrect student responses using an automated machine learning algorithm. Substantiation of these hypotheses will illustrate the potential of using neurocognitive data to drive SALE adaptability.

Hemodynamic data in this study is a specific form of neurocognitive data collected during the activation of neurocognitive process associated with the metabolism of oxygen by neural tissue as the tissue is recruited to complete specific tasks. Hemodynamic responses are the rapid delivery of oxygenated blood to active neuronal tissues (Kisler et al., 2018). As the tissue receives the blood the hemoglobin is deoxygenated to drive neuron metabolism as information related to a task is processed. The ratio between oxygenated and deoxygenated blood is an indicator of relative demand of the cognitive system associate with the process as the student completes the task (Curtin & Ayaz, 2018; Lamb, Hand, & Yoon, 2019; Lamb et al., 2018). Neurocognitive data is data which is derived from the measurement of neural processes such as oxygenation and deoxygenation of blood. The neurons and processes are a part of the structures involved with cognition. These processes and structures are important because they offer a means to examine regions of interest related to specific structural and functional activities during task completion. The processes and related measurements produce high quality data of sufficient volume for analysis and prediction. Neurocognitive data when collected during clearly defined tasks is related to specific brain structures and neurocognitive functions. This task-structure-function relationship arises from the interaction between the learning tasks, the cognitive characteristics of the learner, and the specific structures of the brain. This relationship and the ability of functional brain measures to sample neural processes multiple times a second allows neurocognitive measures to capture moment-to-moment fluctuations in human cognition. More traditional forms of data collection such as test questions, student responses, mouse clicks, etc. do not account for individual moment-to-moment individual variation, are retrospective, and as a result miss key data. Using neurocognitive data allows examination of differences which would only be visible in the moment-to-moment cognitive demand fluctuations. One way to detect these fluctuations is by measuring changes in neurocognitive data found in student cognition as the student completes the task i.e., real-time monitoring. Real-time monitoring is a system of data collection in which the input data is processed and available for immediate use as feedback and adaption within milliseconds, increasing the speed at which a learning system can adapt.

2. Literature review

2.1. Theoretical framework

The underlying framework describing hemodynamic responses to understand student cognitive processes is the Brain Microstate Framework (Lehmann, Pascual-Marqui, & Michel, 2009). Specifically, neuronal tissue responses examined in the brain microstate as measured by fNIRS consist of time blocked measures (within 0.165 seconds) of oxygenation and deoxygenation of hemoglobin in neural tissue associated with the cognitive systems being used by the brain to complete the task. The microstate brain activations are the time-limited information moments related to specific cognitive processes stimulated by specific tasks such as reading, watching a video, or answering a question (Papo, 2013). Despite the promise of this framework for the prediction of student outcomes, it has not been applied directly to predictions of student learning. When tied to a functional task such as a science task, the temporal sequence of the task and subsequent hemodynamic response

form the core of the measurements. Within this framework, the initial response-stimulus activation complex occurs with a consistent latency due to the movement of blood to the neuronal tissue of interest (Zohdi, Scholkmann, & Wolf, 2021). Once the neuronal tissue is activated the fluctuations in oxygenation and deoxygenation can sustain a signal for 0.5–6 seconds as tissue demands change during cognitive processing (Verriotis et al., 2016). fNIRS is able to measure changes in oxygenation and deoxygenation within 0.10 seconds–0.25 seconds. Considering the signal decay rate for the hemodynamic response is 0.5–6 seconds it is possible to capture changes in the hemodynamics particularly in areas of the prefrontal cortex.

2.2. Prefrontal cortex

The Prefrontal Cortex (PFC) is the major driver of conscious cognitive control and is tightly tied to learning in general and learning in science specifically. The PFC is the portion of the cerebral cortex covering the front of the frontal lobe brain. It is the area which partially responsible for processing external information and social behavior using the somatosensory systems and cortical and subcortical motor systems (Friedman & Robbins, 2022). The PFC is also the structure responsible for the processes associated with working memory and executive function (Lamb et al., 2018, 2019). When these systems are used in conjunction with long-term memory systems associated with the limbic and midbrain structures, this action contributes to creation of affect, memory, and development of behavioral actions seen in learning. Directly measuring PFC hemodynamic activity allows educators to access and examine underlying markers of information processing as a student engages in learning tasks across all major forebrain systems. Using information across multiple brain systems assists the researchers in synthesizing information to predict how a student structurally and functionally processes science content. Examination of the PFC during task completion also allows researchers to see the moment-to-moment fluctuations and how students use information from within the environment (Zawacki-Richter, Marín, Bond, & Gouverneur, 2019). Work by McGuire and Botvinick (2010) illustrates that the PFC is responsible for transferring knowledge about learning in task completion. Lamb and Etopio's (2019) study also shows that tasks involving substantial amounts of unstructured processing, such as in the conditions in this study, may be challenging and illustrate elevated levels of cognitive demand. This increased difficultly may arise from the generation of less dynamic responses within the PFC due to the lack of well-defined and specific tasks.

The activation of the neural tissue is related to the stimulus through correlation and examination of repeated responses to the specific stimulus, in this case continued actions involving the VR or video task. Importantly, both the individual response and the global patterns are task specific, repeatable, and consistent across populations i.e., if both tasks are critical thinking tasks, the activation will be identical apart from the intensity as one task may be more demanding than another resulting in greater hemodynamic response. While the hemodynamic responses can range in intensity and onset, the fNIRS sampling rate is from 4Hz to 10Hz (compared to functional magnetic resonance imaging (fMRI) .5Hz–1Hz and electroencephalography (EEG) (256Hz–1024Hz) making it possible to include and capture hemodynamic fluctuations within neural tissue (Cui, Bray, & Reiss, 2010; Lamb, Hand, & Yoon, 2019). An EEG is a device which is used to measure electrical activity across the surface of the brain. In contrast to and EEG, an fMRI measures the blood flow to any part of brain due to increased neuronal activity.

Within this study the hemodynamic fluctuations were measured over the span of the task, rapidly providing important information about the moment-to-moment changes as they occur throughout the task allowing educators to understand how the parts of a task relate to the whole of the task. A critical aspect of the stimulus-response complex is that the response is not present when the student is not actively working on the task. The stimulus-response complex is the related group of stimuli resulting from a specific task as it is combined with a unique hemodynamic response pattern. If the response is present while the student is not working on the task, then the response and task are considered unrelated and not a part of a unique stimulus-response complex. For this reason, the researchers made use of a baseline measure and null condition to measure neural activity without the presence of the task. fNIRS has been shown to provide timely localized region of interest information related to activities such as language mapping (Janecek et al., 2013), mapping of written word generation (Watanabe et al., 1998); and specific and general neurocognitive functioning related to several types of functional educational tasks (Hong & Yaqub, 2019). As hemodynamic responses and task completion occurs in the timeframe of seconds to minutes in this study, fNIRS is capable of acquiring data across the whole task and the subcomponents of the task.

A technique used to mitigate concerns related to overlapping signals is the examination of the hemodynamic response over the length of the task using a moving average. The moving average is calculated for the optodes illustrating hemodynamic response ratios above baseline for the time segment in which the student is working on the task. An optode is a sensor which can measure the concentration of a substance using light, in the case of this study, oxygenated and deoxygenated blood in brain tissue. While this reduces some of the ability to resolve specific hemodynamic responses associated with task subcomponents, it provides an overall measurement of average cognitive demand while the student completed the complex task. The incorporation of a second baseline after the removal of the stimulus allows sufficient time for the signal to stabilize back to baseline levels ensuring the task is responsible for the observed response. Several studies have validated the use of fNIRS for the measurement of oxygenation, deoxygenation, and total hemoglobin concentrations using comparisons to functional magnetic resonance imaging (fMRI) of the PFC (Maggioni, Bellani, Altamura, & Brambilla, 2016). When examined as a region of interest (as in this study) fMRI studies and fNIRS studies illustrate a strong significant correlation up to $r = .83$ (Maggioni et al., 2016) between fMRI data and fNIRS measures of deoxygenated hemoglobin (Cui et al., 2010; Maggioni et al., 2016).

3. Methods

This study used a mixed, blocked-event, counterbalanced design as illustrated by Petersen and Dubis (2012). The three conditions were video on a television screen, virtual reality, and no stimulus (null) all with embedded content questions. After students were prescreened by the researcher, the fNIRS sensor was placed on the student. During the prescreening process the student was told they may ask any questions about the study that they had. The first part of the design is an event related response. An event related response is a response in which a change in hemodynamic response, as a measure of student cognitive activity results directly from a sensory, cognitive, or motor event such as watching a portion of a video or answering a question. The second part of the design was a mixed block approach, in which a Baseline 1 (A-condition), is followed by a Stimulus (B-condition), followed by Baseline 2 (A-condition). The baseline condition "A" consisted of participants sitting quietly and is a "neural rest condition." The stimulus condition "B" had the participants complete a science task and was the "neural active condition." The A_1-B-A_2 approach increased the robustness of the results, increased statistical power, identifies the stimulus-response complex, and provided a within-subject control through examination of changes related to each participant associated with each of the baselines. Mirroring an approach commonly used in EEG studies, this study design allows for detection of moment-to-moment neurocognitive responses. It is important to note that while there are parallels in EEG and fMRI research to fNIRS research, the techniques and activities associated with signal analysis and signal acquisition differ. Responses in which multiple cognitive systems correlate to the tasks can be differentiated through the Hemodynamic Response Function (Seghouane & Ferrari, 2019). Event related

responses allow analysis of individual student's responses as they watch a video, answer a question, and the correlation of the subsequent behaviors related to task completion. For example, the portion of the VR presentation on the opening of the deoxyribonucleic acid (DNA) helix in preparation for replication is treated as a stimulus-response complex allowing researchers to gather information as the student completes the task. DNA is the molecule inside cells which is responsible for storing genetic information for an organism and concepts related to DNA replication are typically taught as a part of the biology curriculum in tenth grade. When treated as a block, the hemodynamic response fluctuations allow the measurement of non-sequential completion of the tasks, timing of stimulus, examination across the whole task, and allows students to work on tasks as they would in the classroom. In addition, because of the use of fiber optics and light the fNIRS is more robust to signal interference and movement artifacts when compared to both EEG and fMRI making it a better measurement device for educational settings (Pinti et al., 2020).

3.1. Data collection

Data acquisition occurred via the fNIRS optodes. Data was aggregated and initial processing happened with *COBI Studio software* version 1.3.0.19. Filtering, data preparation, and signal analysis for statistical examination of data happened with *fNIRS Soft Professional* 4.10. Neurocognitive data consisted of fNIRS hemodynamic responses illustrated in composite images identifying location and intensity via colorized images. The neurocognitive data also included composite images, video, and numeric oxygenation and deoxygenation ratios. A MP160 device was used to synchronize optode signals, video, and student responses. The operator of the fNIRS observed each of the students ($n = 40$) as they worked on the tasks over the span of the study and monitored the fNIRS device placing markers as events occurred (e.g., when the student started and ended the task). The students were not able to see the outputs as the operator was monitoring and marking. The operator marked the beginning and end of baseline data collections to ensure that hemodynamic responses were stable. While the authors acknowledge that classroom conditions are variable, baseline is used to standardized starting points for the within student comparisons. A second baseline allows the researcher to identify if the stimulus and response are connected within the hemodynamic response and not an artifact associated with another classroom activity. It is expected that once the task is completed and the student is sitting quietly that the hemodynamic response would return to baseline. Video of the students completing the tasks was taken to allow researchers to determine if the markers associated with each task were accurately placed. Fig. 1 illustrates the position of optodes and emitters for the participants.

3.2. Participants

Participants in this study were (n = 40) neurocognitively healthy right-handed randomly selected ninth grade students from four high schools ($N = 2096$), 21 males and 19 females taking part in a ninth-grade Earth Science class. The students were naïve to the content and questions used in the null, video, and VR presentations. Students were randomly selected from each school, 10 students per school, from all ninth-grade students taking Earth Science ($N = 1157$). Selection occurred using a random number generator to select a student to recruit. Recruitment occurred in three waves until 40 students agreed to participate. Students were contacted through letters home, classroom visits, and teacher meetings. All 40 participants were from a mix of urban and rural schools. Students gave assent while parents provided informed consent. This project and its procedures are in keeping with the Declaration of Helsinki and other ethical standards. The average age of the participants is 14.6 ($SD = 0.4$). Each participant is at current grade level related to mathematics proficiency, English language proficiency, and reading proficiency. Levels of proficiency were used to ensure that additional cognitive responses beyond the task were not derived from processing difficulties associated with ability in these areas. The researchers pre-screened participants using the Woodcock-Johnson IV Achievement Test (alpha = .90) in the areas of Calculation, Applied Problem Solving, and Quantitative Concepts (Reynolds & Niileksela, 2015). Reading levels were assessed using the Wide Range Achievement Test Third Edition (alpha = .75) (Wilkinson & Robertson, 2006). Participants were identified as neurotypical through extensive interviews and review of histories as suggested in the *Compendium of Neuropsychological Tests* (Strauss, Sherman, & Spreen, 2006). No participant was removed based upon screening. Each participant was on grade level in their current science class, obtaining passing grades, having passed their

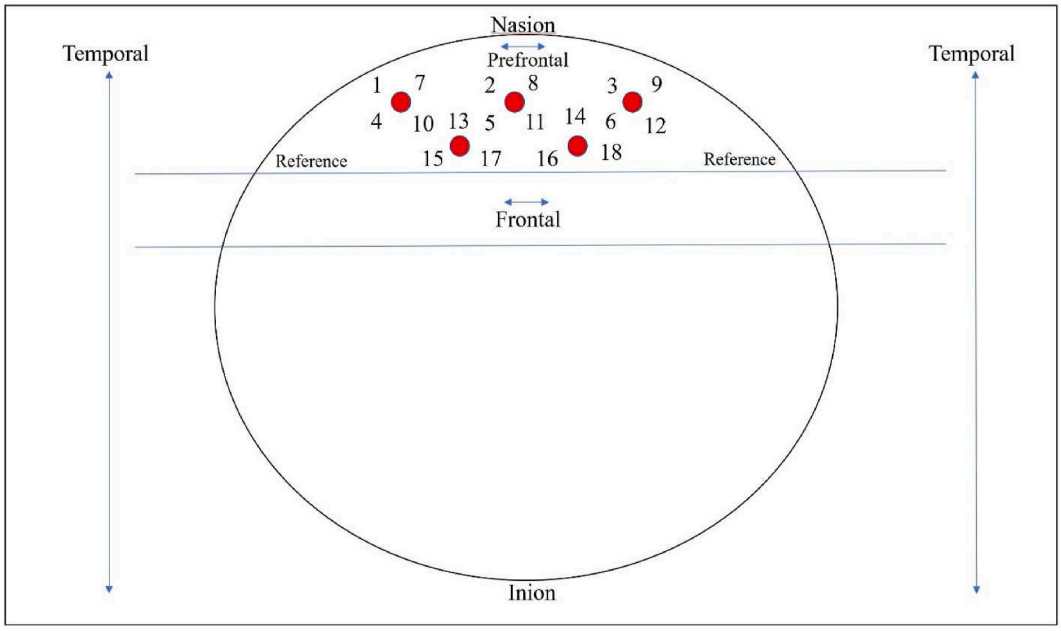


Fig. 1. locations of optodes (sensors) and emitters. Optodes are identified by numbers and emitters are the red circles. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

previous science classes with a "C" or better, and within the top 50% of their current class. Biology content was specifically chosen because, while the students did recall some aspects of the topics, the students did not recall details associated with the topics and the tasks were novel to the students. Fifty-three percent of the students had taken Life Science in previous years (seventh grade year) Fall semester and 47% had taken Life Science in the previous years (seventh grade year) Spring semester. None of the students had taken biology, however all the participants had taken eighth grade physical science.

3.3. Conditions

Within this study, neurocognitive data from the PFC as students engaged with each of the conditions was used to create a prediction of student responses on a science content test. At specific times during the video, VR, and null condition, questions about presented content were given to the students (Cronbach's Alpha = .87). The recorded video condition is a 20-min explanation of DNA replication. During the video instruction, a female instructor provides an explanation of DNA replication. In addition to verbal instruction, pictorial representations were also shown. The video contained multiple (3) color graphics displayed on the screen replacing the instructor. The VR condition consisted of a 20-min 3D immersive video of DNA replication with three dimensional visualizations. During VR instruction, a voice over discussed DNA replication. In addition to verbal instruction, pictorial representations were also shown. The VR content also contained multiple (3) color graphics displayed on the screen replacing the instructor. Students taking part in this condition did not have the ability to interact with the content other than to "walk around" the representations. The lack of interaction was intentional to create parity between the VR and video conditions. The null condition consisted of no content being presented to the students, however, at the same intervals as the video and VR sessions, questions appeared related to the content. Fig. 2 illustrates the timing of the content and the relative amount of content that specifically relates to the question. The time above the content illustrates the time at which the content related to the question started.

3.4. Data analysis

Prior to full analysis of the neurocognitive data there is significant data preprocessing which must occur. Data preprocessing initially starts with removal of gross movements, movements due to respiration, and heart pulsations (Pinti et al., 2020). Artifacts were removed using a 0.14HZ cutoff low pass filter as suggested by Nguyen, Yoo, Bhutta, and Hong (2018). Filtering data using a 0.14Hz low-band filter resulted in a loss of 7% of the data for the VR condition, 9% loss of data for the video condition, and 9% loss for the null condition. In addition to removing movement artifacts, extracranial and extracerebral contributions to the fNIRS signal were separated via regression on a per optode basis. This separation resulted in a total loss of 5% of the data.

Ratios of the concentrations of oxygenated and deoxygenated blood were converted to standardized Z-scores with respect to Baseline 1 to allow for comparison across tasks and individuals. A moving mean (average) was calculated for each of the participants based upon data as they were completing the task. The moving mean statistically smoothed short-term hemodynamic response spikes and filtered-out signal noise ensuring large variations in hemodynamic response did not overweight the analysis.

A mixed model analysis of variance (MX ANOVA) and post-hoc planned comparison by task was conducted using SAS JMP Pro 14. A MX ANOVA is an analysis of variance technique using a mix of a between subject and within subject comparisons with two or more categorical independent variables. The MX ANOVA in this study was used to determine if there are significant differences between correlated means across optodes and tasks (Boisgontier & Cheval, 2016). Specifically, the MX ANOVA was used to examine measures over each of the time points; Baseline 1 (A_1-condition), Stimulus (B-condition), and Baseline 2 (A_2-condition) across conditions. This approach allowed the researchers to examine which optodes illustrated greater hemodynamic response when compared to Baseline 1 and 2. The MX ANOVA and partial eta-squared are used to examine each of the participants per task to determine the effects of the condition on hemodynamics for specific optodes. MX ANOVA is robust to unbalanced repeated measures and is indicated because of the hierarchical clustered nature of the data with time points and A_1-B-A_2 conditions clustered by student. Post-hoc (Tukey HSD) analysis was run to determine which of the treatments illustrated the greatest hemodynamic response. A second post-hoc analysis (Tukey HSD) was used to determine which optodes illustrated hemodynamic response greater than each of the baselines (1 and 2) for each task to identify optodes which correspond to task related activations. A Bonferroni correction was conducted at an alpha equal to 0.05/c where c is the number of comparisons ($c = 7$). Significant p-values must be below 0.007 for the seventh comparison. Lastly correlational analysis was conducted between outcomes on the tasks to ensure task completion was related to the hemodynamic response.

3.5. Predictor development

Predictors were developed using a machine learning algorithm applied to the two data sources, hemodynamic response collected during each condition and test question responses. Analysis of the standardized hemoglobin absorption ratios ($i = 864{,}000$) between the oxygenated hemoglobin and deoxygenated hemoglobin occurred using MX ANOVA. The authors used a Random Forest model along with penalized logistic regression to generate confusion matrices. These two approaches helped the researchers to identify the underlying structure of existing data and generate rules for prediction of observations.

Data analysis in this portion of the study was used to identify the specific node weightings and model with the best accuracy and generalized model fit. Best accuracy and general model fit arise from the predicted test data with the least error. Models that illustrate best data and conceptual fit are identified using multiple measures. The two primary measures are Conforming Capability (Conforming = MMSEtr + MMSEtst) and Generalizing Capability (Generalizability = MMSEtst − MMSEtr). The minimum means square error (MMSE) is a measure of estimated quality of the dependent variable fit values. Models with highest standard errors are not considered viable models. The model with the most generalized architecture is retained after the model with the highest deviations is removed. Model architecture with the closest training and testing data is considered most generalized. Comparison of training and testing data occurs through the evaluation of means. As identified by Al-Nafjan, Hosny, & Al-Wabil (2017) and Xiao et al.

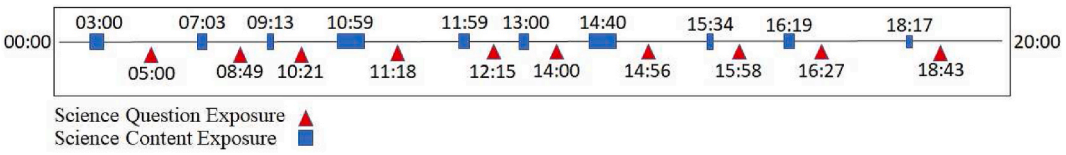


Fig. 2. Timing of content and questions used in this study for each condition.
Note. While the null condition did not have content, questions were still asked at the same time points.

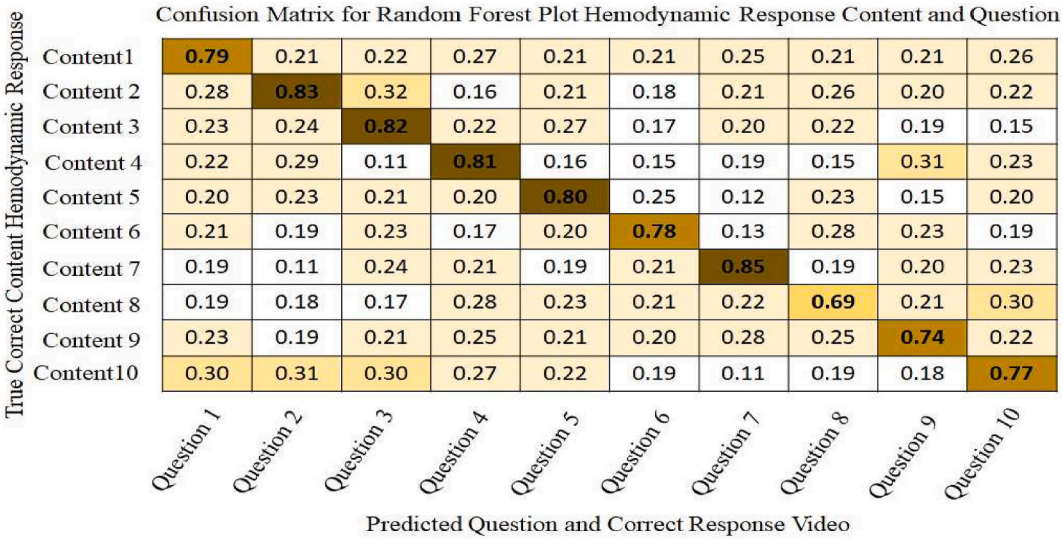


Fig. 3. Example confusion matrix for the video condition.

(2019), ANNs have been used to model cognition related to learning with good reliability. In addition, ANNs have been used with good success in predicting student performance in educational setting when given sufficient data (Rodríguez-Hernández, Musso, Kyndt, & Cascallar, 2021; Wang, Xie, Wang, Lee, & Au, 2021). The model used in this study is an error back propagation model with a Random Forest algorithm. A K-fold cross validation with random data assignment was used. The data was randomly divided into multiple similarly sized slices ($n_1 = 288{,}000$, $n_2 = 288{,}000$, and $n_3 = 288{,}000$). The segments are used for validation and training. K-fold cross validation has been used with success in engineering and other fields to support proposed predictive models (Wijayasekara, Manic, Sabharwall, & Utgikar, 2011). Analysis of each type of question and the associated difficultly occurred by relating incorrect responses with student's indications that they had lost concentration during the task.

4. Results

4.1. Summary of results

Results suggest that neurocognitive responses collected during VR and video conditions is predictive of correct responses on the content test, while signals obtained from the null condition did not predict correct responses. Automatic machine learning classification outcomes of hemodynamic patterns obtained while the students watched the video, VR, and null condition predicted correct and incorrect responses during the content test. Visualization of these outcomes are illustrated using a confusion matrix. The confusion matrix uses a Random Forest algorithm. Within the diagram each row is the predicted model, and the column is the actual students' performance. The training and test models show good model fit for the identification of actual correct and wrong answers. For example, while watching Content Selection 1, the hemodynamic response patterns were predictive of success in answering Question 1 correctly 79% of the time. The model predictions illustrated between a 69% and an 85% success rate (shown in dark brown) in accurately predicting responses on the content questions by the students. illustrates an example confusion matrix from this study (See Fig. 3)

4.2. Results of model development

The Random Forest model obtained satisfactory results with an area under the Receiver Operating Characteristic (ROC) curve of 0.79 (shown in Fig. 4). A ROC curve is a graphical representation of the diagnostic ability of a system, which plots the true positive rate against the false positive rate. The authors also provided a confusion matrix of the Random Forest model showing the instances of the prediction of correctness associated with each condition per case. Each row illustrates the predicted model i.e., hemodynamic response collected during the viewing of the content. Columns illustrate the actual students' performance on the question. The Random Forest model shows good model fit in relation to the identification of actual correct answers and incorrect answers. The random forest has the average accuracy of 0.839 ± 0.042, sensitivity of 0.73 ± 0.071, specificity of 0.71 ± 0.044, and Cohen's kappa coefficient of 0.41 moderate. The authors also shows that the Random Forest model results in outcomes not by chance through a re-randomization test evaluating model information (Cruz-Martinez et la., 2022). This procedure for a re-randomization test occurs through the reshuffling of the model with variance as the outcome variables and re-calculating the area under the curve for the ROC curve (See Fig. 4) using 1000 iterations. The number of cases resulting in a better rival model versus the selected Model outputs and identified the fNIRS optodes that resulted in the best prediction for student participants. These results replicate and support results obtained by Oku and Sato

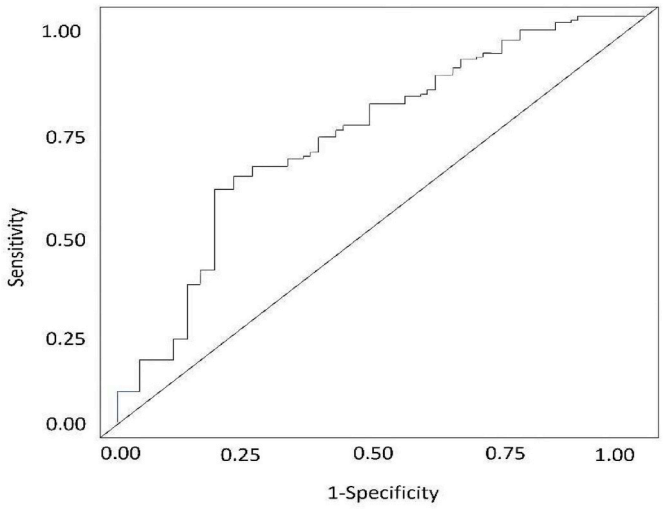


Fig. 4. The composite ROC curve for the Random Forest, plotting sensitivity versus false positive rate at thresholds across conditions.

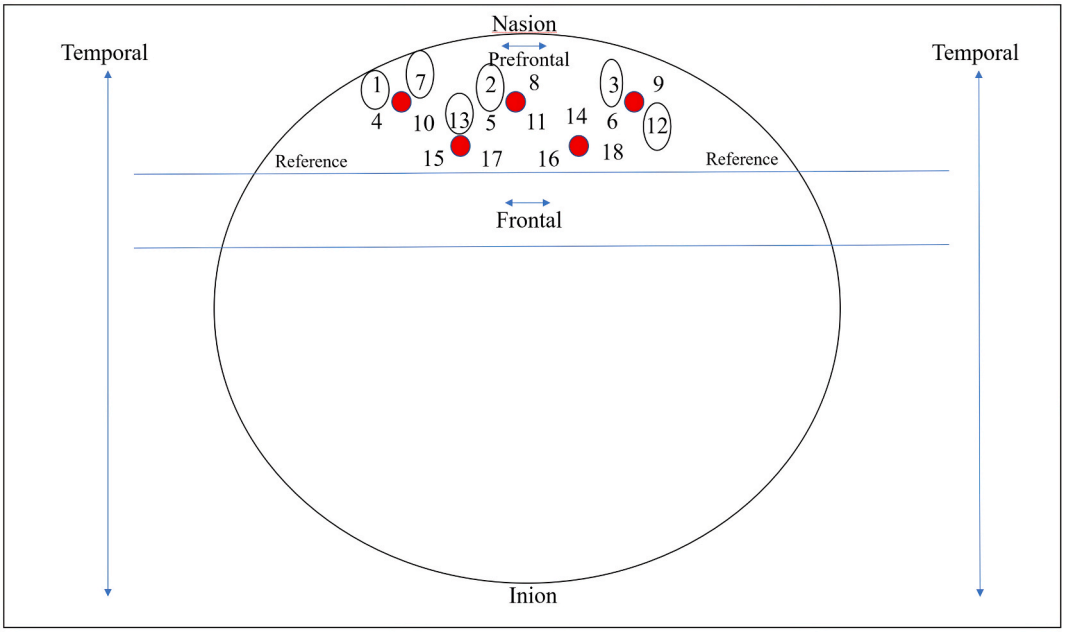


Fig. 5. Resultant optodes with high predictability for the VR and video conditions.

2021 (see Fig. 5).

To ensure good conceptual model fit, it was necessary to identify the neurocognitive responses which linked to each content question. To accomplish this the authors examined video of the students completing the video, VR, and the assessment ensuring that content questions by themselves were insufficient to predict correct or incorrect responses. The authors of the study also examined the frequency of the selected channels in each iteration for cross-validation of the Random Forest model. The authors validated the covariates as Channel 12 (deoxygenated hemoglobin) and Channel 1 (oxygenated hemoglobin) as having the greatest weight in the prediction of correct answers as these optodes illustrated 96% use in 83% of the participants across video and VR conditions. Across VR and video conditions the relevant optodes according to this model are 1, 2, 3, 7, 12, and 13 corresponding to the middle prefrontal and frontal cortex. Once identified changes in optode signals were visible within 300 ms–500 ms during the task. The PFC is primarily responsible for activities in learning such as the cognitive processes of working memory, cognitive flexibility, planning, inhibition, and abstract reasoning (Zgaljardic et al., 2014). Each of these cognitive processes is extremely important when engaged in the learning of science content. Please note the null condition did not exhibit any optodes which stood out and appeared to be relevant in predicting correct responses. This was expected given the lack of stimulus during this condition.

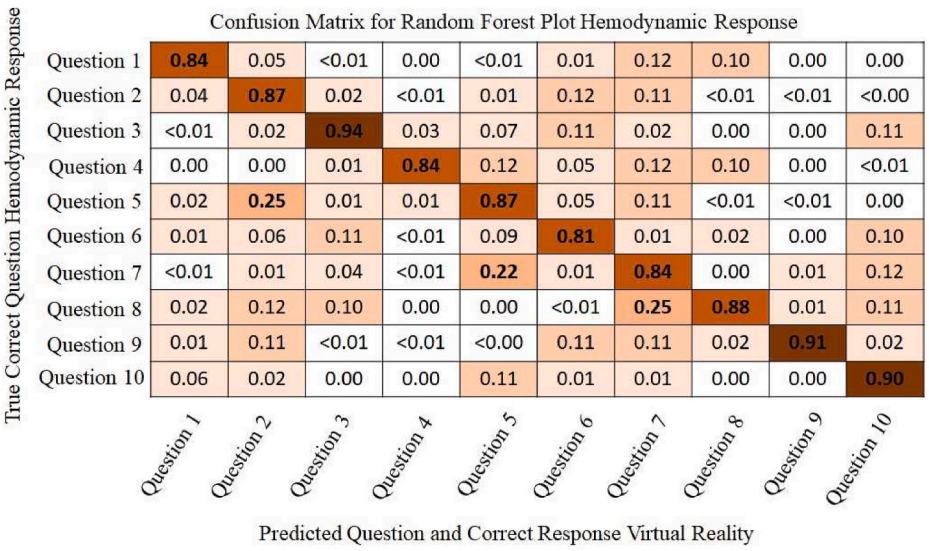
By identifying the optodes which were most explanatory of student responses (by question, by student) with a minimum of error, the authors were also able to assess the levels of student participation in each of the conditions. Fig. 6 shows differentiation by the model of correct and incorrect responses using only hemodynamic responses across questions for each condition. To this end, the developed model indicates there is a slight association between correct answers and the student belief that they engaged with the content. Examination of the confusion matrix illustrates that for the video and VR condition hemodynamic response has a 0.79 and 0.87 probability of correctly predicting outcomes on the content test. The null condition the average for prediction of correct responses was significantly lower at 0.29.

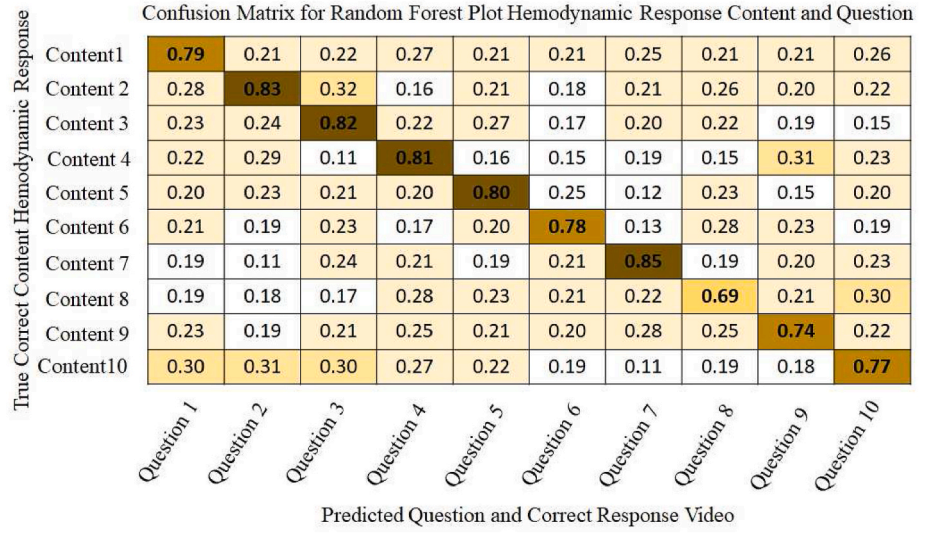
5. Discussion

This study examines how neurocognitive data in the form of hemodynamic response measures may be used to develop machine learning classifiers in real-time to create student level answer predictions as students engage with science content using a SALEs. Error rates in the prediction of student response are below 15% and illustrate a prediction rate of 85% accuracy. These outcomes illustrate that neurocognitive data can be used in real-time and is more predictive than current data sources. This form of data can be used as a cornerstone to understand the degree of student engagement with content, validate newly developed content, and examine the process of assimilation of new content in real-time during content interactions.

The primary goal was to investigate if neurocognitive data collected from the participant's PFC during content interactions could accurately and rapidly predict student answers on a content test. The authors specifically used a configuration of optodes allowing collection of neurocognitive data in a more naturalistic setting and in far more realistic situations as students used a learning system. The results provide evidence that neurocognitive data use for prediction is more accurate (average 85% accuracy) and more rapid (300 ms–500 ms) than what is currently available (average ~55% accuracy and several minutes to hours). Additional studies using high resolution neuroimaging devices may be warranted as they can more clearly identify cognitive systems and the precise locations of subsidiary areas in the brain involved during video and VR use. This study also provides evidence to support current cognitive models of how the PFC is used during learning in digital adaptive environments (Lamb, 2019; Lamb & Etopio, 2019; Lamb et al., 2018). The larger contribution of this study is the successful prediction of content test outcomes solely from a student's interaction with content using neurocognitive data. An implication of this outcome is that this type of data may be used to develop computational models for experimentation and testing of student learning interventions (Lamb 2014, 2016; Lamb & Annetta, 2013; Lamb, Annetta, Vallett, & Sadler, 2014; Lamb et al., 2014, 2014). The digital nature of the data and the speed of data collection allows neurocognitive data to be used to drive adaption of the digital environment and content potentially in milliseconds as opposed to minutes, hours, or days. This reduces the latency of the user response interface, more directly connects with neural processes used in learning, and propels the development of SALEs to increase student learning outcomes.

SALEs are at the forefront of a new generation of computerized learning systems delivering multimedia course content with the intent of





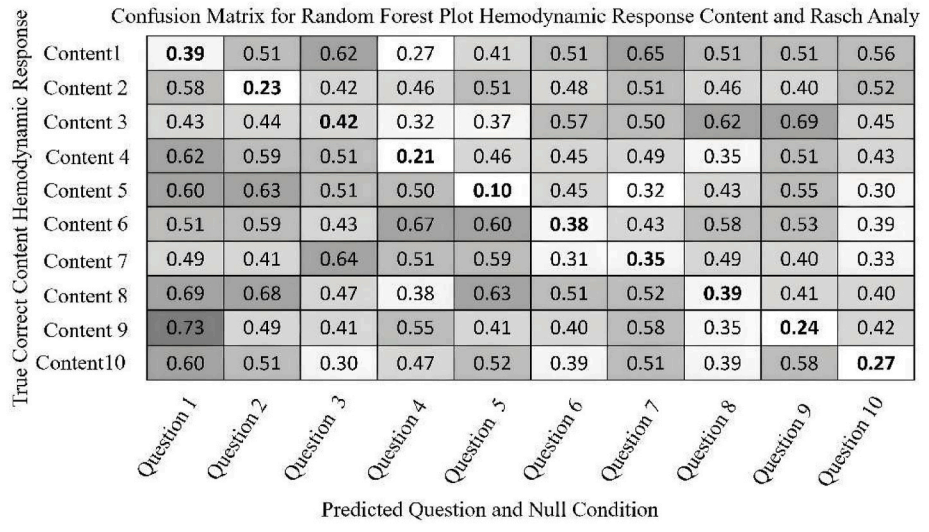


Fig. 6. Confusion matrix illustrating the relationship between content and questions.

providing learners a personalized experience. This traditionally occurs by capturing the learning system's data associated with each user's personal preferences, content knowledge levels, cognitive abilities, and other factors. In most cases this occurs through the collection of data through assessment of content before and after content interaction has occurred. Little if any consideration is given to analysis of the learning during the student's interaction with the content. More importantly much of the learner's cognitive interactions with the content are a "black box. This study illustrates the potential value of neurocognitive data collected, in real-time, during content interaction as an accurate predictive tool for assessment supporting findings by Dahlstrom-Hakki et al. (2019). Combining a SALE with neurocognitive data collection and machine learning classification during content interaction helps in adaption of content to classroom objectives.

Pelánek (2017) encountered challenges with the use of machine learning classifiers using more traditional data. This is because machine learning approaches need large training data sets to identify a student's need. In many cases when using traditional data sets, this is often unavailable. This data collection often takes the form of large numbers of content questions and other forms of data collection which is time and resource intensive for the learner and the digital system. Data taken from assessments and questionaries often involve a significant amount of time for analysis and meaning making e.g., minutes to hours, when used to identify a student's individual needs. The second issue is that the adaptive learning systems are not always able to characterize the learner's individual needs or the student's individual needs do not align with learning objectives used by the system.

These challenges illustrate the two significant problems; (1) large-scale data collection decreases the desire to use the learning system and produces excessive cognitive demand; and (2) there is a lack of technology (until recently) for identification of the learner cognitive states. Researchers have recommended that augmentation of SALEs occur through new data sources and through creation of more accurate classifiers for system training (Cui, Chen, Shiri, & Fan, 2019). Considering the needs of SALEs in terms of data, the use of neurocognitive data solves both concerns. Neurocognitive data as used in this study is of sufficient volume and resolution that machine learning classifiers, the underlying engine of adaptive learning systems, can make use of the data to predict student outcomes. In addition, this study also shows that newer technologies such as fNIRS can be used to assess student cognitive state.

5.1. Pedagogical implications

Measurement of student neurological states (i.e., attentional dynamics and cognitive dynamics) during the learning of science content allows a deeper understanding of student processes and provides critical information for how to best adapt content and questions for the student (Zhai, Yin, Pellegrino, Haudek, & Shi, 2020). Neurocognitive measurement may be used to update content in a learning environment supporting work by Rose and Strangman (2007). Collecting data about cognitive states in real-time with only a 300 ms–500 ms latency in system response allows semi-instantaneous (e.g., within milliseconds as opposed to second or minutes) adaption of the learning system. Using non-invasive functional neurological measurement such as fNIRS in conjunction with machine learning classifiers can allow educators to use SALEs to adapt learning content more effectively and, in less time, than SALEs using other forms of data.

The use of neurocognitive data for the predication of student outcomes in science education has the potential to impact the ability of SALEs to meet student needs, leading to greater individualization. The main findings of this study which impact teacher activities in the classroom are: (1) increased accuracy of the prediction and (2) the speed of prediction i.e., real-time prediction of student learning outcomes. These two findings can result in the development of new models of adaptive learning and assessment of student at risk for failure or missed learning opportunities and the provision of students with content to enhance learning outcomes automatically. Future iterations of these models using this data could provide real-time understanding of student's individual cognitive states as they teach. In addition, this type of system may be useful for assessment of non-verbal students by predicting outcomes on assessments as they engage with the content. With non-verbal students, they would not have to communicate what they learn, it may be possible to simply assess students based upon neural activity.

Content teachers can play a leading role in developing predictive models, using these models to adjust instruction, and provide more meaningful interventions for students. This will occur because the teachers, using these models, are able to more closely observe students and capture more data as they engage in the learning process. The predictive models could be used during instruction to identify specific content and modes of instruction which were effective and not effective based upon classification and ability to predict outcomes on assessments. The ability to assess effectiveness would allow the automatic modification of digital content to remediate students at home and in the classroom by potentially using decision tree learning approaches (Matzavela & Alepis, 2021). Curriculum developers may also make use of this technology to assess the quality of the content and user experiences with the content. By examining the moment-to-moment fluctuations in neurocognitive data it is possible to systematically identify the components of the curriculum and how these components create levels of risk for failure for students at varying levels.

The weakness of this approach for use in the classroom is two-fold. First, to account for and mitigate extracerebral signals while using fNIRS to collect neurocognitive data in the classroom, the predictive model uses statistical methods instead of short-range detectors to reduce invasiveness (Tachtsidis & Scholkmann, 2016). Secondly, while the collection of real-time neurocognitive data via fNIRS is less invasive than other forms of neurocognitive data collection and can occur in the classroom, it still requires a headset to be worn by students. Students taking part in this form of data collection may not be tolerant of this without further development of the technology to reduce the size of the detectors. The primary strength of this form of data is the ability to capture moment-to-moment changes in student cognition throughout a task. This provides significantly more detail about the alignment of specific cognitive systems, levels of cognitive demand, accuracy of the prediction, and can allow collection of data in real-time. These strengths will allow educators in the classroom to iterate their content and their teaching approaches more rapidly to better meet student needs.

5.2. Limitations

The finding of this study is subject to some limitations. First, our sample size was relatively small in terms of predictive studies ($n = 40$), however the amount of data per persons is fairly large, $i = 1200$ reading per person over the 20-min task ($k = 48{,}000$). Second, although a complex task was used, these measures may not be enough to cover the total range of cognitive action which may be relevant in science education or other content areas. Future studies may consider building prediction models for other tasks and content areas such as mathematics, technology, or engineering. Third, although the mixture of students is from a general education classroom and covers a relatively broad range of student performance across several science tasks, there are potential problems in using these prediction models to predict scores of students which have more neurocognitive diversity then is found in a general education classroom. Future studies should attempt to include a more diverse sample of students with neurocognitive differences. Finally, data collection in this study only examined a single video and VR content presentation. Additional work is needed to inform educators about cognition, affect, and behavior performance during each condition and their relationship to hemodynamic response.

6. Conclusion

SALEs can increase student engagement and promote individualized learning if they are able to accurately predict student outcomes. SALEs are more accurately and quickly able to predict outcomes when using neurocognitive data versus other forms of data. These systems reinforce learned concepts and provide a means to deliver on-demand learning based solely on student need without a human in the loop e.g., instructor making the decision. Empirical assessments in studies by Wachtler, Scherz, and Ebner (2018), have shown that video and related quizzes lack adaptabilities which meet all student's needs but can increase knowledge, intensify engagement, and promote attention. Using these forms of digital instruction combined with neurocognitive data may provide greater success for students. The use of neurocognitive data can act as means to drive adaptation. While there is the possibility to measure student performance via classroom assessments, the involvement of students in the execution of learning through the examination of brain states during learning does not occur in current systems. Further studies may want to consider assessing how machine learning classifiers used in SALEs can not only classify neurocognitive states, but how best to leverage these states to present content in the most coherent way for the individual students. Studies into this form of SALE may allow for greater individualization of content and learning experiences.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge the East Carolina University Neurocognition Laboratory for its support and use of the equipment.

References

Abdalmalak, A., Milej, D., Yip, L., Khan, A. R., Diop, M., Owen, A. M., et al. (2020). Assessing time-resolved fNIRS for brain-computer interface applications of mental communication. *Frontiers in Neuroscience, 14*, 105.

Al-Nafjan, A., Hosny, M., & Al-Wabil, A. (2017). Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review. *Applied Sciences, 7*(12), 1239.

Al-Samarraie, H., & Saeed, N. (2018). A systematic review of cloud computing tools for collaborative learning: Opportunities and challenges to the blended-learning environment. *Computers & Education, 124*, 77–91.

Alexander, B., Ashford-Rowe, K., Barajas-Murph, N., Dobbin, G., Knott, J., McCormack, M., … Weber, N. (2019). *Horizon report 2019 higher education* (edition). EDU19.

Boisgontier, P, & Cheval, B (2016). The anova to mixed model transition. *Neuroscience & Biobehavioral Reviews, 68*, 1004–1005.

Brooker, A., Corrin, L., De Barba, P., Lodge, J., & Kennedy, G. (2018). A tale of two MOOCs: How student motivation and participation predict learning outcomes in different MOOCs. *Australasian Journal of Educational Technology, 34*(1).

Carvalho, L., Martinez-Maldonado, R., Tsai, Y. S., Markauskaite, L., & De Laat, M. (2022). How can we design for learning in an AI world? *Computers and Education: Artificial Intelligence, 3*, 100053.

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., et al. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence, 2*, 100027.

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100002.

Cruz-Martinez, C., Reyes-Garcia, C. A., & Vanello, N. (2022). A novel event-related fMRI supervoxels-based representation and its application to schizophrenia diagnosis. *Computer Methods and Programs in Biomedicine, 213*, 106509.

Cui, X., Bray, S., & Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *NeuroImage, 49*(4), 3039–3046.

Cui, Y., Chen, F., Shiri, A., & Fan, Y. (2019). *Predictive analytic models of student success in higher education: A review of methodology*. Information and Learning Sciences.

Curtin, A., & Ayaz, H. (2018). The age of neuroergonomics: Towards ubiquitous and continuous measurement of brain function with fNIRS. *Japanese Psychological Research, 60*(4), 374–386.

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.

Dumford, A. D., & Miller, A. L. (2018). Online learning in higher education: Exploring advantages and disadvantages for engagement. *Journal of Computing in Higher Education, 30*(3), 452–465.

Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. (2016). Predicting student performance using personalized analytics. *Computer, 49*(4), 61–69.

Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology, 47*(1), 72–89.

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction, 28*(2), 127–203.

August Hasan, R., Palaniappan, S., Raziff, A. R. A., Mahmood, S., & Sarker, K. U. (2018). Student academic performance prediction by using decision tree algorithm. In *2018 4th international conference on computer and information sciences (ICCOINS)* (pp. 1–5). IEEE.

Hong, S, & Yaqub, M (2019). Application of functional near-infrared spectroscopy in the healthcare industry: A review. *Journal of Innovative Optical Health Sciences, 12*(6), 1930012.

Janecek, J. K., Swanson, S. J., Sabsevitz, D. S., Hammeke, T. A., Raghavan, M., Rozman, M. E., et al. (2013). Language lateralization by f MRI and Wada testing in 229 patients with epilepsy: Rates and predictors of discordance. *Epilepsia, 54*(2), 314–322.

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence, 2*, 100017.

Khan, M. A., & Khojah, M. (2022). *Artificial intelligence and big data: The advent of new pedagogy in the adaptive e-learning system in the higher educational institutions of Saudi Arabia*. Education Research International, 2022.

Kisler, K., Lazic, D., Sweeney, M. D., Plunkett, S., El Khatib, M., Vinogradov, S. A., … Zlokovic, B. V. (2018). In vivo imaging and analysis of cerebrovascular hemodynamic responses and tissue oxygenation in the mouse brain. *Nature Protocols, 13*(6), 1377–1402.

Lamb, R. Successful use of a novel artificial neural network to computationally model cognitive processes in high school students learning science.

Lamb, R. (2014). Examination of allostasis and online laboratory simulations in a middle school science classroom. *Computers in Human Behavior, 39*, 224–234.

Lamb, R. L. (2016). Examination of the effects of dimensionality on cognitive processing in science: A computational modeling experiment comparing online laboratory simulations and serious educational games. *Journal of Science Education and Technology, 25*(1), 1–15.

Lamb, R., & Annetta, L. (2009). A pilot study of online simulations and problem based learning in a chemistry classroom. *Journal of Virginia Science Educator, 3*(2), 34–50.

Lamb, R., & Annetta, L. (2012). Influences of gender on computer simulation outcomes. *Meridian, 13*(1).

Lamb, R. L., & Annetta, L. (2013). The use of online modules and the effect on student outcomes in a high school chemistry class. *Journal of Science Education and Technology, 22*(5), 603–613.

Lamb, R., Annetta, L., Hoston, D., Shapiro, M., & Matthews, B. (2018). Examining human behavior in video games: The development of a computational model to measure aggression. *Social Neuroscience, 13*(3), 301–317.

Lamb, R., Annetta, L., Vallett, D., Firestone, J., Schmitter-Edgecombe, M., Walker, H., … Hoston, D. (2018). Psychosocial factors impacting STEM career selection. *The Journal of Educational Research, 111*(4), 446–458.

Lamb, R. L., Annetta, L., Vallett, D. B., & Sadler, T. D. (2014). Cognitive diagnostic like approaches using neural-network analysis of serious educational videogames. *Computers & Education, 70*, 92–104.

Lamb, R., Antonenko, P., Etopio, E., & Seccia, A. (2018). Comparison of virtual reality and hands on activities in science education via functional near infrared spectroscopy. *Computers & Education, 124*, 14–26.

Lamb, R., Cavagnetto, A., & Akmal, T. (2016). Examination of the nonlinear dynamic systems associated with science student cognition while engaging in science information processing. *International Journal of Science and Mathematics Education, 14*(1), 187–205.

Lamb, R. L., & Etopio, E. (2019). Virtual reality simulations and writing: A neuroimaging study in science education. *Journal of Science Education and Technology, 28*(5), 542–552.

Lamb, R. L., Etopio, E., Hand, B., & Yoon, S. Y. (2019). Virtual reality simulation: Effects on academic performance within two domains of writing in science. *Journal of Science Education and Technology, 28*(4), 371–381.

Lamb, R., Hand, B., & Yoon, S. Y. (2019). An exploratory neuroimaging study of argumentative and summary writing. In *Theorizing the future of science education research* (pp. 63–82). Cham: Springer.

Lamb, R. L., Vallett, D. B., Akmal, T., & Baldwin, K. (2014). A computational modeling of student cognitive processes in science education. *Computers & education, 79*, 116–125.

Lehmann, D., Pascual-Marqui, R. D., & Michel, C. (2009). EEG microstates. *Scholarpedia, 4*(3), 7632.

Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence, 2*, 100016.

Mädamürk, K., Tuominen, H., Hietajärvi, L., & Salmela-Aro, K. (2021). Adolescent students' digital engagement and achievement goal orientation profiles. *Computers & Education, 161*, 104058.

Magalhães, P., Ferreira, D., Cunha, J., & Rosário, P. (2020). Online vs traditional homework: A systematic review on the benefits to students' performance. *Computers & Education, 152*, 103869.
Maggioni, E., Bellani, M., Altamura, A. C., & Brambilla, P. (2016). Neuroanatomical voxel-based profile of schizophrenia and bipolar disorder. *Epidemiology and Psychiatric Sciences, 25*(4), 312–316.
Matzavela, V., & Alepis, E. (2021). Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments. *Computers and Education: Artificial Intelligence, 2*, 100035.
McGuire, J. T., & Botvinick, M. M. (2010). Prefrontal cortex, cognitive control, and the registration of decision costs. *Proceedings of the National Academy of Sciences, 107*(17), 7922–7926.
McMahon, D., Wright, R., Cihak, D. F., Moore, T. C., & Lamb, R. (2016). Podcasts on mobile devices as a read-aloud testing accommodation in middle school science assessment. *Journal of Science Education and Technology, 25*(2), 263–273.
Moreno-Marcos, P. M., Pong, T. C., Munoz-Merino, P. J., & Kloos, C. D. (2020). Analysis of the factors influencing learners' performance prediction with learning analytics. *IEEE Access, 8*, 5264–5282.
Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: A machine-learning approach. *Higher Education, 80*(5), 875–894.
Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences, 11*(1), 237.
Nguyen, H. D., Yoo, S. H., Bhutta, M. R., & Hong, K. S. (2018). Adaptive filtering of physiological noises in fNIRS data. *BioMedical Engineering Online, 17*(1), 1–23.
Papo, D. (2013). Why should cognitive neuroscientists study the brain's resting state? *Frontiers in Human Neuroscience, 7*, 45.
Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: An overview of learner modeling techniques. *User Modeling and User-Adapted Interaction, 27*(3), 313–350.
Petersen, S. E., & Dubis, J. W. (2012). The mixed block/event-related design. *NeuroImage, 62*(2), 1177–1184.
Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., et al. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences, 1464*(1), 5.
Polyzou, A., & Karypis, G. (2019). Feature extraction for next-term prediction of poor student performance. *IEEE Transactions on Learning Technologies, 12*(2), 237–248.
Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence, 2*, 100018.
Rose, D. H., & Strangman, N. (2007). Universal design for learning: Meeting the challenge of individual learning differences through a neurocognitive perspective. *Universal Access in the Information Society, 5*(4), 381–391.
Sablić, M., Mirosavljević, A., & Škugor, A. (2020). *Video-based learning (VBL)—past, present and future: An overview of the research published from 2008 to 2019* (pp. 1–17). Technology: Knowledge and Learning.
Seghouane, A. K., & Ferrari, D. (2019). Robust hemodynamic response function estimation from fNIRS signals. *IEEE Transactions on Signal Processing, 67*(7), 1838–1848.
Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science, 72*, 414–422.
Simamora, R. M. (2020). The Challenges of online learning during the COVID-19 pandemic: An essay analysis of performing arts education students. *Studies in Learning and Teaching, 1*(2), 86–103.
Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. American Chemical Society.
Tachtsidis, I., & Scholkmann, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: Issues, challenges, and the way forward. *Neurophotonics, 3*(3), Article 031405.
Tatar, A. E., & Düştegör, D. (2020). Prediction of academic performance at undergraduate graduation: Course grades or grade point average? *Applied Sciences, 10*(14), 4967.
Thees, M., Kapp, S., Strzys, M. P., Beil, F., Lukowicz, P., & Kuhn, J. (2020). Effects of augmented reality on learning and cognitive load in university physics laboratory courses. *Computers in Human Behavior, 108*, 106316.
Umlauft, J., & Hirche, S. (2019). Feedback linearization based on Gaussian processes with event-triggered online learning. *IEEE Transactions on Automatic Control, 65*(10), 4154–4169.
Verriotis, M., Fabrizi, L., Lee, A., Cooper, R. J., Fitzgerald, M., & Meek, J. (2016). Mapping cortical responses to somatosensory stimuli in human infants with simultaneous near-infrared spectroscopy and event-related potential recording. *ENeuro, 3*(2).
Wachtler, J., Scherz, M., & Ebner, M. (2018). June). Increasing learning efficiency and quality of students homework by attendance monitoring and polls at interactive learning videos. In *EdMedia+ innovate learning* (pp. 1357–1367). Association for the Advancement of Computing in Education (AACE).
Wang, J., Xie, H., Wang, F. L., Lee, L. K., & Au, O. T. S. (2021). Top-N personalized recommendation with graph neural networks in MOOCs. *Computers and Education: Artificial Intelligence, 2*, 100010.
Watanabe, E., Maki, A., Kawaguchi, F., Takashiro, K., Yamashita, Y., Koizumi, H., et al. (1998). Non-invasive assessment of language dominance with near-infrared spectroscopic mapping. *Neuroscience Letters, 256*(1), 49–52.
Wijayasekara, D., Manic, M., Sabharwall, P., & Utgikar, V. (2011). Optimal artificial neural network architecture selection for performance prediction of compact heat exchanger with the EBaLM-OTR technique. *Nuclear Engineering and Design, 241*(7), 2549–2557.
Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test (WRAT4)*. Lutz, FL: Psychological Assessment Resources.
Xiao, L., Li, K., Tan, Z., Zhang, Z., Liao, B., Chen, K., … Li, S. (2019). Nonlinear gradient neural network for solving system of linear equations. *Information Processing Letters, 142*, 35–40.
Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27.
Zgaljardic, D. J., Durham, W. J., Mossberg, K. A., Foreman, J., Joshipura, K., Masel, B. E., … Sheffield-Moore, M. (2014). Neuropsychological and physiological correlates of fatigue following traumatic brain injury. *Brain Injury, 28*(4), 389–397.
Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: A systematic review. *Studies in Science Education, 56*(1), 111–151.
Zohdi, H., Scholkmann, F., & Wolf, U. (2021). Individual differences in hemodynamic responses measured on the head due to a long-term stimulation involving colored light exposure and a cognitive task: A SPA-fNIRS study. *Brain Sciences, 11*(1), 54.