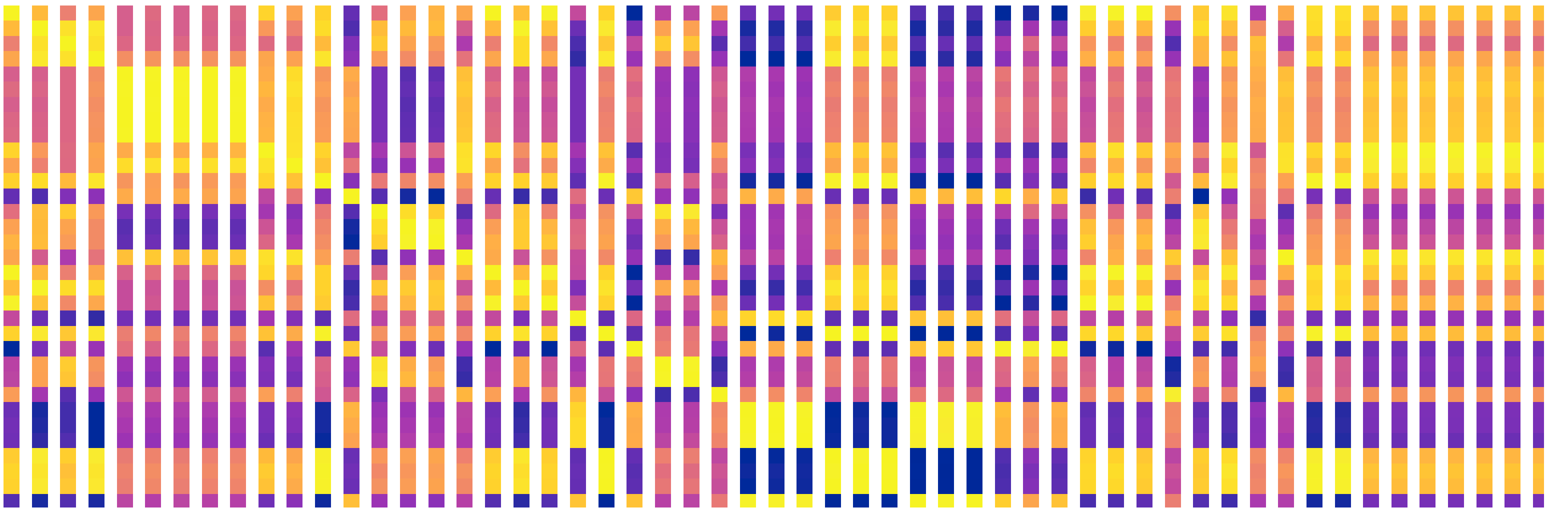
NOTTINGHAM
TRENT UNIVERSITY

MULTIMODAL MULTISENSOR ATTENTION MODELING

A thesis submitted
to Nottingham Trent University in
fulfillment of the requirements for the degree of Doctor of
Philosophy



This thesis was written by
Mohammad Hossein Taheri
M.S., Swansea University, Swansea, UK, 2012
B.S., Guilan University, Rasht, Iran 2010

October 2014 - May 2020

ABSTRACT

Introduction: Sustaining attention is one of the most important factors in determining successful outcomes and deep learning in students. Existing approaches to track student engagement involve periodic human observations that are subject to inter-rater reliability. Our solution uses real-time Multimodal Multisensor data labeled by objective performance outcomes to track the attention of students.

Method: The study involves four students with a combined diagnosis of cerebral palsy and a learning disability who took part in a 3-month trial over 59 sessions. Multimodal Multisensor data were collected while they participated in a Continuous Performance Test (CPT). Eye-gaze, electroencephalogram, body pose, and interaction data were used to create a model of student attention through objective labeling from the Continuous Performance Test outcomes. To achieve this, a type of continuous performance test is introduced, the Seek-X type. Nine features were extracted including High-Level handpicked Compound Features (HLCF). Using leave-one-out cross-validation, a series of different machine learning approaches were evaluated.

Research questions:

RQ1: Can we create a model of attention for PMLD/CP students using the CPT?

RQ2: What are the main correlations found in the CPT outcomes and the Multimodal Multisensor data?

Results: Overall, the random forest classification approach achieved the best classification results. Using random forest, 84.8% classification for attention and 65.4% accuracy for inattention were achieved. We compared these results to outcomes from different models: AdaBoost, decision tree, k-Nearest Neighbor, naïve Bayes, neural network, and support vector machine. We showed that using a multisensor approach achieved higher accuracy than using features from any reduced set of sensors. Incorporating person-specific data improved the classification outcome, compared to being participant neutral. We found that using High-Level handpicked Compound Features (HLCF) can improve the classification accuracy in every sensor mode. Our approach is robust to both sensor fallout and occlusions. The single most important sensor feature to the classification of attention and inattention was shown to be eye-gaze. We have shown that we can accurately predict the level of attention of students with learning disabilities in a real-time approach that is not subject to inter-rater reliability, human observation, or reliant on a single mode of sensor input. In total, 2475 separate correlation tests were carried over 55 data points using Pearson's correlation coefficient. Data points from the SDT, CPT outcomes measures, Multimodal Multisensor features, and participant characteristics were assessed longitudinally for cross-correlation significance. A strong positive correlation was found between participant ability to maintain sustained and selective attention in the CPT to their academic progress in school (d'), $P < .01$. Participants who showed more inhibition in tests had progressed further in their academic assessments $P < .01$. The Seek-X type CPT also showed specific physiological characteristics, including body movement range and eye-gaze that were significant in P scales such as 'Reading' and 'Listening' $P < .05$. We found that participant bias was overall liberal $\overline{B''_D} < 0$. Participants showed no significant bias change during the sessions, and we found no significant correlation between bias (B''_D) and sensitivity (d').

Conclusion: An approach to labeling Multimodal Multisensor data to train machine-learning algorithms to track the attention of students with profound and multiple disabilities has been presented. We posit that this approach can overcome the variation in observer inter-rater reliability when using standardized scales in tracking the emotional expression of students with such profound disabilities. The accuracy of our approach increases with multiple modes of sensor input, and our method is robust to sensor occlusion and fall-out. Multiple sources of sensor input are provided, to accommodate a wide variety of users and their needs. Our model can reliably track the attention of students with profound disabilities, regardless of the sensors available. A system incorporating this model can help teachers design personalized interventions for a very heterogeneous group of students, where teachers cannot possibly attend to each of their individual needs. This approach could be used to identify those with the greatest learning challenges, to guarantee that all students are supported to reach their full potential.

Keywords—Affective computing in education, affect detection, attention, continuous performance test, engagement, flow, HCI, interaction, learning disabilities, machine learning, multimodal, multisensor, physiological sensors, Signal Detection Theory, selective attention, sustained attention, student engagement.

TABLE OF CONTENTS

LIST OF ACRONYMS

Acronym	Definition
AAC	Augmentative and Alternative Communication
ANN	Artificial Neural Network
BCI	Brain-Computer Interface
CC	Correct Commission
CO	Correct Omission
CP	Cerebral Palsy
CPT	Continuous Performance Test
EEG	Electrocardiography
EMG	Electromyography
ERP	Event-Related Potentials
FAR	False Alarm Rate
H	Hit
HCI	Human-Computer Interaction
HLA	High-Level Attention
HLCF	High-Level Compound Feature
HLI	High-Level Inattention
k-NN	k-Nearest-Neighbor
LA	Learning Activity
L-O-O	Leave One Out
MM	Multimodal Multisensor
NN	Neural Network
PMLD	Profound and Multiple Learning Difficulty
ReLU	Rectified Linear Unit
RQ	Research Question
SDT	Signal Detection Theory
SSPI	Severe Speech and Physical Impairment
TI	Total Imitations
TN	True Negatives
TO	Total Omissions
TP	Total Presses
TP	True Positives
TT	Total Targets
VEP	Visually Evoked Potentials
WC	Wrong Commission
WO	Wrong Omission
ZPD	Zone of Proximal Development
ZPF	Zone of Proximal Flow

LIST OF FIGURES

LIST OF TABLES

DEDICATION

Dedicated to my beloved parents, Nahid Ansari and Bahram Taheri for their continuous support throughout my life and their everlasting personal enthusiasm and encouragement in my academic career. And a special thanks to my lovely fiancée Meg Wills, for her support and help in proofreading this manuscript.

ACKNOWLEDGEMENTS

I want to give my thanks to my Professors; Nasser Sherkat, David Brown and Caroline Langensiepen, for their guidance and help in the complete process of this Ph.D. and my academic career.

Chapter 1. INTRODUCTION

It is often a challenge to keep children attentive in learning activities, especially if the activity requires them to retain focus and active participation for a continuous period of time. Researchers reported that students with learning disabilities do not display any significant attention deficiency compared to non-disabled students – these students can complete the same activities if given more processing time [1]. Despite this outcome, student engagement can vary greatly depending on the activity, and understanding when the student is engaged, and when they are not, is not a straightforward task.

While research has focused significantly on the ability of children with learning difficulties to recognize [2], perceive [3] and interpret [4] emotional cues, there is little to no research on the recognition of the emotional state of these students. The importance of carers being able to interpret the emotional cues and states of such students has been documented in [5]. It is found that carers made significantly more critical and 'fundamental attribution' [6] errors in the emotional expression of their clients with learning disabilities in comparison to their clients without learning disabilities. This affects the quality and quantity of their client's treatment [5] and has a negative effect on the provisional treatment [7], [8]. Currently, carers rely on their expert understanding and personal experience of the students to interpret their voices, expressions, and gestures. Dependent on the personal experience with a particular client, a carer's internal modeling of the emotional expression of that client can vary widely and demonstrate inter-rater reliability issues.

One of the main ways to measure engagement in students with special educational needs is to use the Special Schools and Academies Trust (SSAT) Engagement Scale [9]. The engagement profile scale is a classroom tool developed through SSAT's research into effective teaching and learning for children with complex learning difficulties and disabilities. It allows educators to focus on the child's engagement as a learner and create personalized learning pathways [10]. The authors describe seven components of engagement namely, awareness, curiosity, investigation, discovery, anticipation, persistence, and initiation. Teachers assign a score out of four for each component giving a total score out of 28. One potential issue with the use of this scale is that teachers assign a subjective rating to each component, which will be subject to inter-rater variability.

The scale has been used to assess the impact of new technologies in special education – especially in studies investigating the suitability of humanoid robots to support learning in students with Profound and Multiple Learning Disabilities (PMLD). The approach of using an engagement scale to create personalized learning pathways has been examined by others [11]–[13].

One way to overcome the variation in observer inter-rater reliability in tracking emotional expression is to introduce a reliable indicator of that emotion. In this research, a robust methodology for tracking attention levels of children with PMLD or Cerebral Palsy (CP) is proposed using Signal Detection Theory (SDT) [14]. The application of this theory gives quantifiable information on the improvement of deterioration or attention in response to a Continuous Performance Test (CPT) specifically adapted to the abilities of such students [1], [15]. Performance in the CPT has been shown to provide objective labels to train machine learning algorithms [16]. Sensor data (e.g., on eye-gaze and body pose) is collected whilst the students are participating in a CPT. After obtaining a labeled dataset, machine learning models can be applied to the data so that in the future new unlabeled data can be presented to the model and attention can be inferred.

Many traditional interactive systems use devices such as a keyboard and mouse and are constructed to emphasize the transmission of explicit messages while ignoring implicit information about user interaction. The emerging science of affective computing can only be accelerated with the abundance of sensor data [17], [18] and wearables [19]. These multimodal human cues [20]–[22] provide the Multimodal Multisensor data points necessary for enhanced emotional modeling. Multimodal Multisensor data has been instrumental in determining user affective states [20], [23]–[29] including engagement [30]–[32].

There are a number of challenges to developing such a model including understanding the relationship between the terms used in educational contexts (e.g., 'flow' and 'engagement'), developing appropriate CPTs suitable for the abilities of students with the most profound learning disabilities, selection of appropriate sensors and features derived from these data streams from which emotional states can be inferred, finding a suitable population of end-users to collect data with to train the machine learning algorithms, and finally comparing the performance of a range of machine learning methods to track attention. This work addresses each of these challenges.

1.1 Research questions

RQ1: Can we create a model of attention for PMLD/CP students using the CPT?

RQ2: What are the main correlations found in the CPT outcomes and the Multimodal Multisensor data?

1.2 Outline of thesis

The thesis starts by introducing affect states of learning and the challenges that carers and teachers face in monitoring learner affect state, especially for PMLD and CP learners. It does on to describes objective alternatives to tracking affect state, namely the CPT. The literature review discusses the affect states of learning in the Zone of Proximal Flow theory and relates them to learner skill and learning challenge level. A theoretical framework for monitoring learner performance in a learning activity is discussed and examples are provided of how to track learner performance in a learning activity with known challenge levels. To answer the core research questions a CPT is developed that tracks Multimodal Multisensor data while the learner takes part in a series of signal detection trials of varying difficulty. The methodology, participant characteristics and data collection findings are explained in detail. Empirical lessons during the data collection pilots are presented and adaptations to the method are developed for the PMLD/CP user group. Machine learning results and comparisons are discussed. The best Machine Learning methods for each sub-set of sensors is presented with accuracy ratings. Cross-correlation analysis is also completed that explores relationships between CPT outcomes measures and participant characteristics. Significant correlations between the two are identified. The appendix includes documents used to gain ethical approval, information packs, parental consent forms, complete correlation results and papers published from this work.

Chapter 2. LITERATURE REVIEW

The field of affective computing evolved from the field of Human-Computer Interaction (HCI) in an effort to reduce user frustration [33]. Affective computing expands HCI by including emotional communication along with appropriate means of managing affective information [33]. Current HCI designs, usually use traditional interface devices, such as a keyboard and mouse, and are designed to emphasize the transmission of explicit messages, while ignoring implicit user information, such as changes in an affect state. However, changes to the user's affect state is an essential component of human-human communication. Some affect states encourage human actions, and others enrich the meaning of human interactions. Subsequently, HCI, which ignores the user's affect states, compromises contextual information available in the interaction process that can augment the experience. As a result, such interactions are frequently perceived as cold, out of context, and socially inept. The advancement of the human-computer paradigm is dependent on future user interfaces that are not only sensitive to affect but are anticipatory and based on naturally occurring multimodal human cues [20]–[22]. The emerging science of affective computing can only be accelerated with the abundance of sensor data [17], [18] and wearables [19], which augment the user interface with Multimodal Multisensor data points necessary for enhanced emotional modeling. Multimodal Multisensor data has been instrumental in determining user affect state [20], [23]–[29] including engagement [30]–[32], [34].

> *"The essential role of emotion in both human cognition and perception, as demonstrated by recent neurological studies, indicates that affective computers should not only provide better performance in assisting humans but also might enhance computers' abilities to make decisions. Affective computing, coupled with new wearable computers, will also provide the ability to gather new data necessary for advances in emotion and cognition theory"* [19].

A significant amount of research has been carried out in the classification of human behaviors [23], [26], [35]–[38]. Many of these studies model human behavior patterns to learn about the user's state or predict future behaviors. This has been the motivation for the field of "affective computing", which is the affective response to HCI with consideration of the temporal user state [33].

Affect is any feeling or emotion, which is the direct experience and consciousness of a particular emotional state (as in a person's feelings of elation upon accomplishment) [39]. Some affective states encourage human actions, and others enrich the meaning of human interactions. A distinction has also been made between negative and positive affect states, for example, 'flow' is a positive affect state while 'boredom' and 'frustration' are considered negative affect states [40], [41].

Along with cognition and conation, affect is one of the three traditionally identified components of the mind. These three divisions are classically referred to as the "ABC of psychology" which are "affect," "behavior," and "cognition" [42]. Cognition is relating to the part of mental functions that deals with logic, as opposed to affective which deals with emotions. Conation refers to the connection of knowledge and affect (emotion) to behavior and is associated with the issue of "why." The conative, as opposed to the cognitive or affective, relates to purposeful, but not necessarily ultimately rational, action.

Like cognitive objectives described by Bloom's taxonomy [43], affective (or feeling) domain objectives can also be divided into a hierarchy (according to Krathwohl's revised taxonomy [44]). The area of the affective domain is concerned with feelings or emotions and social/emotional learning and skills. Krathwohl describes the affect domain as the following [44]:

1. Receiving
   This refers to the learner's sensitivity to the existence of stimuli – awareness, willingness to receive, or selected attention.
   - feel, sense, capture and experience
   - pursue, attend, perceive
2. Responding
   This refers to the learner's active attention to stimuli and his/her motivation to learn – acquiescence, willing responses, or feelings of satisfaction.
   - conform, allow and cooperate
   - contribute, enjoy and satisfy
3. Valuing
   This refers to the learner's beliefs and attitudes of worth – acceptance, preference, or commitment. An acceptance, preference, or commitment to a value.
   - Believe, seek and justify
   - respect, search and persuade
4. Organization
   This refers to the learner's internalization of values and beliefs involving (a) the conceptualization of values; and (b) the organization of a value system. As values or beliefs become internalized, the learner organizes them according to priority.
   - examine, clarify and systematize
   - create and integrate
5. Characterization
   This refers to the learner's highest level of internalization and relates to behavior that reflects (a) a generalized set of values; and (b) a characterization or a philosophy about life. At this level, the learner is capable of practicing and acting on their values or beliefs.
   - internalize, review and conclude
   - resolve and judge

2.1 Engagement

Engagement is a concept of the significant importance of HCI, not only to inform the design and implementation of interfaces, but also to allow more sophisticated interfaces, capable of adapting to users. Although the notion of engagement is being actively studied in a diverse set of domains, the term has been used to refer to concepts such as interest, sustained attention, immersion, and involvement [45]. Student engagement (participation in learning) was found to be the most reliable feature for determining successful learning [34], [46], [47], and to result in major educational outcomes such as persistence, satisfaction, and academic achievement [48]. Active personalized learning was shown to encourage participation and engagement, not only in the classroom, but also in extra-curricular activities and work-related learning in the local community [49]. As the tutor or the technological learning facilitator forms a better understanding of the learner's strengths and challenges, they are in a better position to go through scaffolding objectives, involving the choice of skill to train at a given moment and choice of learning activities, while preserving the learner's interest and engagement [50].

According to Carpenter [51], the process of engagement is a journey that connects students and their environment (including people, ideas, materials, and concepts) and enables learning and achievement. Students who are disengaged can become frustrated or bored, which can have an adverse effect on achievement and lead to disruption of learning, for the individual learner, as well as for other students when learning takes place in a collective/collaborative environment like a classroom.

The term "engagement" is used in at least two ways. First of all, it can be used in the sense of launch, bearing in mind the initiation of contact. For example, a user can interact with a machine, moving into a specific range to which the machine responds. This does not imply any necessary duration of interaction, but rather is a step - the user can work with the machine for a short time, and then decide to leave. In the longer term, engagement also refers to the concept of activity [52]. In this regard, engagement seems to imply a more sustained engagement. In the literature, engagement is described in different ways: as a process; as a phase in a process or as an overall process; as experience; as a cognitive state of mind; an empathic connection; as a perceived or theorized indicator that describes the general state of the interaction. However, most engagement studies reveal two basic foundations: attention and emotional involvement [52]. Selective attention of a stimulus seems to be necessary for the most basic form of engagement to take place. At a glance, this form of engagement can be limited to a relevant potential incentive that turns out to be no more interest. A more sustainable form of attention places greater demands on commitment and also allows for the possibility of emotional participation [52].

O'Brien *et al.* [53] define engagement as the following. "Engagement is a quality of user experiences with technology that is characterized by challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect." Specifically, when we discuss engagement in relation to technology, we use it to show a degree of involvement, alongside immersion, presence or fun [54].

Engagement, when termed as a process, can be seen as a series of different phases through which it can progress. They can relate to the intensity or level of participation of the participants in relation to the focus of the interaction. In the study of interaction with humanoid robots, Sidner *et al.* [55] refer to engagement as a process by which individuals in an interaction start, maintain and end their perceived connection to one another. It is a natural starting point to consider engagement as consisting of at least these three broad phases. O'Brien and Toms [53] refer to four phases of engagement: a point of engagement, sustained engagement, disengagement, and re-engagement. This is illustrated in **Table 1**.

Table 1. Attributes of the 4 stages of engagement from [53].

Point of engagement attributes	Period of engagement attributes	Disengagement attributes
Interest Motivation Novelty Aesthetics Specific or experiential goal	Attention Interactivity Control Novelty Challenge Feedback Interest Aesthetic and sensory appeal Positive Affect	Challenge Interruptions Negative affect Positive affect Usability Perceived time

Re-engagement

The concept of re-engagement raises the critical issue of when an engagement can be considered to end. In some circumstances, the student may leave his or her chair or start looking around the classroom. However, in other cases, it may be harder to determine. For example, if the student looks away briefly, it may just mean that they have been temporarily distracted. In some cases, looking away may actually signal engagement, such as during shared attention scenarios, when looking at an object under mutual interest in a collaborative activity [56]. The social environment has also been used in [57], among other indicators, to help identify the engagement phase according to the stages of "present," "attending and interested," "engaged and interacting." This is an important consideration during mobile scenarios, for example, where robots and users are free to move around the environment, highlighting the important role of the context of the interaction [53].

Therefore, if we consider these interactions as emotional cues, individually they carry undetermined outcomes in a view to engagement state. In the examples, the context of the interaction (looking away, at what?) and the user condition (mainstream or autism or…?) define the emotional cue construct. With a naive view of context, emotional cues are not a global predetermination of engagement. In a set context, emotional cues are personal to the user condition. To summarize, the model of human engagement is influenced by the three C's of engagement.

$$\text{Engagement influences} \begin{cases} emotional\ Cues \\ interface\ Context \\ human\ Condition. \end{cases}$$

Without being consciously aware of it, we evaluate the emotional cues we receive from people we interact with on a day-to-day basis. This is an early years skill [58], modeled in the teen years from parents [59], with girls showing greater ability [60]. Unconsciously, we create a model of their engagement as a response to our interactions. We take great sensitivity to the context of the interaction, was the scene a casual or a serious one. We condition our response to the receiver’s condition, if they are joyous, our response reflects that, if they are mourning, we are cautious and selective. Similarly, in more complex interactions, if our receiver has autism, a learning disability, or PMLD we formulate our response accordingly to suit the condition of the interaction. Ultimately, we store this model, our perception and attributions of it to memory - only to retrieve it in our next interaction with that person, and to use it like a map, to best guide our engagement experience. Significantly, our perception shapes this model, and it must be evaluated through this lens. The subjectivity of our perception and attributions to the engagement, cannot be ignored [6]–[8], they condition our response. The subjectivity of our perception introduces a challenge in observer-rated engagement values in studies. Also, participant self-rated values can suffer from intra-participant variability. Another issue being, observer-rated methods cannot be automated and require a dedicated human observer, ideally a single one for all participants.

According to Peters *et al.* [45] engagement is experienced in two common modes. First of all, it can be seen in the sense of starting when referring to the initiation of an interaction. Selective attention to a stimulus is necessary for this most basic form of engagement. This may be demonstrated by a quick glance at a potentially relevant stimulus that proves to be of no further interest. In a longer-term sense, engagement has also been seen as the concept of being actively occupied with an interaction. In this respect, the engagement seems to imply a more sustained involvement. In this form of interaction, a more sustained form of attention is required and also allows the possibility of affective involvement.

The later mode of engagement, which involves a more sustained experience has the potential to lead to immersion and interest in this study. We are interested in tracking the optimal engagement experience. Csíkszentmihályi describes the optimal experience as flow when the person is in a state of mind where their awareness and activities merge.

2.2 Observational methods of engagement tracking

One of the main ways to measure attention and engagement for students with special educational needs is the use of the Special Schools and Academies Trust (SSAT) Engagement Scale [9]. The engagement profile and scale is a class tool developed as part of the SSAT research on the effective teaching and learning of children with learning difficulties and complex disabilities. It allows teachers to focus on the child's engagement as a learner and create a personalized learning path [10]. The authors describe seven components of engagement, namely awareness, curiosity, investigation, discovery, anticipation, persistence and initiation (seen in **Figure 1**). The variable success in the determination of student affect state introduces noise in observatory data collection methods, also the imprecise affective state labeling in this method casts doubt in the validity of affect data labeled by the teachers.

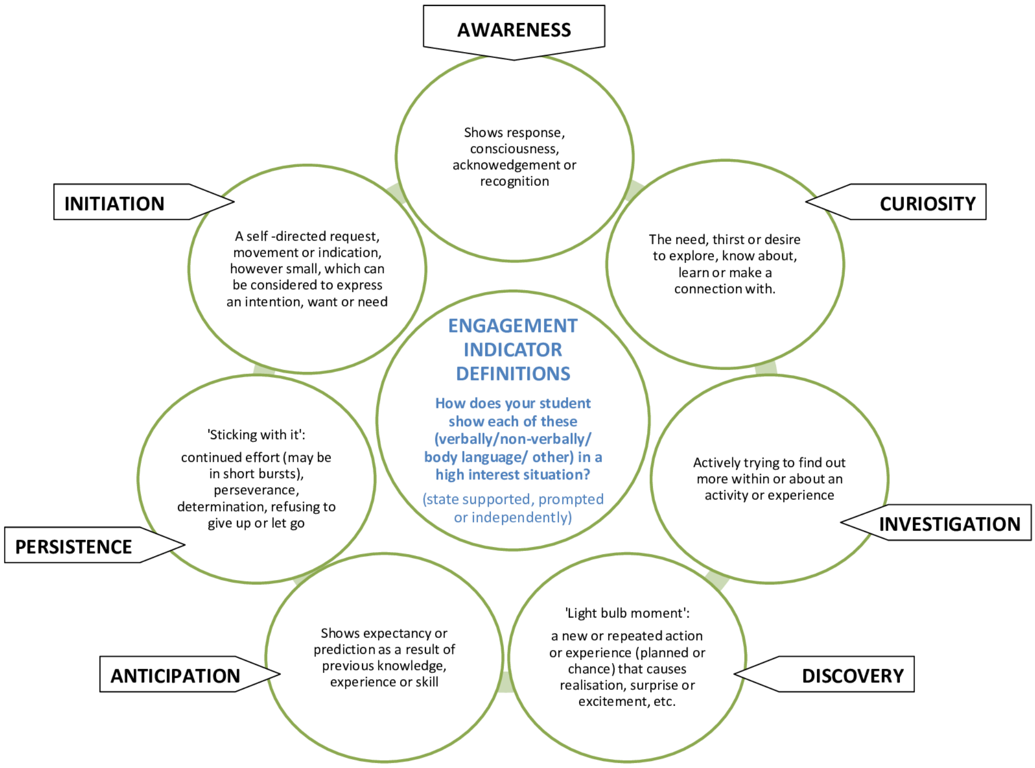


Figure 1. Special Schools and Academies Trust (SSAT) Engagement Scale.

The scale has been used to assess the impact of new technologies in special education – especially in the suitability of humanoid robots to support PMLD learning [11], [12]. Assisted learning [61]–[63] is of particular interest to users with varying degrees of learning difficulty. An example of using affective computing in a special needs educational setting includes the application of an intelligent agent in assisted learning by monitoring the user's response to different learning routes by finding the optimal learning pathway. This has an impact because customizing the learning pathway could allow the system to compensate for user disability and adapt the learning experience to suit the user's more receptive learning pathways (focusing on their abilities, not limitations).

Research in the underlying language of interactions in which people may engage with computers plays a significant role in the design and implementation of smart interfaces for a variety of applications, from learning to assistive [52]. Such interfaces should be capable of adapting to the individual user needs and acting appropriately according to the context of the situation and the requirements. Being able to monitor, track and react contextually to users' interest and engagement, plays a vital role in achieving this.

2.3 Flow, performance, attention and engagement in learning

Engagement's crucial role in learning was recognized by Carpenter [62], stating that "Sustainable learning can occur only when there is meaningful engagement". Learner engagement in the classroom is the single most reliable indicator of deep learning [46], [47], [64] and learner satisfaction [65]–[67]. In the absence of learner engagement, deep conceptual learning is also not present [34], [51], which is an essential attribute to long-term learning and new skill achievement [51]. Chen *et al.* presented a model of engagement in games (2005) that related skill, challenge level and attention to engagement [68]. They later claimed it was not ready (2007, 2011) [69], [70]. In education, the use of the term 'engagement' is more familiar to teachers than flow. D'Mello and Graesser [71] see considerable overlap between the two terms: "We conceptualize engagement/flow as a state of engagement with a task such that concentration is intense, attention is focused, and involvement is complete" (p.146). The relationship between engagement, flow and attention has been illustrated in Bianchi-Berthouze's [32] engagement model, modified from the original model developed by Chen *et al.*, see **Figure 2**.

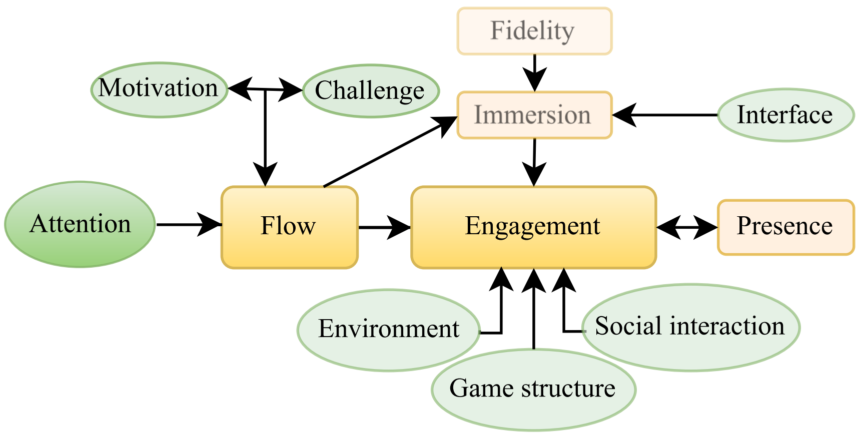


Figure 2. Relationship between attention, flow and engagement shown in diagram form [32].

Contrary to engagement, the concept of flow is well defined in Csíkszentmihályi's works [67], [72]. While in flow, according to Csíkszentmihályi, the person's skill level is matched with their challenge level [64], [66], [67] *making flow the optimal psychological state of engagement, resulting in immersion, concentrated focus and deep learning* [65], [66]. This results in concentrated focus and immersion in the task which leads to deep learning and High-Levels of work satisfaction [67].

Flow's intense experiential involvement is responsible for three additional reported qualities: the fusion of action and awareness, a sense of control, and an altered sense of time [73]. During flow, attention resources are fully absorbed in the task, so that objects outside the immediate interaction do not come into consciousness [73]. Attention is so totally absorbed in momentary activity that there is little to devote to mental processes that contribute to the duration of time [74]. As a result, people who are deeply immersed, usually report that time passes quickly [75]. William James noted that boredom seems to increase when "we grow attentive to the passage of time itself" [76].

A characteristic feature of flow is the intense empirical commitment to momentary activity. Attention is completely absorbed in the task at hand, and the person performs at their greatest capacity [73]. The ability of a person to sustain attention is often combined with self-control and inhibition, which enormously increases performance [73], [77].

Flow experiences are relatively rare in everyday life, but almost any activity is capable of producing them, provided that certain conditions are met. Previous research suggests three conditions of fundamental importance. Firstly, flow tends to occur when the activity contains a clear set of goals [73]. Secondly, a balance between perceived challenges and perceived skills, otherwise known as the concept of "optimal arousal" [40], [64], [78]–[80]. However, this is the perception of challenge, not the actual objective presence [73], [81]. When perceived challenges and skills are balanced, attention is completely invested. This balance, however, is inherently fragile; if challenge gradually exceeds skill, one typically becomes anxious or frustrated; if skill begins to exceed challenge, one relaxes and then becomes bored [40], [64], [73]. This is a push and pull dynamic. The equilibrium of skill and challenge is also represented in the Zone of Proximal Flow (ZPF) theory [40], [64]. We adapt and modify the engagement diagram from Chen *et al.* [68] and Bianchi-Berthouze's [32] in **Figure 3**, with a focus on learning challenge and learner skill, using the definitions of flow, attention and flow from the works of Csíkszentmihályi [73], [82], [83] and the ZPF theory.

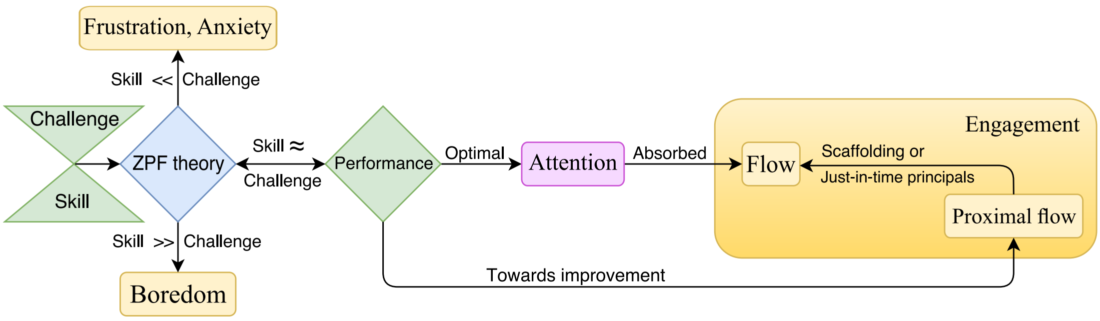


Figure 3. Visualization of the relationship between performance, attention and flow using the Zone of Proximal Flow theory [40], and definitions from Csíkszentmihályi theory of flow [73]. Affect states are shown in orange and attributes from the learning activity are shown in green.

High levels of performance usually depend on targeted attention focused on specific challenges and clear feedback [84]. Unsurprisingly, studies have found a strong positive relationship between flow and performance, and between attention and learning performance [85]. For example, flow is positively associated with artistic and scientific creativity [86], [87], effective teaching [66], learning [88] and peak performance in sports [89], [90]. This leads to the advancement of skill.

As a person masters challenges in an activity, they develop their skill, and the activity ceases to be as challenging as before. To continue experiencing flow, they must identify increasingly greater challenges. Thus, over time, the balance between challenge and skill enhances competence. Experiential goals thus encourage growth and stretch a person's existing capacities (Vygotsky's Zone of Proximal Development) [91].

This positive relationship between flow and skill development has been demonstrated in several studies [90] in which students were tested in a school environment. In a longitudinal study, teenage students still committed to pursuing their talent area were compared to their peers who had already disengaged. Four years earlier, those who were still committed had experienced more flow and less anxiety than their fellow students while engaged in school-related activities; they were also more likely to identify their area of talent as a source of flow [88]. In another longitudinal study, students talented in mathematics showed that those who experienced flow in the first part of the course performed better in the second half, achieving a higher grade point average (GPA) [73], [92]. Other longitudinal research suggests that the matching of challenge and skill in daily life may protect against negative outcomes. Teenagers who had experienced high adversity at home or school but felt successful when engaging in challenging extracurricular activities were much less likely to have problems years later [93].

To summarize, flow is the optimal state of engagement, where engagement meets productivity [65], [66]. Maintaining flow in learning is especially significant because it is the most reliable indicator for determining successful learning [36], [45]-[48]. This results in immersion, concentrated focus and deep learning [65], [66]. One is in flow when one is engaged [32] and steady performance has been maintained at the comfortable limits of one's skill limitations [64] for the duration of time. Therefore, performance trend tracking can be used as an indicator of flow [64]. This approach has been used in [34], [40], [64], [95], [97]–[100] as a model for relating learner affect to user performance in a pre-defined activity/task with a known challenge. Conclusively flow, *a sub-state of engagement* [32], [67], [101], is a suitable measure to track and assess the quality of an experience; firstly it can be objectively monitored through performance tracking, secondly, through its monitoring, engagement is also established. Flow is the optimal state of engagement, where engagement meets productivity. Maintaining flow in learning is especially significant because it is the most reliable indicator for determining successful learnings [46], [47], [102]. In the absence of learner engagement, deep conceptual learning is also not present [51], which is an essential attribute to long-term learning and new skill achievement [51].

Flow is central to classroom performance and the achievement of learning outcomes [45]-[48] which is closely linked with attention [16], [32], [64]. During flow, attention is completely absorbed in the task at hand, and the person's performance is maximized [73], [85]. The ability of a person to sustain attention often coincides with inhibition, which increases performance [73], [77]. In the next section, we discuss the affect states of learning and how learner skill, interacting with learning challenges influences learner affect.

2.4 Affective states of learning

Developing an affect sensitive system requires a firm understanding of the relationship between skill, challenge and how it influences the learner's emotional state. This helps clarify the relationship between learning and affect state and the impact a system could have on learning and engagement. There is now an accumulation of evidence to indicate the link between affect and cognitive performance and decision-making [103]. The goal to learn and understand is associated with an increase in positive emotions like the enjoyment of learning as well as a decrease in negative emotions like boredom. Affect can direct attention and influence the level of that attention. According to Thompson and McGill [104], affect functions as a motivator, influencing the tendency to approach or avoid a situation as well as how information is processed.

The existence of the link between affect state and achievement suggests that a learning session may be improved if the teacher is sensitive and responsive to the emotional state of the learner [105]. However, the success of this strategy depends on the skill and experience of the human tutor, and there is evidence to suggest that, especially with students with special needs, teachers may find it particularly challenging to determine affect state [106].

Herein, we focus on the affect states identified by D'Mello and Picard to be significant in learning [107]. They identified frustration, boredom, and flow to be the most important emotions to skill acquisition. The concept of the Zone of Proximal Flow (ZPF) proposed by Basawapatna *et al.* [40] reflects these affect states in a two-dimensional diagram of learner skill and learning challenge. They combine independent learning limits and scaffolding from Vygotsky's Zone of Proximal Development (ZPD) with Csíkszentmihályi *et al.*'s [108] theory of flow. Carpenter [109] and Iovannone *et al.* [110] see engagement as the single best predictor of successful learning for children with intellectual disabilities. In other studies the use of multimodalities in tracking learner engagement has been explored [21], [34], [53], [95], [97]–[101], [111]–[121]. These concepts underlie the learning vision in providing an engaging learning environment in which students with diverse needs and varying levels of ability are supported by assisted learning. However, these studies do not investigate the use of Multimodal Multisensor affect tracking on PMLD and CP students.

The relationship between affect state and learning achievement is crucial for the development of the affect-based learning platform. Classroom-related affect states are linked to the students' goal structure and their adoption of specific achievement goal orientations. The goal to learn and understand is associated with an increase in positive emotions like the enjoyment of learning as well as a decrease in negative emotions like boredom. Adopting a performance-approach goal—that is, the goal to be better than others—was found to be associated with positive emotions. In contrast, the adoption of a performance-avoidance goal—that is, a goal not to appear incompetent, stupid, or uninformed in comparison to others—was related to a negative emotion like anxiety and hopelessness [63], [122]. However, the relationship between goals and affect might not be unidirectional but a reciprocal one as proposed in Linnenbrink and Pintrich's bidirectional model [123].

In 2002, Linnenbrink and Pintrich described a model of affect in which goal achievement is reciprocally related to the learner's emotional state. In this model (see Figure 4**)** the learner's personal goals are profoundly influenced by their perception of the Learning Activity (LA) challenge. This perception, in turn, has a direct influence on their affective state. Based on the broader literature, positive moods predict goal endorsement while negative moods predict avoidance goal endorsement. Personal attribution of success or failure in an activity can also affect performance interpretation, which in turn affects task involvement [124]. Wong [125] found that autonomy orientation [125] was positively related to absorption [124] in school-related activities.

This relationship between skill and affect states has been more specifically described in Csíkszentmihályi's Theory of Flow [72], where learner skill and their perception of the task challenge lead the learner to a variety of affect states, which he presented in Figure 5. Importantly, not all emotions are relevant to learning and parts of the theory of flow are less relevant to the scaffolding process in identifying optimal learning experience and the moment where the learner requires scaffolding intervention. Sidney D'Mello and Rosalind Picard [107] conducted a study on the relevance of emotions to learning and found 'frustration', 'boredom', and 'flow' to be the most important emotions to skill acquisition. This has reduced the focus of the theory of flow to the most relevant and influential states of affect for learning.

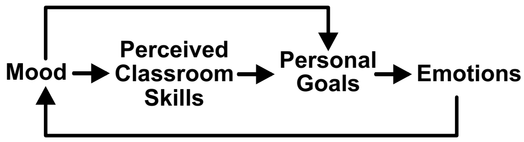


Figure 4. Linnenbrink and Pintrich's asymmetrical bidirectional model of achievement goals and affect.

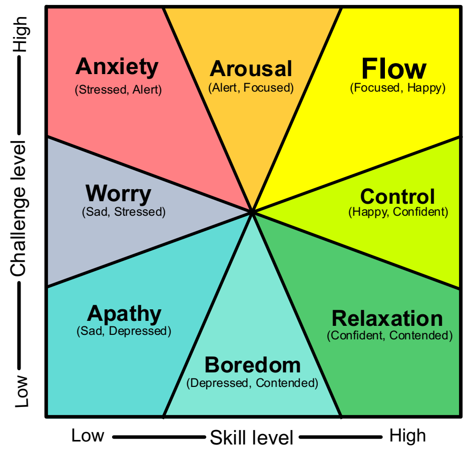


Figure 5. Csíkszentmihályi's theory of flow states[1].

In 1978, Vygotsky investigated the advancement of cognitive understanding by becoming interested in the process [126]. The boundaries of learner skill were broken into segments, (1) where learners have the capacity to learn independently and (2) assisted learning (or instructional scaffolding) from a tutor or a more knowledgeable peer. This second segment is known as the 'Zone of Proximal Development'.

Only later, in 2013, was it that Basawapatna *et al.* [40] combined learner skill, independent learning limit and scaffolding in the 'Zone of Proximal Flow' (ZPF) state change diagram. Critically this work provided the first state change diagram to reference both Vygotsky's ZPD and the affective states from Csíkszentmihályi's theory of flow. Moreover, to adapt this diagram to facilitate an educational platform, knowing the limitations of the individual learning is important in designating when the platform should mediate and deliver scaffolding intervention. To this aim, the ZPD limit to independent learning from Vygotsky's theory has been applied to the ZPF diagram, and we introduce a new diagram seen in **Figure 6** as the complete affective state diagram for learning. The learner's skill level is displayed as the X-Axis, and the task challenge is displayed as the Y-Axis. Unlike Csíkszentmihályi's flow

[1] Frank Vandeven, accessed http://frankvandeven.com/getting-into-the-flow-what-does-that-mean/ on 7/3/2018

diagram or Vygotsky's ZPD, a single ZPF graph can be used to track the learner's progress in a learning activity and any permutations of the level of skill or task challenge.

A learning experience with a learning platform comprises interactions with the learning material, the 'challenge' of activity, as depicted in the diagram of **Figure 7** and will consist of learning material with different levels of challenge. The maximum level of challenge observed by an external expert judge is used as a baseline for the highest level of challenge in the graph.

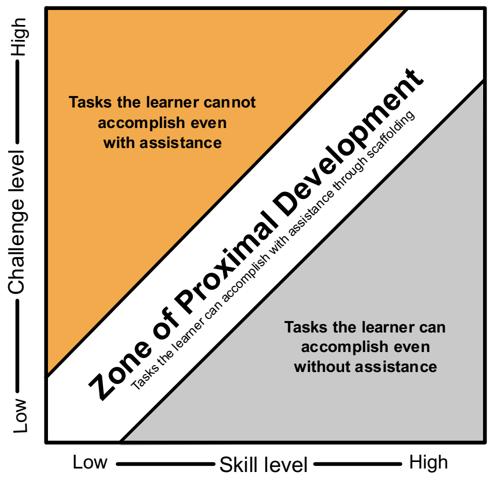


Figure 6. Morsink's graphical representation of The Zone of Proximal Development[2].

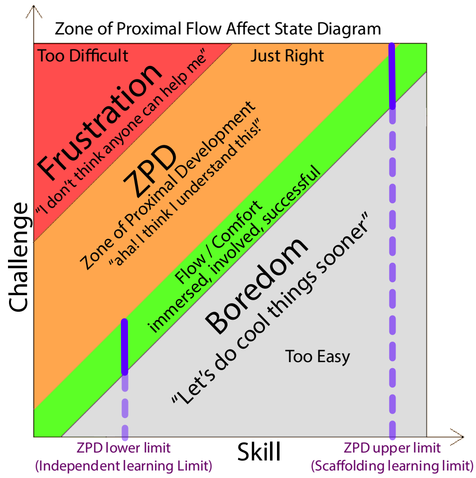


Figure 7. State diagram from the Zone of Proximal Flow theory adapted from Basawapatna *et al.* [40] to include both the ZPD lower independent and upper scaffolding learning limits.

In this way, the graph can plot more than one learner. Two ballet students learning the same ballet move could be plotted on the same ZPF graph - but importantly setting a global, not relative ground truth allows the system to influence the user's movements in the graph with only one independent variable, 'challenge'. The ground truth is set against objective measures that can be tested (by the expert or the indicators the expert sets the system to monitor) and this monitoring is achieved through performance analytics (correct and incorrect responses and response time measures) and affect state tracking. To calculate learner accuracy and success, completion time (learning achievement completion time is the time taken to answer a question) is tracked alongside the learning material challenge level to determine learner performance in relationship with activity challenge, affect state and learner skill level.

[2] Paul Morsink, *TILE-SIG Feature: The "Digitally Enhanced" Zone of Proximal Development*, http://literacyworldwide.org/blog/literacy-daily/2013/09/20/tile-sig-feature-the-digitally-enhanced-zone-of-proximal-development, accessed on 7/3/2018

2.4.1 Frustration

According to the Zone of Proximal Development theory [126], frustration is where the learner cannot achieve new learning even with assistance. Studies have found that actors who perceive that they lack the skills to take on effectively the challenges presented by the activity in which they are participating experience frustration. Simply put, if a learner feels incompetent in a given situation, he or she will tend not to be motivated [73]. This is a negative experience, and its gravity pulls the learner further into frustration, in a deteriorating cycle that hampers the learning process. In this state, the learner is exposed to a hopeless feeling; his or her emotional state could be represented by the statement "I do not think anyone can help me".

2.4.2 Vygotsky zone of proximal development

The ZPD refers to 'the state of arousal where the learner can perform an action or skill with the aid of a skilled or knowledgeable tutor or in collaboration with more capable peers'[91]. This achievement is limited by the ZPD upper limit. However, this limit is dependent on the skill of the 'more knowledgeable peer' or scaffolding tools. Better tools achieve better results as do more knowledgeable peers induce and encourage higher levels of skill achievement in others due to their access to higher levels of knowledge. This zone limit has been illustrated in **Figure 7**, with the vertical line. While in this zone, the student with assistance can acquire a higher level of skill [127]–[130]. In this zone, the level of challenge provides the optimal arousal and engaging experience for the learner to obtain new skills. In this state, the most engaging learning experiences for the learner can happen; it is where optimal and deep learning opportunities manifest themselves. According to [131] "deep learning is a committed approach to learning. It is a process of constructing and interpreting new knowledge in light of prior cognitive structures and experiences, which can be applied in new, unfamiliar contexts". Deep learning results in better quality learning and profound understanding. While in this zone the student with the assistance of the tutor (instructional scaffolding or assisted learning) acquires higher skill and is encouraged to learn and mentally develop [127]–[130].

2.4.3 Mihaly Csíkszentmihályi flow

Csíkszentmihályi first described flow in 1997 [63] as the state where the learners are fully immersed, feeling involved and successful. Flow is a delicate state where the skill level and task challenge levels are balanced. This state represents the learner state where the learner is functioning within their independent capacity, i.e., where the learners find themselves in their comfort zone, both in terms of the learning challenge or learning styles. Flow is also the state where new learning materializes as a new skill in the mind of the learner, which provides the learner an opportunity for reinforcement learning, that carries a successful emotional feeling.

Skill advancement in flow, however, is limited by the learner's lower limit of ZPD (the maximum a learner can achieve independently), which has been shown with a vertical line in **Figure 7**. Therefore, in order for the learner to achieve new learning outside their independent capacity, the learner must eventually leave flow and be led to ZPD, to pursue new learning opportunities (i.e., acquire a new skill or to complete competence of a partially acquired skill). In either case, in flow or while in ZPD, the learner is limited to the upper level of ZPD, which is dependent on the scaffolding tools and scaffolder.

2.4.4 Boredom

Boredom is the state where the learner is not challenged sufficiently. This state can manifest through the addition of a dry skill base through lecture-style teaching, or by providing interactive activities that do not challenge the learner outside what they have already learned. Boredom is a negative feeling, and its gravity pulls the learner further into this state, leading to learner disengagement and stifling the learning progress. In this state, the learner's emotional state could be represented by the statement: "let's do interesting things sooner."

In boredom, the low level of challenge relative to skill allows attention to drift. Particularly in contexts of extrinsic motivation, attention shifts to the self and its shortcomings, creating a self-consciousness that impedes engagement of the challenges. Goetz and Hall review the development of learners' boredom and call it an emotion that is frequently experienced by learners and can undermine their learning and performance [132].

2.5 Framework for monitoring learner skill and learning challenge

Measuring student skill is important because it provides an evaluation tool to compare student competency between different learning activities. The student skill can be measured in a local or global way. In local assessments, the student's performance is compared in one Learning Activity (LA) to determine the level of student achievement and affect state. In global comparison, student performance can be compared between LAs and various time intervals. These comparisons will be used in determining system adaptation.

2.5.1 Learning challenge

From the pedagogical point of view, this is a measure of the skill required to complete an activity. In that respect, the unit of Difficulty is the same as the unit of skill. If this was to be visualized in a graph (see **Figure 8**), any challenge above the diagonal of the graph would represent an increased challenge and anything underneath the diagonal would represent a lower challenge.

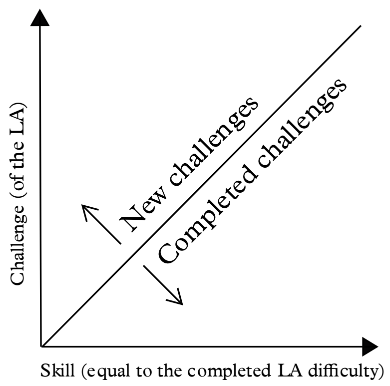


Figure 8. LA challenge vs. user skill.

For quantifying student performance, Difficulty is an internal measure quantized numerically from 1 to 10 that represents the LA challenge setting. This Difficulty does not necessarily match 1-1 with the student's perceived challenge of the task, but it is a simple way of representing a LA's variable challenge levels.

Linear LA challenge

LA challenge can be changed by adding more steps to the workload (see **Equation 1**) or by adding more complexity (see **Equation 2)**

$$x + y \ \boldsymbol{can\ become}\ x + y + z \qquad \textbf{Equation 1}$$

$$x + y \ \boldsymbol{can\ become}\ xy + wz \qquad \textbf{Equation 2}$$

Adding more steps or complexity can be translated into different subjects, for example in Geography, challenge can be increased by enumerating the question.

Which country is in Africa? *[Difficulty 1]*

a) Lesotho b) Ecuador c) Guyana d) Trinidad and Tobago

Can become,

Which three countries are in Africa? *[Difficulty 3]*

a) Lesotho, Togo, Djibouti

b) Ecuador, Surinam, Eritrea

c) Guyana, Burundi, Mauritania

d) Trinidad and Tobago, Morocco, Togo

[Answer – Lesotho, Togo, Djibouti are all in Africa]

Alternatively, complexity can be increased by requiring more detail

Which country is in Africa? *[Difficulty 1]*

Can become,

Which country is in Africa and on the equator? *[Difficulty 2]*

Alternatively, the mental workload can be increased by offering more possible answers

Which country does not belong in this group? *[Difficulty 3]*

a) Georgia b) Mongolia c) Cyprus

Can become

Which country does not belong in this group? *[Difficulty 4]*

a) Georgia b) Mongolia c) Cyprus d) Armenia

[Answer – Georgia is the only transcontinental country in the list.]

These are simple examples of how to achieve a simple linear progression in question and answer challenge - however, in more complex procedural skill acquisition tasks, the number of steps required to complete the task could represent the level of challenge, and this can be translated to almost any type of activity, for example, the number of steps in solving a physics problem, playing a musical score or performing a specific ballet move.

2.5.2 User Response Value

The user response outcome can be either 'correct,' 'missed' or 'wrong.' Polarities are assigned to the correct and incorrect answers so that the sum shows a better representation of user performance over a window of time (see **Equation 3**). This means between two users with the same amount of correct answers, the user with fewer wrong answers is awarded a higher accuracy.

$$User\ Response\ Value \begin{cases} +1 & if\ correct \\ 0 & if\ missed \\ -1 & if\ wrong \end{cases} \qquad \textbf{Equation 3}$$

Response Time

Response Time is the amount of time it took the user to respond from the moment a *new* challenge was presented to them.

User Response Accuracy

User Response Accuracy is an LA dependent measure for scoring user answers in the LA. It is calculated by dividing the sum of User Response Values by the total number of response opportunities, including the ones that have not been used (see **Equation 4**).

$$User\ Response\ Accuracy = \frac{\sum User\ Response\ Value \begin{cases} +1 & if\ correct \\ 0 & if\ missed \\ -1 & if\ wrong \end{cases}}{Total\ number\ of\ response\ opportunities} \qquad \textbf{Equation 4}$$

User Response Rate

User Response Rate is the ratio of challenges the user responds to (with either a correct or a wrong response) to the number of missed challenges (see **Equation 5**). The challenges included in this evaluation are the only ones that require a user response distinctively. For example, no response at all is required for a correct response; this challenge cannot be included in this assessment.

$$User\ Response\ Rate = \frac{User\ Responses\ that\ are\ correct\ or\ wrong}{User\ respose\ opportunities\ that\ are\ missed} \qquad \textbf{Equation 5}$$

2.5.3 Skill

Skill is an LA dependent user attribute that represents the average level of challenge the user has successfully mastered in each LA.

Measuring Skill

To calculate user Skill in a LA, User Response Accuracy for each LA challenge level will be looked at not as a percentage but a value towards the completion of the LA challenge scale (see **Figure 9**) calculated in **Equation 6**. In this way, 100% of LA challenge is looked at as 100 points, and 90% of LA challenge is looked at as 90 points, and so on. The sum of these points becomes the average skill performance of the user in this LA.

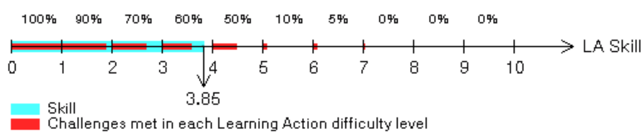


Figure 9. Average LA Skill.

$$user\ Skill\ in\ a\ LA = \frac{\sum_{Per\ LA\ difficulty} User\ Response\ Accuracy}{100} \qquad \textbf{Equation 6}$$

Using **Equation 6** for the above example shown in **Figure 9** we have:

$$user\ Skill\ in\ a\ LA = \frac{100\ +\ 90\ +\ 70\ +\ 60\ +\ 50\ \ +\ 10\ +\ 5}{100} = 3.85$$

User Skill can be used to compare user knowledge between two LAs from the same activity. Given that an LA can be implemented through different learning tasks the user's Skill in every LA is equal to the maximum value of Skill in its tasks (LA_i) seen in **Equation 7**.

$$user's\ Skill\ in\ a\ LA =\ Maximum\ of\ \{LA_1Skill, \quad LA_2Skill, \quad LA_3Skill \dots\} \qquad \textbf{Equation 7}$$

Prior Skill

Prior Skill is a user attribute that represents the user's previous level of knowledge. Prior Skill is both LA and time-dependent and is measured by calculating user LA Skill for a range of time in the past. A pedagogical expert that has knowledge could define this range and past experience of the user's learning rate. For learners with slower learning rates and longer LA Completion Times (see **Equation 10** for LA Completion Time) this time range should be respectively longer as well.

Aptitude

Aptitude is a user attribute that represents the ratio of current Skill as a percentage value of Prior Skill and it is LA dependant (see **Equation 8**).

$$Aptitude\ in\ a\ SLA = \frac{Skill}{Prior\ Skill} * 100\% \qquad \textbf{Equation 8}$$

This user attribute will be used alongside other sensor information in the 4 state Zone of Proximal Flow (ZPF) classification. Because Aptitude is a reflective value on past best performance, it has information on the state of the current user performance. Aptitude highlights whether the user's Response data is affected by boredom, fatigue or inattention.

Performance

User Performance is a local assessment of the speed of LA achievement. In other words, it demonstrates the rate of successful task completion. This is a descriptive attribute of user Skill and is dependent on the LA and the Difficulty settings involved. We define performance in **Equation 9**. While Skill gives a value to the overall user knowledge achieved in an LA, Performance gives a rate for those achievements.

$$Performance = \frac{Difficulty}{Response\ Time} \times User\ Response\ Value \qquad \textbf{Equation 9}$$

Performance can be used to compare the rate of user Skill achievement between LAs, and even between learners or different scenarios. Due to its dependency on LAs, it cannot be used to compare user Skill between two learners in two different scenarios - but it can be used to compare the ability of each user in their own scenario.

For example:

User A has a Performance of 70 in LA_1, and user B has a Performance of 50 in LA_2. This does not mean User A is more intelligent or capable than User B. However; it does mean that User A is performing better and achieving more of LA_1 than User B is achieving in LA_2. This comparison reflects on the qualitative experience of User A, where User A is more confident and capable in LA_1 than User B is in LA_2.

Completion Time

Completion Time is the time the user spends on the LA (see **Equation 10**). Specifically, the total time the user spends responding to answers, inclusive of correct, missed and incorrect responses before the current LA has been completed to a satisfactory level in which the system can move to the next LA. The user does not need to take a linear progression through the LA; they can skip LA Difficulty levels or even drop down some.

$$SLA\ Completion\ Time = \sum_{LA} \sum_{Difficulty\ level} Response\ Time \qquad \textbf{Equation 10}$$

Overall Performance

The Overall Performance is a global assessment of the average speed of task completion for mixed LAs (**Equation 11**). Overall Performance is dependent on the LAs involved and can be used to compare a single user performance between LAs and different sessions.

$$User\ Overall\ Performance = \frac{\sum_{SLA} \sum LA\ Difficulty}{\sum SLA\ Completion\ Time} \qquad \textbf{Equation 11}$$

It cannot be used in mixed LA comparisons unless the LAs in both samples are the same. In addition, similar to Performance, Overall Performance cannot be used to compare two learners, unless the LA samples are the same.

2.5.4 Performance comparisons

Local comparisons

Local comparisons are used to visualize the current user performance to previous steps. To determine if the user is performing better than moments ago, or if their performance is degrading. This type of comparison is useful in making an adaptive adjustment while the user is still interacting with the learning experience. User Performance can be evaluated in real-time for the current interaction and then a near point comparison could establish if the user's performance is degrading or improving. A comparison of user performance in two different levels of difficulty (A and B) is shown in **Figure 10** for a single LA.

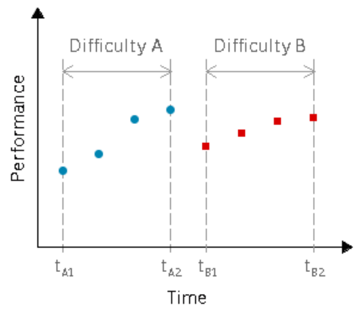


Figure 10. Local comparisons can be made within the same LA with different Difficulty levels. In this graph, a user's performance is shown for the same LA first with Difficulty A, and later with Difficulty B.

This comparison is restricted to interactions with LAs with the same challenge from the same LA, making it less useful to compare performance differences between LAs with two different levels of Difficulty. For cases where LA challenge comparison is required, a Global comparison method is more appropriate.

Global comparisons

A global comparison is used to compare user performance between sessions. It is different from a local comparison because it facilitates the comparison between user sessions of mixed LA Difficulty. Global comparisons could be used to determine three evaluation types.

1. Performance comparisons
2. Progress comparisons
3. Response Accuracy comparison

Performance comparison

Global performance comparison can demonstrate improvement or a decline in the user performance from one LA challenge to the next, or it could be used to make comparisons between user performance from one LA to the next. An appropriate index for this comparison is Overall Performance. Overall Performance has restrictions on the LA sample if a comparison is being made, both sides of the comparison are required to use the same LAs as seen in this example (**Equation 12**).

$$Overall\ Performance\ session\ 1_{LA_1} > \ Overall\ Performance\ session\ 2_{LA_1} \qquad \textbf{Equation 12}$$

Progress comparison

Progress comparisons can be used as a tool to understand what range of knowledge from one LA has been completed and compare that to another LA. An appropriate index for this comparison is Skill. With Skill, user progress in an LA can be evaluated. Skill can be used to compare progress on mixed LA paths. This information could be used in evaluating the non-linear learning progression and to determine the level of progression in each Learning Goal.

For example, if the user has completed some of LA_1 and LA_2 belonging to Learning Goal A and LA_3 and LA_4 belonging to Learning Goal B, the user Progress between the two Learning Goals could be evaluated using the total Skill acquired (**Equation 13**).

$$LA_1Skill + LA_2Skill > LA_3Skill + LA_4Skill \qquad \textbf{Equation 13}$$

Response Accuracy comparison

Response Accuracy can be a useful feature for detecting user affect state and determining when a user has reached an accuracy threshold. This threshold can be used to assess when a user has shown significant correct responses to challenges. When a user reaches the required accuracy threshold, they have acquired that Difficulty level in Skill. This threshold can be set differently for LAs in varying Learning Goals.

A comparison can also be used to evaluate the different levels of LA Difficulty with each other (as described in 0), or between learners to determine which user has progressed further in a collaboration scenario.

Reporting values

In any comparison, either local or global there may be a difference in sample quantity. For a robust evaluation, sample quantity must be taken into account for any reporting and comparison. For example, for a similar confidence level of 95%, a sample quantity of 50 values results in a confidence interval of only ± 3.24, while a sample size of 5, would nearly triple that interval to ± 10.24.

The range for the true population mean of 5 samples is 90 ± 10.24, while for 50 samples it is only 90 ± 3.24.

Therefore, with more samples, the precision of the assessment is much higher, and the evaluation, in turn, becomes more representative. This demonstrates the importance of always reporting any evaluation or comparison with confidence intervals.

The formula for confidence interval is shown in **Equation 14**:

$$Confidence\ Interval = z^* \frac{\sigma}{\sqrt{n}} \qquad \textbf{Equation 14}$$

Where σ is the standard deviation and n is the sample size. z^* (or z-score) is the confidence coefficient and can be looked up from Table 2 for the desired level of confidence.

Table 2. Confidence level to Confidence Coefficient lookup table

Confidence Level	z^*
.7	1.04
.75	1.15
.8	1.28
.85	1.44
.9	1.645
.92	1.75
.95	1.96
.96	2.05
.98	2.33
.99	2.58

Because a normal distribution never changes to simplify things, often a lookup table is used to find out the z-score of data points in a normal distribution.

2.5.5 Tracking flow through performance

Supporting learner engagement is important for deep learning and skill achievement. Students who are not engaged can become frustrated or bored, which can have a negative effect on achievement and lead to disruption in the classroom, which influences the learning of others. Importantly, many learning processes depend on a simple 'text' for the transfer of knowledge and evaluation. A single mode of learning can have limitations. For example, for a dyslexic student that has a reading-related learning disability, the single source of information transfer, therefore, becomes a problem [50]. Universal Design for Learning recognizes this problem [133] by embracing the pupil learning diversity by offering multiple means of learning accessibility. Multi-media learning platforms can use audio, audio text, video and tangible objects in a smart learning environment to offer the student a choice of the most accessible formats. This allows for multiple means of recognition, expression, and engagement [50], [133]. Multi-media approaches to learning resources are best demonstrated in the use of computer-mediated learning where educational games are developed around the learning outcomes and aims. This approach, which is not new, was shown in a systematic review of 129 papers by Connelly *et al.* [134] and showed playing educational games impacts across a range of areas including engagement, cognitive ability and, most commonly, knowledge acquisition and content understanding.

Zone of Proximal Flow state transitions for optimal learning

How do we apply affect knowledge to a learning platform? Guided by the affected state and the ZPF state diagram of a learner (**Figure 11** and **Figure 12**), the appropriate level of learning material challenge in order to maintain an optimal learning condition where both flow, and Skill Achievement, are maximized, can be determined. 'Flow' and 'ZPD arousal' learning states are the active learning states of the learner. New skill acquisition and skill uptake maximization happen in ZPD while maintaining the learner in the state of flow and providing the opportunity for reinforcement learning (as visualized in Figure 11), which can solidify skills acquired during the learning process and enhance the learning experience itself. Although new skill is not acquired in flow, a slow parallel growth over the long term, with the increase of the level of challenge, introduces an increase in learner skill. This is however limited to the lower ZPD independent learning limit, and to increase skill further beyond that, the learner must enter the ZPD.

Any learning material challenge adaptation processes need to maintain the learner in the optimal path, as portrayed in Figure 11. The optimal path is the one with the shortest forecasted achievement time and one that facilitates the most positive affect states. This path must take the learner through arousal and avoid boredom or frustration. It should start at the lower limit of the ZPD (familiar base) to avoid unnecessary repetition while allowing the learner to remain in the state of flow to enable the reinforcement of acquired knowledge (reinforcement learning) visualized in Figure 12**.** This loop of leaving flow and entering the ZPD (shown as a snakelike pattern in Figure 11) should continue until the maximum possible skill achievement is obtained (highlighted by the upper ZPD limit). The caveat being that the 'familiar base' (starting concept) should be challenged for specific learners with a disability to re-evaluate previously established learning outcomes.

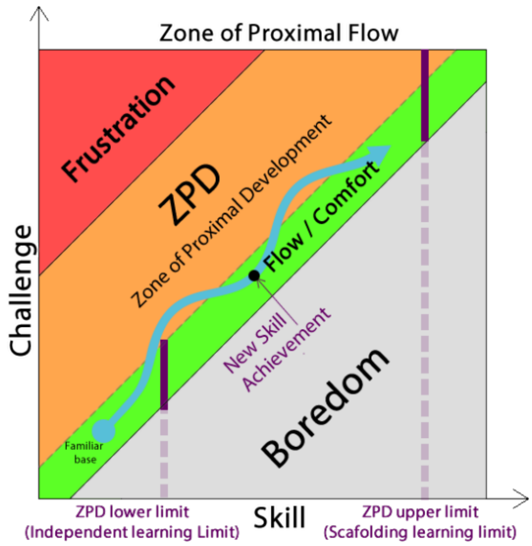


Figure 11. Optimal learning experience loop in ZPF diagram adapted from Basawapatna et al [40].

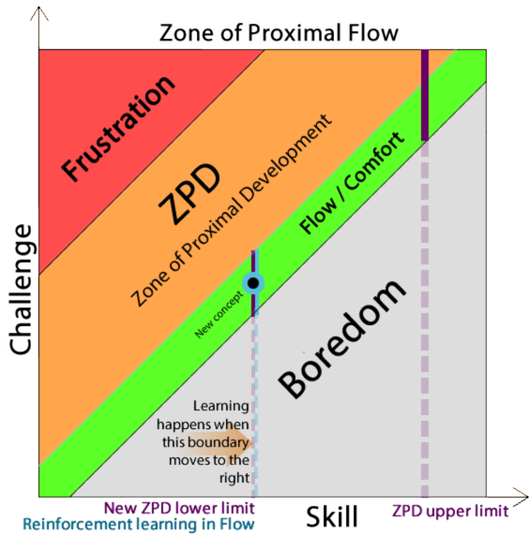


Figure 12. State change is paused while reinforcement learning takes place adapted from Basawapatna et al [40]. The ZPD lower limit is moved to reflect the new skill achievement.

This process is a delicate one, when the learner is in flow, a continuous effort to push the learner out of their comfort zone and into the ZPD zone by challenging them to higher levels of challenge will stimulate the learner. However, if the learner is projected too far into arousal, the learner becomes frustrated. By monitoring the learner affect state and learner skill level, a learner can oscillate between the ZPD and the state of flow with new skill materialization. As a result, a new concept materializes as the learner's skill with just-in-time principles as displayed in Figure 12 as a circle in the green area. Adaptive learning requires continuous monitoring of learner affective state and LA progress. The learning path is far more engaging and optimal, and the learner is always in a positive affect state. Maintaining the learner in the state of flow provides the opportunity for reinforcement learning.

To conclude, the broader literature shows that there is a strong relationship between learning goal outcome achievement and a learner's affect state. Positive affect state has been shown to encourage greater learner outcomes and sustained engagement, which lead to deep learning and long-term skill retention. Prior skill, learning challenge and learner performance are quantifiable characteristics of the learning experience. They can be used to determine the learner affect state through the mapping of the Zone of Proximal Flow (ZPF) state diagram. This is important to this work, as it provides a quantifiable base to track learner affect by measuring their performance, at a known level of challenge. In the next section, we discuss existing methods of affect assessment for learners with disabilities, which are prone to inter-rater reliability issues and cannot be automated.

2.6 Challenges with understanding learner affect with disabilities

Abrams (1986) stated [135], "The vast majority of children with learning disabilities have some emotional problem associated with the learning difficulty." Traditionally, however, teachers have prioritized the diagnosis and remediation of learning disabilities [136]. Empirical data suggest the critical need to deal with the emotional aspects of learning difficulties and of course this depends on reliable methods of determining this state. PMLD and CP require assistance with a variety of daily activities and, depending on the level and types of difficulty, have challenges with their cognitive and motor functions and speech. Care workers and teachers rely on their interpretations of the levels of engagement and interest to personalize these experiences for them. This is even the case when using well-defined engagement Scales (e.g., SSAT Engagement Scale [137], [138]) as the ratings for each component of the scale are still subject to inter rater interpretation.

Studies have considered self-reported affect states as the ground truth for inter-rater agreement studies [41], [139]. These studies have looked at the level agreement and correlation between self-rated affect states and peers, clinicians, and long-term partners. The level of correlation even though significant between the 40th and 70th percentiles [41], [139], still leaves room for improvement. In addition, self-rated affect states may carry bias or not be representative of the true affect state. Therefore, an automated method that would base its ground truth on self-rated affect states would thus be impacted by such bias and unknown reliability factors. The validity of a machine learning method based on clinician, or peer-rated affect states would inherit even greater bias, reliability, and interrater reliability uncertainty, as it is one more level separated. Importantly, a machine learning method with 100% classification accuracy trained with clinician-rated affect data would at best achieve around 70% correctness of the self-rated affect states. Furthermore, the self-rated affect states may themselves have a bias or be unrepresentative. This creates a problem for both the clients and care workers as it has been shown that observation is not a reliable method of determining a person's mood and affect state [6]. This can only be more intensified with PMLD or CP, as their behaviors, body language and voice may not have the same cues as mainstream people.

Moreover, the levels of skill and experience between care workers and teachers vary widely, as does their capacity and accuracy of interpretation of others' behaviors. This uncertainty of interpretation and inaccuracy in the observation of the affect state of a person experiencing PMLD or CP (mood and emotional well-being) can be detrimental to their quality of life [5], [7], [8]. Hence, the well-being of a student with PMLD or CP can be improved if their levels of interest and engagement could be determined and tracked by more independent and repeatable means, such as using technology, and in our case sensors. This added interpretation of a student's state of affect is not meant to replace teachers' or carers' interpretation, but more to augment this judgment. Monitoring a person's level of interest and engagement in an activity allows carers, teachers, and parents to be responsive to those levels. An objective approach to the reporting of engagement is the use of a standardized test to monitor for indicators of flow.

We demonstrate the possibility of tracking and then modeling body movements, eye-gaze, Electroencephalogram (EEG) and interaction data from learners with PMLD and CP to track their level of attention, as a good indicator of what interests them and positively influences the quality of that experience. In this study, we investigate the ability of sensor-based technology to detect and track sustained attention in a repetitive demanding activity, with a Multimodal Multisensor platform. This allows us to make inferences on the attention level of the student throughout the length of the activity through their responses to the challenges presented in the repetitive activity.

2.7 Continuous Performance Test and Signal Detection Theory

The CPT was first introduced by Swanson as a standardization of the SDT in the 80s [15][140] to study vigilance and sustained and selective attention in children with learning difficulties. The CPT is reported to be the most popular measure of sustained attention or vigilance [15]—the ability to sustain attentional focus and remain alert to stimuli over time [141], [142]. The basic paradigm of a CPT involves selective attention or vigilance for an occurring stimulus [1]. A rapid presentation of continuously changing stimuli (like letters, numbers or images) is presented on a display among which there is a designated "Target " pattern [143], which the participant is tasked to respond to (e.g., a button press) [142], [144]. The duration of the task varies, but the task is intended to be of sufficient length to measure sustained attention. The CPT slide is also called a SDT trial or signal [145]. The SDT Target is also called an SDT primer [146]. The SDT noise trial is also called a distractor cue [146]. In this work, we present two types of distractor cues, the Imitation and the contrast, see section 3.2, Experimental platform.

Significant correlations between CPT scores and teacher ratings of inattention, impulsivity, and hyperactivity have been established [147]–[149], thus providing some evidence of the affective relation and validity of these measures as indicators to the user's affect state. Therefore, there is an opportunity to use the CPT scores to label a Multimodal Multisensor data stream. This could then be used as an objective measure of learner attention to support teachers. From the CPT, typically three scores are derived: (a) the total number of Hits (Correct Commissions), (b) the number of Target stimuli Missed (Wrong Omission), and (c) the number of non-Target stimuli to which the subject responded (False Alarms). The total Hits and Correct Omissions are seen as measures of sustained attention whereas False Alarms are linked to impulsivity [150], [151].

Signal Detection Theory or SDT is the basis for learner performance analysis in the CPT. SDT, a theory initially developed to qualify radar detectors by Marcum in 1947 [14], was later theorized by Peterson, Birdsall, and Fox in 1954 [152] and adapted as a psychological theory by Tanner and Swets in the same year [153]. It was only in the 90s that SDT was expanded with analytical methods to evaluate bias in detail, by Swets and Pickett [154] and by Macmillan and Creelman [155]. A CPT is made of SDT trials. The trials can either be signal + noise or just noise. The amount of noise in the trial is what determines the challenge of the trial. On signal trials, yes responses are correct and are designated hits. On noise trials, yes responses are incorrect and are termed False Alarms. The hit rate is the probability of responding *yes* on signal slides and the false-alarm rate is the probability of responding *yes* on Imitation Target or Contrast slides [145], [156]. An example of a CPT outcome is shown in **Table 3**. For the rest of this section we use the numbers in this table to calculate the SDT analytical methods.

Table 3. Confusion matrix example outcomes of a CPT.

	⇓ What is on screen ⇓		
⇓ Student Response	⇓ Target ⇓	⇓ No Target ⇓	Total participant decisions ⇓
Participant Press ⇒	**Hit or Correct Commission (CC)** = Participant presses and there was a Target on the screen. 25	**FAR or Wrong Commission (WC)** = Participant presses and there was No Target on the screen. 7	**Total Presses (TP)** ⇓ 32
Participant Omission ⇒	**Miss or Wrong Omission (WO)** = Participant Omission and there was a Target on the screen. 3	**Correct Omission (CO)** = Participant Omission and there was No Target on the screen. 13	**Total Omissions (TO)** ⇓ 16
Total slides⇒	**Total Targets (TT): 28**	**Total Imitations (TI): 20**	**48**

SDT has often been used as a means to assess learner characteristics [148], [149], [157]–[160], especially attention, vigilance and inhibition in learning disabled and CP population [1], [15], [140], [143], [150], [158], [160]–[172]. The CPT and commercial versions ('TOVA' and 'QbTest') of it have been used to detect attention and inhibitory control and to diagnose Attention Deficit Hyperactivity Disorder (ADHD) [16], [173]–[176].

In this study, we have adopted the basic requirements for the CPT and the constructs of the methodology but adapted the test to facilitate our PMLD and CP learners. Using Signal Detection Theory (SDT) [14], [15], [140], [152]–[155], quantifiable objective data on the improvement or deterioration of attention is collected and analyzed using SDT analysis theory detailed in [154], [155].

2.7.1 Hit rate

Hit rate (H) is the probability of a yes response given the Target is present. H can vary between 0 and 1. H is calculated in **Equation 15**.

$$H = P\,(yes|present) = \frac{\text{Correct Commissions}}{Total\ Target\ slides}$$ **Equation 15**

For the example from the confusion matrix: H = 25/28 = .89

2.7.2 False Alarm Rate

False Alarm Rate (FAR) [also shown as F or FA] is the probability of a yes response given the Target is absent. FAR can vary between 0 and 1. FAR is calculated in **Equation 16**.

$$FAR = P\,(yes|absent) = \frac{\text{Wrong Commissions}}{Total\ non\ Targets\ slides}$$ **Equation 16**

For the example from the confusion matrix: FAR = 7/20 = .35

2.7.3 Sensitivity d'

Sensitivity was introduced by Swets and Green in [177], and is a measure of the quality of participant performance in a CPT or, in other words, the ability of a participant to maintain sustained and selective attention is represented by d'. It is measured in standard deviation units [154] and is the distance between the center of the normal distributions of Correct Commissions (CC) and Correct Omissions (CO) seen in **Figure 13**. The formula for d' can be seen in **Equation 17**, where z is the z-score or the z transformation, a function that transforms probabilities to normal distribution units. A perfect participant would have a Hit rate (H) of 1 and a False Alarm Rate (FAR) of 0 and would have a d' of +∞. Elliott tabulated a measure of sensitivity (d') in [178].

$$d' = z(H) - z(FAR)$$ **Equation 17**

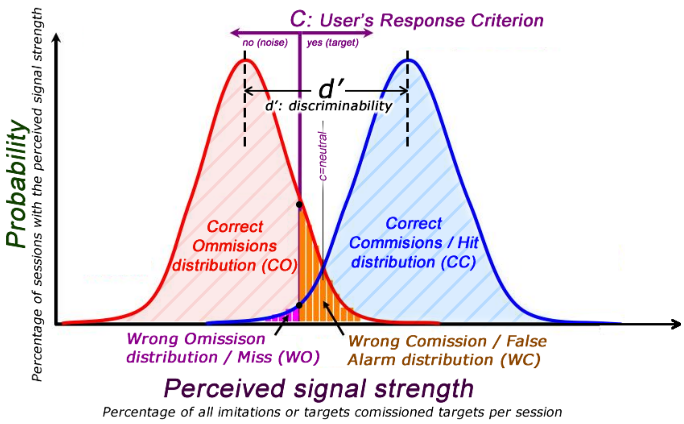


Figure 13. Response criterium of the participant.

A z-score is a measure of how many standard deviations below or above a population mean a data point is. The area under the normal distribution from the left tail is equal to the probability of a complete distribution falling within this range. It is logical that the total area under the normal distribution is 1 (100/100) because 100% of the total distribution exists under the general area of the normal distribution. The z-score of any given raw value (x), in other words, is the probability that all values are equal to or less than x. This probability, as discussed, is also equal to the area under the standard normal curve from the left tail to the vertical value of x, as shown in **Figure 14**. A table of z-score conversions is seen in **Table 2**.

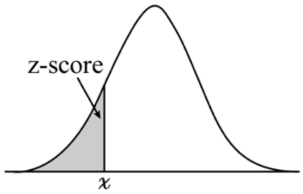


Figure 14. A z-score is a measure of how many standard deviations below or above a population means a data point is.

Alternatively, there are commonly built-in functions that calculate the z-value in excel and popular programming languages. For example, in the Python programming language [179] there is the property density function [33], which can be used to calculate d' as shown in **Equation 18**.

```
import scipy.stats as st

def d_prime(H, FAR):
  ZH = st.norm.ppf(H)
  ZFAR = st.norm.ppf(FAR)
  d_prime = ZH - ZFAR
  return d_prime

# returns 1.6119
print(d_prime(0.89, 0.35))
```

Equation 18. Python function that uses the scipy.stats library to calculate $\boldsymbol{d'}$ using the z-score values of H and FAR.

In the absence of a built-in function for z-value, an algorithm for approximating the z-values has been presented by Bromphy (and described in his 1986 paper) [178]. The maximum absolute error of z-values estimated by the program is .00020, the maximum absolute error of Beta bias values calculated by the algorithm is .0003 [180] and the values of d' have a maximum absolute error of .0004. Bromphy's algorithm from the original in Basic is translated to Python in **Equation 19**. Using the values from the confusion matrix (H = .89, FAR = .35), we have d' = 1.61.

```python
import math

h = float(input("What is the Hit rate:"))
fa = float(input("What is the False Alarm Rate:"))
# Bromphy's estimation algorithm
def z_value(prob):
  if prob > 0.5:
    prob = 1 - prob
    k = 1
  else:
    k = -1

  if prob < 0.00001:
    z = 4.3
  else:
    r = math.sqrt(-math.log(prob))
    z = (((2.321213*r + 4.850141)*r-2.297965)*r-2.787189)/((1.637068*r+3.543889)*r+1)

  y=1/math.sqrt(2*math.pi)*math.exp(-z*z/2)
  z=z*k
  return [z,y]

# Calculate the Bommphy estimations for d-prime and bias
print('ZH Bromphy=', z_value(h)[0])
print('ZF Bromphy=', z_value(fa)[0])
print('d prime estimation Bromphy=', z_value(h)[0]-z_value(fa)[0])
print('beta estimation Bromphy=', z_value(h)[1]/z_value(fa)[1])
```

Equation 19. Bromphy d', Beta bias and C calculation translated from Basic to Python.

2.7.4 Bias B''_D

Bias represented by B''_D [181] is a way of measuring whether a participant is liberal or conservative in their Commissions against the signal. It is a good way of contrasting the participant outcome result against how much risk they are willing to take. For example, low mistakes could mean the user has not made many attempts (if they are conservative) and a high Hit rate could mean they are always pressing the button (if they are liberal). Neutral bias (no bias) is indicated by $B''_D = 0$. When $H + FAR = 1$, B''_D (bias) is always neutral (or 0). Negative numbers represent liberal bias, positive numbers represent conservative bias, and the maximum in either direction is 1.0. The formula for B''_D is represented in **Equation 20**. Bias outcomes are shown in **Table 4**. The outcomes of bias for learning disabled students are reported to be more conservative and mainstream students are more liberal [1], [143], [151].

$$B''_D = \frac{[(1-H)(1-FAR) \; - \; H \times FAR]}{[(1-H)(1-FAR) \; + \; H \times FAR]} \qquad \textbf{Equation 20}$$

Table 4. Bias outcomes of SDT.

$B''_D < 0$ (Liberal)	$B''_D = 0$ (neutral)	$B''_D > 0$ (Conservative)

If H and FAR are plotted in a two-dimensional plot, the diagonal across would be where neutral bias is separated from the conservative and liberal biases. The dynamic of how bias interacts with H and FAR is shown in **Figure 15**. The higher the Hit rate and the False Alarm Rate the more liberal the participant is and the lower the Hit rate and the False Alarm Rate, the more conservative the participant is. SDT trials that have outcomes of equal Hit rate and False Alarm Rate are considered to have a neutral bias (B''_D).

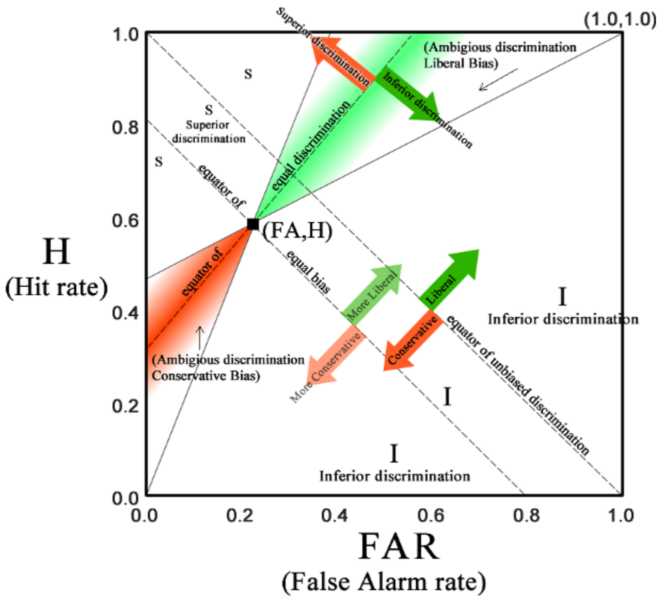


Figure 15. Bias shown with respect to H and FAR.

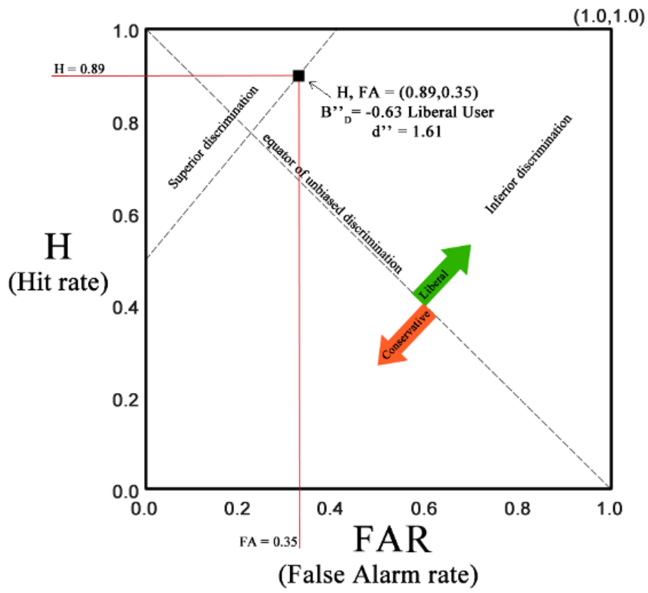


Figure 16. Bias graph for H value .89 and FAR value .35. The graph shows that the user has a liberal bias.

Neutral bias (no bias) is indicated by $B''_D = 0$. When H + FAR = 1, B''_D (bias) is always neutral (or 0).

Negative numbers represent liberal bias, positive numbers represent conservative bias, and the maximum in either direction is 1.0.

Using the values from the confusion matrix (H = .89, FAR = .35) for B''_D we have:

$$B''_D = [(1\text{-}0.89)(1\text{-}0.35)\text{-}0.89*0.35]/[(1\text{-}0.89)(1\text{-}0.35)+ 0.89*0.35]$$

$$=(0.11*0.65\text{-}0.3115)/(0.11*0.65+0.3115)$$

$$= \text{-}0.24/0.383 = \text{-}0.63$$

The Bias graph for this example shown in **Figure 16**, describes a user with better than chance signal detection ability and a liberal bias.

2.8 Motivations for tracking attention using CPT outcomes

In the Literature review, we discussed the challenges of tracking engagement levels of learners with PMLD/CP. There are paper based methods [182], [183] but they require subjective observer or peer rated values that are subject to inter-rater reliability. This has been shown to cause attribution errors [6] which in turn lowers PMLD and CP care quality [5], [7], [8]. These assessment scales are only suitable as a global measures, and are not suitable to moment to moment tracking of learner engagement state [182]. To assess learner experience quality, we look to flow.

There are many studies in the literature that relate flow a closely related affect state of engagement to greater learning outcomes [73]. One is in flow, when one is engaged [32] and steady performance has been maintained at the comfortable limits of one's skill limitations [64]. Maintaining flow in learning is especially significant because it is the most reliable indicator for determining successful learning [36], [45]-[48]. This results in immersion, concentrated focus and deep learning [65], [66]. Flow is central to classroom performance and the achievement of learning outcomes [45]-[48] which is closely linked with attention [16], [32], [64]. During flow, attention is completely absorbed in the task at hand, and the person's performance is maximized [73]. The ability of a person to sustain attention, is often coincides with inhibition, which increases performance [73], [77].

Tracking attention is important to this work as it is an indicator of performance, and flow, which is linked to greater academic outcomes. When Attention is completely absorbed in the task at hand, the person performs at their greatest capacity [73], [84]. Studies have found direct correlation between attention and academic performance[85]. The ability of a person to sustain attention, is often combined with self-control and inhibition, which enormously increases performance [65], [66], [73], [77]. Studies have found a strong positive relationship between flow and performance. For example, flow is positively associated with artistic and scientific creativity [86], [87], effective teaching [66], learning [88] and peak performance in sports [89], [90]. This positive relationship between flow and skill development has been demonstrated in a number of studies in which students were tested in a school environment achieving a higher grade point average (GPA) [73], [88], [90], [92].

Therefore, performance trend tracking can be used as an indicator of flow [64]. This approach has been used in [34], [40], [64], [95], [97]–[100] as a model for relating learner affect to user performance in a pre-defined activity/task with known challenge. Continuous Performance Tests are test that track sustained attention (vigilance), selective attention [147], [160], [165], [184]–[186] and inhibition [147], [184], [185], [187], [188]. Commercial versions of the CPT; QbTest [189], Test of Variables of Attention (T.O.V.A.) [176] and Conners CPT3 [190] are used to detect Attention Deficit Hyperactivity Disorder (ADHD) by tracking hyperactivity and inattention [191]. The SDT and CPT outcomes; Misses and False Alarms, mean reaction time, d' sensitivity and B''_D bias have also been used to understand learner cognitive processing ability [1], [192], [193]. CPT outcomes such as Misses indicate inattention and False Alarms measure impulsivity [150], for a more detailed overview see section 2.7. Greater CPT outcomes have also been related to academic achievement [158], [194]. Others have used the concept of continuous performance monitoring of game outcomes as a method to infer affect state [40], [64], [80], [97], in combination with the ZPF theory [40]. The CPT has also been used in combination with body pose and head tracking data to track attention [195], in another study CPT was used to label multimodal sensory data with objective attention labels [196].

For these reasons we propose a Multimodal Multisensor platform incorporating a CPT. The development of the Multimodal Multisensor platform using the CPT test outcomes as objective labels for the sensor data, will assist the creation of a model of participant attention. Multimodal Multisensor data is collected and labeled using Signal Detection Theory outcomes from a Continuous Performance Test. Features are then extracted and used in a machine learning model to classify participant moment-to-moment attention state (**RQ1**). Longitudinal SDT and CPT data is analyzed for correlations with participant physiological characteristics and academic scales (**RQ2**). The development of an Multimodal Multisensor sensory platform addresses the limitation of previous approaches with PMLD/CP learners, that required observer rated outcomes, that provided only global outcomes. Consequently, this approach minimizes subjectivity and provides an objective, automated measure for attention classification.

2.8.1 Multimodal advantages and studies

The detection of affects is an important problem of pattern recognition that has inspired researchers from various fields. A study assessed 90 peer-reviewed Multimodal Multisensor (MM) systems, the review indicated that the state of the art mainly consists of person-specific models (62.2%) that fuse audio and visual (55.6%) information to detect acted-out expression of basic emotions (52.2%) and simple dimensions of arousal and valence (64.5%) with feature- (38.9%) and decision-level (35.6%) fusion techniques [38].

Multiple sources of sensor data can be combined to improve the identification and classification of affective states [38]. One example of such an approach demonstrated improved classification results when, comparing single modality data to multimodal data in visual emotion recognition from face and body sensor data [197]. Over 85% of the studies reviewed resulted in the improvement of classification outcomes compared to single modality [38]. This provides important evidence that multimodal classifiers outperform single modal alternatives.

While single modality detection involves the use of a single mode of sensor data (e.g., eye-gaze, facial features, body pose), Multimodal Multisensor systems fuse two or more modalities for affect detection. However, collecting Multimodal Multisensor data is not without its challenges. Björn Schuller proposes additional stages [198] in "Multimodal Affect

Databases—Collection, Challenges, and Chances." The chapter discusses the challenges of collecting and annotating affect data, particularly when more than one sensor or modality is used. Schuller's 10 steps highlight the most important considerations and challenges, including (1) ethical issues, (2) recording and reusing, (3) meta-information, (4) synchronizing streams, (5) modeling, (6) labeling, (7) standardizing, (8) partitioning, (9) verifying perception and baseline results, and (l0) releasing the data to the broader community. D'Mello *et al.* report the main challenges as the following: (a) deciding which modalities to combine; (b) collecting Multimodal Multisensor training data; (c) handling missing samples, fusing different sample rates, and retaining modality interdependence when training models; (d) determining the fusion strategy between different modalities; and (e) deciding how to evaluate Multimodal Multisensor affect models [38].

Moreover, the process of recovering the user intent through multiple different input sources and their potential combination, known as "multimodal input fusion," presents several challenges to be overcome before multimodal interfaces can be experienced to their fullest. Firstly, designing computer architectures to manage parallel Multimodal Multisensor input streams efficiently, as well as to maintain a reliable and stable connection between sensors and platform across the multiple connection types (LAN, Bluetooth, USB). Secondly, the type of data to be managed by a Multimodal Multisensor system may originate from a variety of different sources. For example, in this study, an EEG device connected via Bluetooth is sending surface skin voltage readings along the forehead, and an eye-gaze tracking device is relaying coordinates relative to the eye-gaze focal point at a fixed corner on the monitor screen. The redundancy of Multimodal Multisensor data is one of the strengths of multimodal interfaces.

However, Multimodal Multisensor systems have many advantages over their single mode counterparts. The advantages of a Multimodal Multisensor approach to affect detection include (a) capture of higher dimensionality of human affective expression, (b) data redundancy against occlusion and data fall-out, and (c) a potential solution to signal noise which is more prevalent in single modality approaches [199]–[201]. The ability to create a higher dimensionality in the feature space by using multiple signals and their interdependence gives way to the creation of models that more accurately mimic the true nature of human affective expression [38]. Also, single modal systems are affected more severely by occlusion or data dropout [38]. For example, in an affect detection system that uses facial expression, the model becomes unreliable when the user's face is turned away from the system or is covered (occluded) by an object. However, a Multimodal Multisensor system would be able to use other sensors to recover sensory data and provide a more continuous affect detection by basing their decisions on the existing channels.

Studies have surveyed the existing Multimodal Multisensor studies and presented meta-analysis [21], [25], [38], [102], [202]–[204]. Examples of some of these affect sensitive Multimodal Multisensor systems were included in the context of the literature review but here we list some of the most relevant studies:

Self, peer or expert annotated studies

1. The system of Lisetti and Nasoz [205], which combines facial expression and physiological signals to recognize the user's emotions, like fear and anger, and then to adapt an animated interface agent to mirror the user's emotion.
2. The multimodal system of Duric *et al.* [206], which applies a model of embodied cognition that can be seen as a detailed mapping between the user's affective states and the types of interface adaptations.
3. The proactive HCI tool of Maat and Pantic [207], which is capable of learning and analyzing the user's context-dependent behavioral patterns from multisensor data and of adapting the interaction accordingly.
4. The automated Learning Companion of Kapoor *et al.* [208], which combines information from cameras, a sensing chair, a mouse, a wireless skin sensor, and task state to detect frustration in order to predict when the user needs help.
5. The multimodal computer-aided learning system in the Beckman Institute, University of Illinois, Urbana- Champaign (UIUC) [23], where the computer avatar offers an appropriate tutoring strategy based on the information of the user's facial expression, keywords, eye movement, and task state.
6. Picard *et al.* modeled affect states of participants using an auto tutor [209]. Eye-gaze, body sitting posture and textual communication were used in combination with a k-NN classifier to achieve 70% affect state classification.
7. Features extracted from keyboard and mouse typing behavior has been used in a study to detect affect state [210]. A self-assessment scale was used to label participant affect states as 'boredom', 'frustration', 'distraction', 'relaxation' and 'engagement'. Using a J-48 classifier they obtained an accuracy of 80.48%.
8. Lane position, vehicle dynamics and a combination of eye gaze and head dynamics allowed a prediction of lane changing intent to a true positive rate of 87.3% and a false positive rate of 0.39% [211]. The study found that during an attention shift, head movement precedes eye-gaze shift, which is consistent with the biological model of attention shifts.
9. D'Mello *et al.* had 28 participants interact with an auto tutor while facial image data, body pressure patterns and participants' verbal responses to the auto tutor give way to a data set with 29 Multimodal Multisensor features [212]. Expert trained judges hand labeled the data as boredom, confusion, engagement/flow, and neutral. Their classifier achieved an average of 48.65% accuracy with a kappa of .335.
10. In a study by Monkaresi *et al.*, 22 students participated in a structured writing activity while heart rate estimation from facial image data, face animation units from Microsoft Kinect, and local binary patterns of the eyes and mouth were utilized to predict student self-reported engagement states. Their classifier achieved an overall area under the AUC coverage of 75.8%.
11. In a follow-up study D'Mello *et al.* had a group of 20 students interact with an online math learning platform. In this study student "behavioral engagement" (on-task, off-task) [213], [214] and "emotional engagement" (satisfied, bored, confused) [215] were labeled using a combination of participant self-labeling and expert labeled data. To achieve this, two modes of data were collected in the background, 3D facial image data from an Intel® RealSense™ F200 camera [215] (facial expressions, emotions) and interaction data with the learning platform (performance, response time). Their walk-up-and-use (generic) [216] random forest classification achieved 48.12% and their person specific classification achieved 76.37% accuracy.

The final listed work has some similarities to ours, however with three key differences. Firstly, the emotional engagement ground truth labels are hand labeled by the participants and human experts which is susceptible to inter-rater reliability issues. Secondly, the learning platform activity outcomes are an input of classifier model, this makes the model depended on the activity outcomes – hence, its predictions are never in real time, or generalizable to activities other than the exact one in the study. Lastly, while this study explores two modes of data input (one dependent), our study explores 4 independent modes of data, making it more resilient against data fallout, and independent of the data collection activity. In our study the labels are produced by the objective outcomes of the CPT. This makes the ground truth immune to inter-rater reliability issues which human observer labeled data is prone to. In the next section we explore studies that have used CPT as an objective method of labeling affect data.

Studies using objective labels via CPT outcomes

1. Body posture and head position from the Kinect were combined and labeled using objective outcomes of a CPT to detect inattention and hyperactivity characteristics in children with ADHD [195].
2. Theoretical study that proposes the use of multimodal sensory observations labeled objectively with CPT outcomes of attention in a dynamic Bayesian networks to detect 'fatigue', 'nervous' and 'confused' affect states [196].

Multimodal interaction systems aim to support the recognition of naturally occurring forms of human language and behavior through the use of recognition-based technologies [217], [218]. These systems represent initial efforts toward the future human centered Multimodal Multisensor HCI. Most of the methods rely on observer or listener judgment and results can vary depending on the level of expertise, experience and personal style of the observer or listener. This can cast doubt on the validity of such methods as they are susceptible to inter-rater reliability. This can cause attribution error which lowers PMLD and CP care quality [5], [7], [8]. Others that use the CPT as an objective method, are using a single sensor, or are proposing theoretical studies.

In the next section, we describe our study methodology, our use of a CPT to objectively label Multimodal Multisensor data to model learner attention levels.

Chapter 3. METHODOLOGY

As discussed in the previous chapter, there are several approaches to labeling the affective state of learners with PMLD. However, these are based on paper-based tests, which are subject to intra-rater reliability issues. There are also Multimodal Multisensor sensor-based approaches, but these rely on a teacher or student self-ratings to label data.

Our approach is to address these issues by labeling the sensor data using a CPT because the correlation between the outcomes of a CPT and teacher ratings of inattention, impulsivity, and hyperactivity have been established [147]–[149], thus providing some evidence of these measures as indicators of the user's affective state.

A gamified *platform* is proposed that monitors the sustained and selective through performance tracking using SDT [14] measures and outcomes. For the remainder of the work, we will refer to this attention tracking platform as *'the platform'*.

The participant is required to pay continuous *attention* to a computer screen where an *interactive* game provides them with a pre-defined signal detection *challenge*. The participant is in *control* of the response they give, and *feedback* is given to them regarding the correctness of their response to the *challenge*. This is the basis for Swanson's CPT [15]. The CPT is an integral component of *the platform,* and we have therefore created a version, the 'Seek-X' type. This test has been created to be used specifically as *an objective tool* for attention tracking using the CPT outcomes to label multisensor data.

We have named this CPT Seek-X type because the participant is asked to *seek* the Target image between other non-Target images acting as a matrix of noise. Seek-X type exercises engage eye-gaze as a crucial element of answering the SDT challenge. The Seek-X type CPT is a version of the Type-X non-rare Target type, see **Equation 21**:

$$\text{CPT types} = \begin{cases} By\ challenge \begin{cases} Type - X \longrightarrow Seek - X \\ Type - AX \longrightarrow Seek - AX \end{cases} \\ By\ target\ frequency \begin{cases} Rare \\ Non - rare \end{cases} \end{cases} \qquad \textbf{Equation 21}$$

In summary, the period of sustained attention is marked by participants' *attention* and *interest* being maintained in an interactive interaction. Maintaining sustained attention indicates the key foundation for recognizing flow. For this reason, this work explores classical methods for attention tracking using a neuropsychological test that measures a person's sustained and selective attention (the CPT) [15]. The CPT is reported to be the most popular measure of sustained attention or vigilance—the ability to sustain attentional focus and remain alert to stimuli over time [141], [142]. The first attempt to objectively evaluate the relationship between maintaining attention in students with learning disabilities using CPTs was introduced by Swanson in [15], [140] and later expanded by Eliason and Richman [1]. Using SDT [14], [15], [58]-[62], quantifiable objective data on the improvement or deterioration of attention are collected and analyzed using SDT detailed in [154], [155].

We explore modeling the user's attention indicators without any assumptions about the activity in front of them. A scalable CPT was developed to be appropriate to the capacity of the PMLD / CP participants. Participants were asked to sit in front of a PC and press a button when they saw the predefined signal (a predefined Target) on the screen, and not press the button when any other signal was displayed on the screen. This happened while the sensors recorded their real-time sensor data, including upper limb positional data, EEG brain waves, and eye-gaze data.

3.1.1 Data collection

Four participants were recruited for data collection (see section 3.4, Participants). They took part in a 13 weeklong study with up to four sessions weekly per participant, depending on participant availability. This process started on the April 4th, 2016 and finished on the July 21st, 2016. In the first two weeks, a pilot study was carried out to tweak the test difficulty and make any final alterations to the experimental platform. These alterations are discussed in section 3.2.1, Evolution of the CPT design.

Each session included 48 challenges. Each test lasted between 6-32 minutes depending on participant readiness or other setting-up challenges. Every session recorded nearly 4 minutes of data. A total of 59 sessions of the CPT were carried out (an average of 15 sessions per participant). A series of 48 slides with pauses in between were displayed for each participant.

This CPT design was based on Rosvald and Mirsky's original paper [193]. Recommended time alterations to the experiment length were made to match the shorter length activities that students with PMLD are accustomed to at school [15]. The CPT was therefore shortened to about 4 minutes for our participants, and the whole process took between 6-32 minutes. This is compared to other research, which suggests a 30-minute CPT for neurotypical participants [219].

The difficulty of the CPT was also adapted for each participant by making the maximum response time (slide display time + blank slide display time) shorter or longer or by adjusting the image matrix grid size. These times are initially 1.8 s and 1 ± 0.1 s, respectively, and are increased or decreased depending on participant capacity. The difficulty of the CPT was adapted for each participant by making the maximum response time shorter, the maximum response time is the sum of the 'stimulus duration' and 'interstimulus interval' [154]. These durations (seen in **Table 5**) were estimated in a series of pilot study tests completed in the initial two weeks of the study, where the aim was to reach close to the 85% rule for learning, where the participant makes around 15% mistakes and 85% correct responses [220] when in flow. The platform setup can be seen in a school setting in **Figure 17**. The Seek-X type CPT slide timeline is demonstrated in Figure 18.

Figure 17. The experimental platform seen in a school setting.

It is important in SDT that the participant can demonstrate they understand the difference between the Target and noise, given enough time. To establish this, the game objective was re-introduced to the participants at the start of every session using a paper-based mockup to test the participants' understanding of the challenge and validate their response (discussed further in Empirical lessons learned from the pilot).

Table 5. Participant characteristics and CPT settings adjusted for participant capacity.

Participant alias	Age	P scales mean	Slide display time / Stimulus duration (s)	Blank slide display time / Interstimulus interval (s)
Will	18	6.93	1.8	1.1 ± 0.1
Jen	19	19.45*	1	1.1 ± 0.1
Mark	16.75	3.7	8	2.1 ± 0.1
Rick	19	6.76	1.8	1.1 ± 0.1

*Jen is enrolled in the National Curriculum.

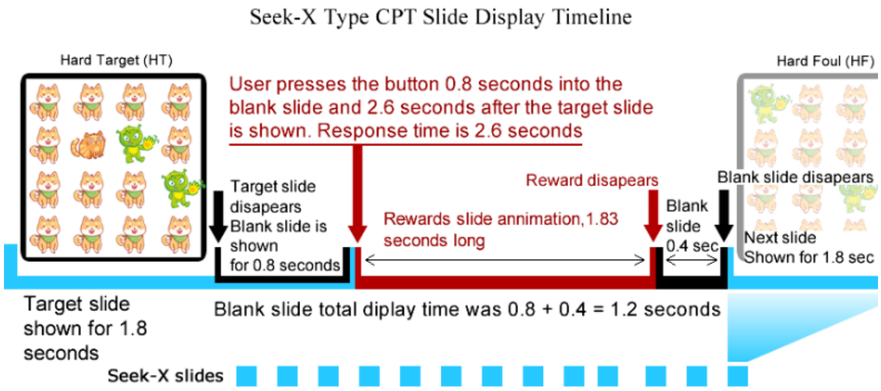


Figure 18. Seek-X type CPT slide timeline.

3.2 Experimental platform

The platform tracks participant performance in a repetitive game, which rewards them with exciting visual and audio feedback when they answer correctly, but ultimately fatigues them by being exhausting over a long period. The participant is required to pay attention to the game dynamic, which challenges them to pay selective and sustained attention to the elements on the screen and respond appropriately. This induces different states of affect, with lower levels of valence, as the game carries on and the learners' attention capacity naturally decreases. During this game, real-time Multimodal Multisensor data is collected within the experimental platform, which is used later to create a machine learning model of flow. The experimental platform was developed in MATLAB 20016a [221] to collect data from various consumer-grade sensor hardware. The experimental platform and the relative participant position are visualized in **Figure 19**.

The new type of CPT, Seek-X type, was designed for this study. The reasons why this new type was created are explained in section 3.2.1, Evolution of the CPT design. Each slide has a mixture of three images, comprising of the Target image, the Target Imitation and the Easy or Hard Contrast image, as seen in Figure 20. The Target Imitation bears a close resemblance to the Target image (similar colors, general shape), however, the Contrast images can be identified with less effort.

Note: In different sections of this work the images used to display the Target , Imitation and distractor slides for the Seek-X type CPT are interchanged. This is due to copyright reasons. Some images however are screenshots of videos taken of the actual set-up and for those reasons the original images are used. The CPT images represented in Figure 20**,** are thus replacement royalty-free images, which are used in most of this document. The orange cat, is replacing the cartoon character Wally (from the popular children books 'Where's Wally?' [222]), and the Imitation image shown by the orange dog is replacing the original red dog from the popular children's book 'Clifford the Big Red Dog' [222], the orange sandwich is replacing the red wool hat and the green alien cartoon is replacing Cookie Monster (from the children's program sesame street [223]). The original images from the actual study can be seen in **Figure 22**, **Figure 53** and **Figure 37**.

The experimental platform was tested under lab conditions to establish its robustness and data collection ability. Many of the shortcomings of the experimental procedure were discovered and improved on before the platform was tested in the pilot study at the school. The results indicate that the sensor data for multiple channels can be adequately captured in real-time and labeled by the system. This was described in section 4.1.5, Labeling and data fusion. The final experimental platform can be seen in operation in this YouTube video [224]. Data recordings were collected in a varied lab environment, and the noise and inattention conditions were not consistent between sessions.

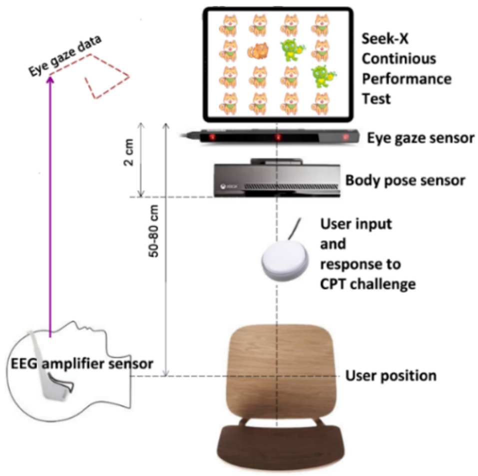


Figure 19. The Multimodal Multisensor experimental platform with the eye-gaze, body pose, EEG sensor and the CPT.

Target image	Target Imitation image	Hard Contrast image	Easy Contrast image

Figure 20. CPT images and the image types.

The ratio of the mixture of the main image to the filler image in all slide types is always 6 to 10 (plus one Target image on signal trials) or 7 to 10 (on noise trials). We found that for our test user group a grid of 4 x 4 introduced enough difficulty to allow for participant responses, without being so easy that the participant would not make any mistakes when fatigued.

The distribution of the Hard Target (HT) pattern among the other random patterns has an occurrence probability of 50%. The other CPT occurrences are standardized [193] as Hard Foul (HF), Easy Target (ET) and Easy Foul (EF). These patterns and their corresponding labels are seen in **Table 7**.

Table 6. The distribution of patterns in the Seek-X type CPT

Pattern	HT	HF	ET	EF
Distribution	50%	25%	12.5%	12.5%
CPT Label	Hard Target: Target image mixed in with imitations Targets with a few Hard Contrast images.	Hard Foul: Imitation Target images with some Hard Contrast images.	Easy Target: Target Image mixed in Easy Contrast images with some Imitation Targets.	Easy Foul: Imitation Targets images with some Easy Contrast images.

The participants were seated in a chair in front of a 20″ computer monitor with a width resolution of 1980 pixels by 1080 pixels, at a controlled distance of 50 cm to 80 cm from the screen. Each participant was asked to press the keyboard spacebar, or a big button if wheelchair-bound, whenever they saw the Target image on the screen, and not to press the button when they did not see the Target image on the screen. During this activity, participant eye-gaze, body pose, EEG measurements and button interaction data were continuously recorded.

The participant was then presented with 48 instances of images displayed in a controlled random sequence on the screen. This meant that no more than two of the same type of slides (shown in **Table 6**) could ever be displayed in sequence. Each image was displayed for a stimulus duration (slide display time) followed by a blank slide displayed for an interstimulus interval. The images were shown in a 4 x 4 grid, at a size of 280 x 192 pixels, 300 pixels apart horizontally and 100 pixels apart vertically. The images were shown in the center of the screen and had 300-pixel margins from the sides and 100-pixel margins from the top and bottom. This layout is shown in **Figure 21**.

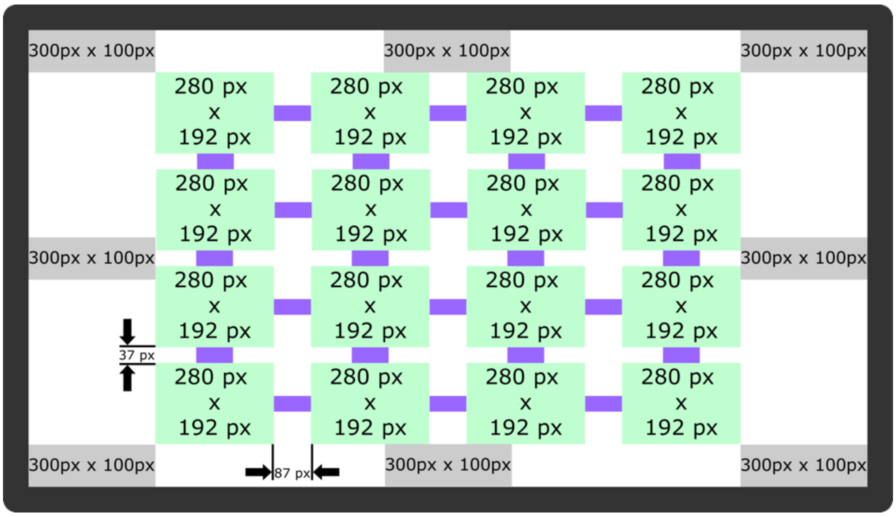


Figure 21. Screen layout for Seek-X type CPT, the grey blocks are margins from the sides of the screen, the light green rectangles are where the images are randomly allocated, and the purple rectangles represent the space between each image in the 4 x 4 grid. The screen resolution was 1980 by 1080 pixels.

Real-time eye-gaze position using Tobii EyeX [225], body pose data using Kinect v2, EEG data from the Muse headband [226] and interaction data from the USB button is recorded in MATLAB 2016a [221]. The Muse EEG headband streams 16-bit voltage data in microvolt (μV) units at 500 Hz, which is equal or comparable to medical-grade EEG specifications [227]. The Tobii EyeX eye-gaze tracking controller [228] uses near-infrared light to track the eye movements and gaze point of a participant [229]. It works in variable light conditions and allows for participant head movement while maintaining accuracy, which is crucial for our target user group. It has a frequency of 70 Hz and uses backlight assisted near-infrared (NIR 850 nm + red light (650 nm)) to achieve a 95% tracking population [230]. The Kinect 2 sensor [231] is a motion-sensing peripheral for body tracking. Using structured light and machine learning it can infer body position [231]. Kinect 2 is reported with an average depth accuracy of under 2 mm in the central viewing angle and increases to 2-4 mm in the range of up to 3.5 mm [232]. The furthest distance captured by Kinect 2 is 4.5 mm, where the average error typically increases beyond 4 mm. The experimental platform was designed to replicate the majority of the CPT variations reported in relevant studies [1], [15], [143], [144], [149], [160], [185], [193]. The features extracted from these sensor data streams are described under feature extraction. Methods of data and memory sharing were developed to synchronize the four independent Matlab instances (see section 4.6.1). This ensures that the four Matlab instances start and stop recording the five channels of information (eye-gaze, EEG, body joint data, interaction data and CPT outcomes) in synchrony. This helps streamline the experimental platform maintenance, making the participant experience less distracting.

3.2.1 Evolution of the CPT design

In the initial two weeks of the study, before data collection began on April 21st, 2016, a pilot study was completed. The platform suitability was tested as a tool for data collection from the PMLD/CP participants. Observations were made that altered the design of the experimental methodology. The alterations impact directly on the design of the CPT.

Figure 22. The platform setup running the Seek-X type CPT seen in a pilot study here used two PC's running MATLAB R2016a in parallel, sharing memory over a local LAN network.

It was determined that the CPT-AX type test, which requires users to recall (the previous slide value), deduce (if the previous slide with the current slide is the signal) and respond (by pressing the button) added too much skill requirement for our user base and the user would be left in a state of continuous learning and readjustment. This could have had a severe consequence of alienating the user early on in the study. The test was altered to become a non-rare type CPT. In this version of the CPT, the signal is shown more frequently than non-Target slides (see **Table 6**, 62.5% of the total slides are Target slides and 37.5% of the slides are non-Target slides). This type of CPT focuses on sustained attention and in called a non-rare type CPT, see **Equation 21**. The new type of CPT, the Seek-X type is fully described in section 3.2, Experimental platform. We now explore the original CPT design that was used in the pilot study sessions.

A-X type CPT

In the initial design of the CPT, the participant was instructed to press the button in front of them only if they see the A then X image appear on the screen (the CPT A-X test). The A image is represented by the green tree image and the X image is represented by the racing car. In this standard version of the test, the participant is asked to press the space bar whenever they see the X (the probe), but only if the X was preceded by an A (the cue). The distribution of the A-X pattern among the other random patterns has an occurrence probability of 70%. The other CPT occurrences are standardized [193] as A-Y (image A followed by image Y), B-X and B-Y. These patterns and their corresponding labels are seen in **Table 7**. All possible signal variations and their significance to the attentional state tested in the CPT are demonstrated and discussed in **Table 7**. The platform in its developmental stage can be seen in **Figure 23** running the A-X Type CPT.

Table 7. The distribution of patterns in the A-X test

Pattern	A-X	A-Y	B-X	B-Y
Distribution	70%	10%	10%	10%
CPT Label	True Positive (Correct)	Foul (Impulsivity, selective attention loss)	Foul (Inattentiveness, sustained attention loss)	True Negative (Incorrect)

Figure 23. CPT type A-X is shown in the development stage of the experimental platform.

All permutations of a standard A-X type CPT are visualized in **Figure 24**. The predetermined correct response to the signal Tree-Car (A-X) is a button press; any other patterns should not be responded to.

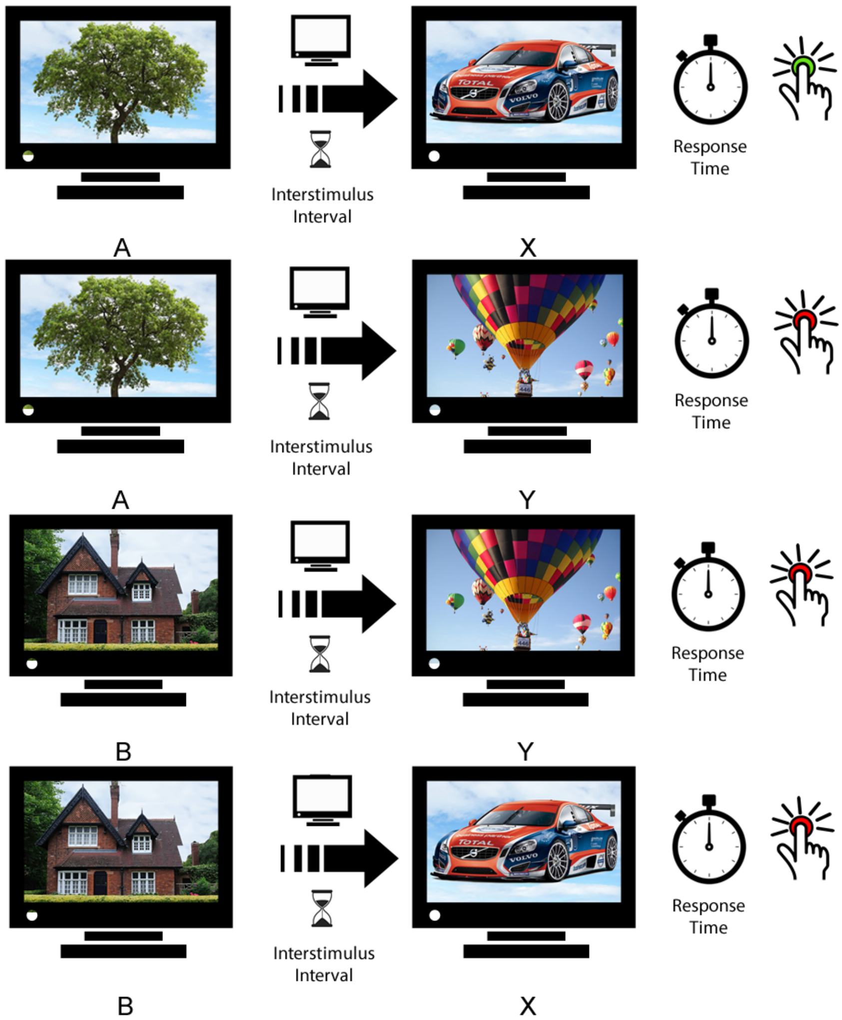


Figure 24. All outcomes of an A-X type CPT.

When the user responds to the A-X signal, it is called a signal detection. If the user responds to any other pattern (e.g., A-Y), it is called a Commission error, which relates to impulse (inhibition) control. If the user responds to B-X, this demonstrates a lack of sustained attention (vigilance) - they saw the car but forgot or did not pay attention to the previous signal (slide). This is also called a Commission error. When the user responds to B-Y it is called an omission error. In this case, there was no cue (tree) or probe (car).

3.3 Ethics

Ethical approval was granted by the Nottingham Trent University School of Science and Technology ethical committee in the month of September 2015. The information pack (both complete and 'easy read' versions) were submitted with the 'ethical issues' document (all attached in Appendices). Information leaflets (attached in Appendices) were sent to a special needs school, where they were distributed by the administration staff to the parents.

Two user group meetings were held with the school alumni, teachers and headteacher in the month of March 2016. In these user group meetings, the motivation, method, and aims of the research study were explained. A demonstration took place in the second user group meeting where the platform was tested and initial feedback on the design was collected from the teachers and participants. The initial system can be seen in this video [32]. We realized the importance of communicating 'simplicity, clarity and trust' through the text of the information packs from the feedback received from the initial user group meetings. Initial feedback on the experimental design and methodology was requested from the user group which consisted of PMLD school leadership and alumni (mature PMLD students from the school).

The length of the information pack, initially 2 pages long, was changed after advice that it should be summarized to a page or less. The experimental description was simplified to cover the main procedural step-by-step process of the experimental methodology, and scientific evaluation of sustained and selective attention was removed from the document. Scientific terminology like 'CPT' was replaced with an 'easy read' description of what the Continuous Performance Test is. Pictures of the EEG headband were also added to the document to demonstrate the consumer-friendly aspects of the technology used to collect the data. Emphasis on the user's well-being was reinforced, and their ability to leave the experiment at any point and retain their anonymity was made clear.

The updated information pack (available in Appendices) was distributed to the guardians of eight students. The students were selected for their appropriateness for the study by the headmaster of the school. Out of those eight, we received permission from four guardians for their child to participate in the research. Three boys and one girl, aged 16, 18, 19 and 19 participated in the data collection study. Two of the four participants are prone to seizures. During the 13-week study, we found that they did not react negatively to the CPT, despite it being quite an intensive activity.

Communication with the teachers was maintained throughout the study. E-mail reminders were sent to the teachers, two days before the data collection day to confirm the student participation and also to coordinate the time for the data collection session. Often due to illness, absence or other teacher-student commitments, sessions would be postponed. Email updates of the student experience and initial data analysis were also presented to the teachers as the study progressed.

At the beginning of the session when the participant was taken away from their classroom and walked towards the room where we were working, our priorities were to ensure that they understand what activity they were taking part in and to relieve any anxiety they felt. To put them at ease, we talked about their week, their life, their weekend, any movies they may have seen etc. The participant was always walked to and from our workroom, to ensure they returned to the activity they had been taken away from so that with consistency and security were maintained. The participant was not taken away from fun activities, like free time outside or from group participation activities. It was important to reinforce the study was a fun activity, not one that causing the participant to miss out on other activities that they may enjoy.

Knowing that the user is not the sole participant in the study (and so not being singled out or punished by being taken out of the class from their peers) was important so they would not have any negative associations with our research. We ensured that the participants knew of their friends who were also participating in the study. When possible, participants were offered the chance to observe each other take part in the experiment, as this helped demonstrate the experimental process to other participants and also reduced the feeling that participation made them stand out from other students in a negative way.

We worked with the teachers to make the participants feel like the activity was extracurricular and they were not being singled out as a punishment but were selected because they're unique and are offering us incredible value through this study. This was reinforced frequently by reminding the participants how well they were doing in the study before and after the study started and concluded. Any discussion that reflected negatively on student performance or any performance comparisons between participants was always avoided. "You're doing the best", or "come on, your friends can do it" is considered counterproductive. The participant was instead reinforced positively with their own progress throughout the study.

The participants were always asked if they wanted to continue with a new session, especially as the CPT can be strenuous and is designed to be tiring. With the consideration that two of the four participants are prone to seizures, it was important to re-establish student health and consent in participation before moving on to a new session. In one case the student asked for the study to not continue to the second recording on the same day. In another case, the recording was cut short due to the CP student signaling that they needed to go to the restroom.

3.4 Participants

Participation for the study was arranged through Oakfield School, Nottingham [233]. Oakfield is a school and sports college for people with a learning disability. The definition of learning disability according to the National Joint Committee for Learning Disabilities [234], [235] is as follows:

> *"Disabilities are a heterogeneous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, writing, reasoning or mathematical abilities…"*

The student population in Oakfield School is a *heterogeneous* one; studies with participants with learning difficulties are challenging [1]; a selection process is required to make sure the participants are in the same performance scales (P scales) [236], [237]. A meeting was initially held with the head of the Oakfield School about the experimental methodology and its desired research outcome. The feedback was incorporated into an 'easy read' version of the participant information. This was then circulated to eight students which Oakfield School recommended to be a good fit for this study based on P scales ranges and their ability to perform in the CPT.

Participants were selected based on their performance in scales, which represent a set of descriptions used to record and assess the progress of children who have special educational needs (P-scales) [236], [237] (see Table 5). Permission for the study was given by Nottingham Trent University's ethics committee. A further request was made to gather more participants in the study by sending the participant information pack to any remaining students and their guardians who may be unaware of this study.

Three boys and one girl aged 16, 18, 19 and 19 with PMLD and CP enrolled to participate in the research study. They are given pseudonyms; Will, Jen, Mark, and Rick. In a further follow-up meeting, details of the four participants were requested. The descriptions of the participants are the focus of the remainder of this section.

A breakdown of each participant's Age, P scales, specific communication strengths or constraints has been recorded at the start of the data collection process (see **Table 5**). The 4 nonhomogeneous participants had different ranges of capability. This ranged from extreme mobility restriction to some moderate learning difficulty. A nonhomogeneous participation base is an accurate representation of the population and is expected when conducting studies with PMLD.

Will is 18 years old, has a diagnosis of global development delay (GDD) and learning disability. These impact on his speech, language and social interaction with others. This means his ability to concentrate on a single activity for an extended period is limited, which in turn limits his sustained attention. His body mobility is not restricted but is slightly imprecise. His speech can be difficult to understand, and he is limited in the selection of words. His capability in conducting particular tasks in quick succession is good; however, he struggles to maintain sustained attention. Will's P Scale graph is shown in **Figure 25**.

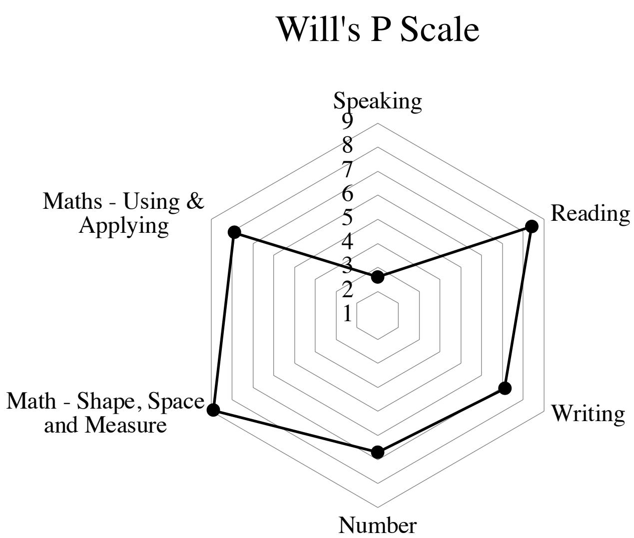


Figure 25. Will's P Scale demonstrates that his speaking ability is considerably lower than his other abilities.

Jen is 19 and has a rare form of epilepsy. She is one of the more capable learners at the school. She is very cooperative and shows an interest in being involved in the study. She also talks about music and theater and has interests in fashion and celebrities. Her performance is highlighted by concentration and commitment to the activity. Jen has completed her P scales and is not enrolled in the national curricular. Jen's National Curricular Level graph is shown in **Figure 26**.

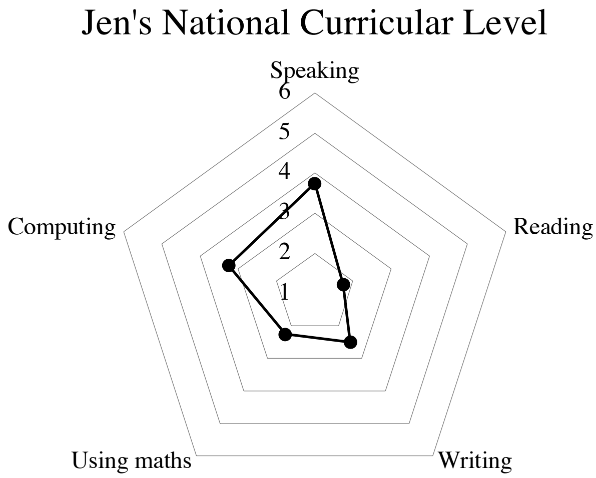


Figure 26. Jen's abilities surpass the P scales grading, and she is on a mainstream rating scale.

Rick is 19 and has a global delay, a rare form of epilepsy and a severe learning difficulty. Rick has problems processing information and communicating. His attention is usually committed to a single concept (an activity, a memory, a sound). He is incredibly reliant on routine, and he will try to avoid any disruptions to it. He enjoys loud motor sounds, power tools, and garden work. He often reflects on activities he has done in the past or will do in the future with single words or short phrases. His mobility is not constrained but is delayed and processing time needs to be allowed for any responses. Physical objects and sounds help him associate with new concepts. Rick's P Scale graph is shown in **Figure 27**.

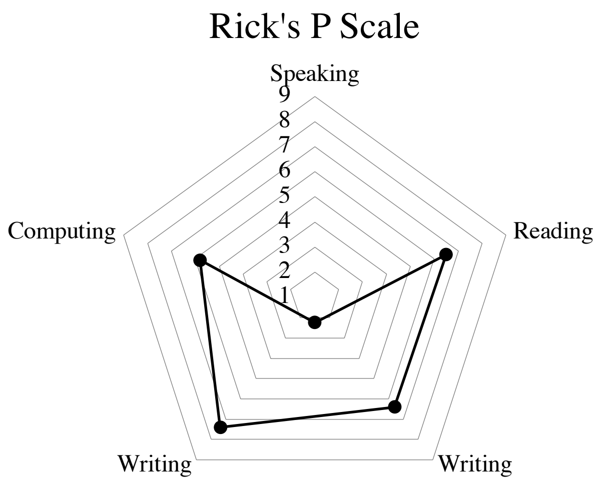


Figure 27. Rick's P scales demonstrate his strengths in written languages and computer-based activities.

Mark is 16 years old and has myotonic dystrophy; this makes his muscles very weak. Myotonic dystrophy is a progressive and life-limiting condition. Mark uses a wheelchair and is at risk of chest infections and sudden heart failure. He uses a specialized CP wheelchair for body support and transportation. The wheelchair supports his body frame and keeps him upright and secure with a safety belt. His head is rested against his right ear on a padded headrest. His mobility disability is extreme; however, he has some imprecise movement in his neck and arms. At school, he uses both eye-gaze technology and switches to interact with computer interfaces. Mark uses his voice to communicate; he likes sharing his sense of humor, he laughs when things go wrong, and makes the sound 'uh-oh' to signal mistakes. He enjoys making choices and can become frustrated when he is not offered choices. Mark likes interacting with computers but shows sensitivity to anything resting on his forehead like the EEG headband. Because of his CP, he required a member of staff to be present during the study. Mark shows a definite progression with communication and is now very accepting of and participating in a wider variety of activities, events, and opportunities in school. He has responded well to unforeseen changes in his daily routine even with little notice. Mark's P Scale graph is shown in **Figure 28**.

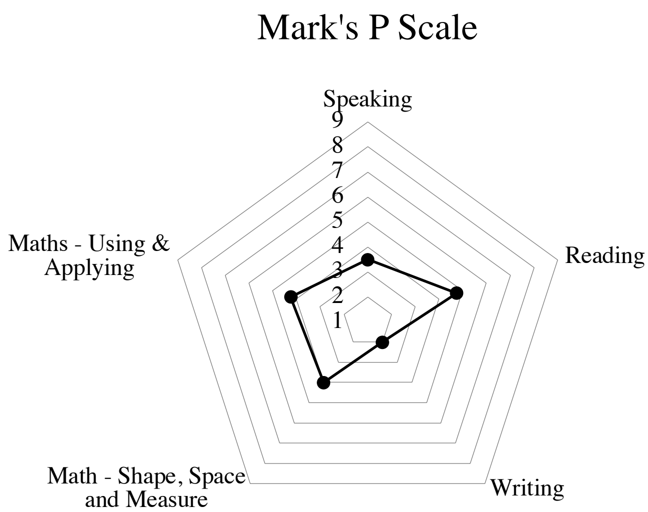


Figure 28. Mark's P scales indicate that he is our most profoundly disabled participant.

Chapter 4. RESULTS AND DISCUSSION

This chapter starts with describing each sensor and the data collected from it. We present our data processing and feature extraction methods. An overview of the features extracted from each sensor is discussed. We then describe the statistics of the data collected and the methods used to develop the machine learning results. Algorithms used in the machine learning are presented and machine learning indexes for comparison are defined. The classification results are then presented in a table for subsets of the data and are compared using the machine learning indexes. Cross-correlation results between CPT outcomes and participant characteristics are presented. Empirical findings from the eye-gaze, body pose, and EEG data are presented.

We show (regardless of the classification method), that there is a relationship between affective state and the Multimodal Multisensor data features. In this study 2615 samples, over the length of 59 sessions, were collected and classified into two categories ('attention' and 'inattention') using nine features (7 low-level and 2 High-Level Compound Features). The individual features will be explained in section 4.1, Data Feature processing, and the High-Level Compound Features will be explained in section 4.1.5, Labeling and data fusion.

, however, a list of the features can be seen in **Table 8**. In total there were 2051 'attention' samples and 564 'inattention' samples.

Table 8. Summary of data points and sources.

Feature	Attribution
Eye scanning	Eye-gaze
Eye dwelling	Eye-gaze
Eyes off screen	Eye-gaze
EEG Alertness	EGG
Body fidgeting	Body pose
Single fast press	Interaction data
Max press count	Interaction Data
HLA	High-Level Compound feature
HLI	High-Level Compound feature

This presents a problem, where the algorithm has a bias towards the majority class while neglecting the minority class. This often coincides with achieving high accuracy results, so it is often overlooked. As our data set is unbalanced, we look into addressing this in section 4.2, Unbalanced datasets.

The data was assessed both as a group and person-specific, the results of both are compared. The aim of classification is to determine the affective state by predicting the CPT outcome. With two classes, the random classifier classification AUC to beat is 50%. As the data set is unbalanced, the dataset was balanced before machine learning took place to improve the classification of the minority class, this method normalizes sample size across all machine methods. The overall approach used to evaluate the fit of the different architectures was leave-one-out (L-O-O) cross-validation. Scoring metrics insensitive to unbalanced datasets were used to competitively compare the performance of the different machine learning approaches. The evaluation parameters used for determining the comparative performance of the machine learning architectures were Area Under the ROC Curve (AUC), Negative Log-likelihood and kappa. The software used to create this architecture is Python 3.7 and two high-performance computers, which ran in parallel over several weeks. The two PCs were both equipped with Intel i7-7700HQ 2.80 GHz CPUs, and 16 GB of DDR4 RAM. The CPU was benchmarked at 82 Gigaflops, with 15 GB/s memory transfer rate and 1 GB/s SSD disk transfer rate.

4.1 Data Feature processing

In this section, an overview of the features extracted from each sensor is discussed. Critical to the process of inferring the affect state of a student with PMLD/CP from labeled Multimodal Multisensor data is the process of data fusion. The process of detecting the user affective state through multiple input sources and their combination, known as "Multimodal Multisensor input fusion" [38], [102], [238] presents a number of challenges. Firstly, the type of data to be managed by a Multimodal Multisensor system may originate from a variety of different sources; these sources may have different sampling frequencies and different code bases. Temporal time syncing of this data and sampling it to a common denominator of the frequencies is a requirement. Making a platform compatible with all the different codebases, while running effectively and reliably is a challenge. Secondly, due to the computing demands of high-frequency sensor data streams, two computers were used in tandem to facilitate the data collected. These computers were connected to a local network, in which two simultaneous Matlab 2016a instances shared memory and coordinated the Multimodal Multisensor data collection and the CPT experimental process. This photo of the platform is shown in **Figure 17**.

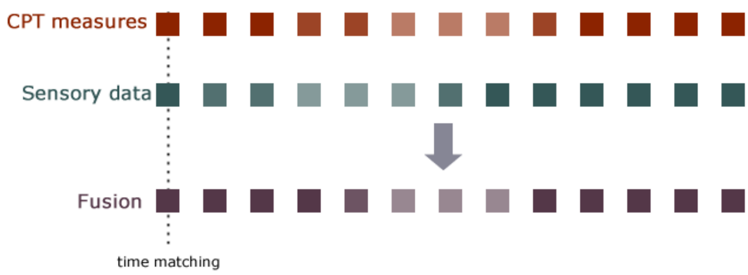


Figure 29. CPT provides objective labeling for the Multimodal Multisensor data

The purpose of incorporating an objective measure of attention, namely the CPT, in the experimental platform is to provide a reference label in the machine learning stage. Without these labels, there would be no objective measure or automated way of performing a supervised learning method on the data. An overview of how the data streams were captured and labeled against CPT attention indicators is shown in **Figure 29**.

4.1.1 EEG

Brain-Computer Interfaces (BCIs) represent a novel mode of communication that has been used in emotional classification [239], and cognitive aware applications [240]. EEG frequency has been used as a feature to determine active brain state [241]–[243]. BCIs are also considered unique in Augmentative and Alternative Communication (AAC) as they do not require physical movement from a user. This makes BCIs a suitable AAC method for people with Severe Speech and Physical Impairments (SSPI) [244], [245], or CP [246]–[249] who do not have access to conventional means of communication including speech and typing [248]. EEG has also been used in Multimodal Multisensor applications in combination with eye-gaze and facial expressions to detect affect [102], [113], [250]–[254], however, these all use expert, peer or self-annotation as labels for the data.

The quality of a BCI — to offer a direct mode of information from the brain — makes it especially ideal as an element in potential real-time affective user state detection [255], computer interaction for rehabilitation [256] and in brain multimedia interaction [257]. A BCI can also be a complementary source of information towards Multimodal Multisensor interaction systems as well, used in conjunction with other modalities such as gesture, facial expressions, gaze and body posture [258]–[260].

Accurate classification of EEG signals into commands or user affect state is the primary challenge of BCI-AAC practice and has sparked the interest of the research community [239], [244], [261]–[264]. EEG classification has been attempted with different techniques and data specifications. Some different EEG features used are Visually Evoked Potentials (VEP), Frequency Averages, P300's and timeline classification [265]–[267]. Other EEG studies use Event-Related Potentials (ERP) as a mechanism to classify EEG [268]–[272]. These are distinct rises in the wave, almost mechanical in their nature. In contrast to the rest of the wave, their event is directly related to an "inducer" from the outside world. This is when an outside stimulus (visible blinking lights, sound or tactile stimuli) correspond to noticeable potential changes in the EEG signal, which are repeatable, reliable and do not require intention on the user's part [258]. This allows the user a control mechanism, which is focused on one of their degrees of freedom. The user can then choose to focus their attention on the stimulant, and the system recognizes the EEG evoked potential as a response [246]. Using ERP however is not a suitable method in this work, as the objective is to find a feature that corresponds to their activity (focusing on the SDT challenge) and not to find features in the EEG that are intentionally induced by the participant (like looking at a blinking light to induce an ERP response). Due to the different approaches to EEG data acquisition, method and definitions of task achievement accuracy, it is difficult to achieve a direct comparison between the methods used in these studies.

A more practical method is to use EEG time-series samples from a baseline activity to train a classifier. The classifier can use statistical and machine learning pattern recognition methods to classify the sample with an output pair of confidence and accuracy percentages [262], [273],

[274]. In this method, a mental task is agreed upon, and the receptive controlled practice of this mental task is recorded. The classifier trains itself on these samples. Some mental tasks include imagining motor planning to stimulate the motor cortex, mental math calculations to stimulate brain activity, relaxation and imagining emotional events to simulate emotional response [257], [258], [262], [275], [276]. This method is more suitable for our experiment, as the mental task will be naturally evoked by the CPT itself when the participant maintains attention in order to perform in the SDT trial.

While the use of EEG signal is becoming more and more attractive in developing practical use natural Brain-Computer Interface (BCI), accuracy still remains a challenge [245]. EEG is challenging due to the complexity of the signal as a data source [247], [277], [278]. While some studies require invasive procedures using implanted electrodes or electrode grids to access the brain tissue [277], [279]–[281], several other studies have presented different methods of tackling the classification challenge using EEG headset amplifiers [282]–[288]. Although the implanted electrodes allow a more direct recording of the neurophysiological activity of the cortex and, therefore, have a better signal-to-noise ratio, they are currently reserved for patient populations [289]. Despite EEG amplifiers being far less invasive and more suitable for this study, using them creates an additional challenge to preparing the EEG data for classification. To do this we use a combination of signal preprocessing methods.

A combined signal preprocessing approach based on methods discussed in previous studies (grounding and Kalman filtering) is taken [290]–[293]. EEG Kalman filtering has been shown to be useful in removing electromyography (EMG) induced artifacts [262], [290]–[295]. A robust Adaptive Autoregressive (AAR) model with an order of six detailed in [293] was used. The AAR model estimate of the EEG Kalman filter was utilized to reduce the impact of Electromyography (EMG) spikes from body movement, eye blinks and other facial muscle movements. These EMG spikes are isolated in a few samples, which makes the data ideal for AAR Kalman filtering. In **Figure 30**, we see that it has removed the EMG artifact that can be seen between samples points A and B, enhanced the EEG spikes, and revealed an EEG peak between C and D. We demonstrate how Kalman filtering smooths out real EEG data from this study and reduces spikes in **Figure 31**. The original signal is shown in the dark blue, with the filtered signal shown in the lighter pink. Removing the EMG spikes from eye blinks, facial and neck muscles, improves the quality of the EEG data by enhancing the EEG features and reducing noise and artifacts caused by the EMG interference.

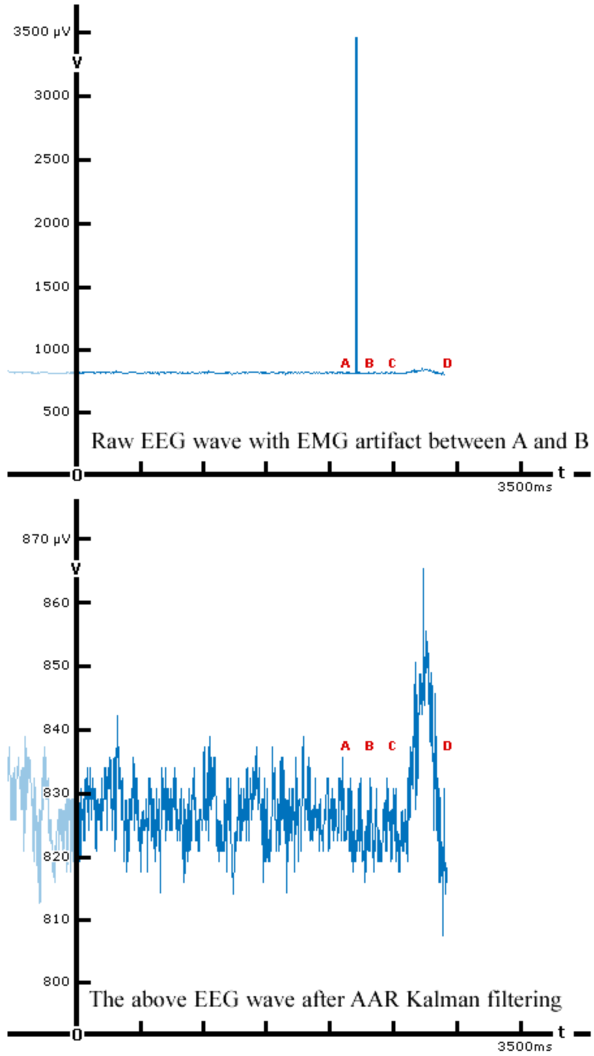


Figure 30. AAR Kalman filtering reduces EMG noise and enhances EEG spikes.

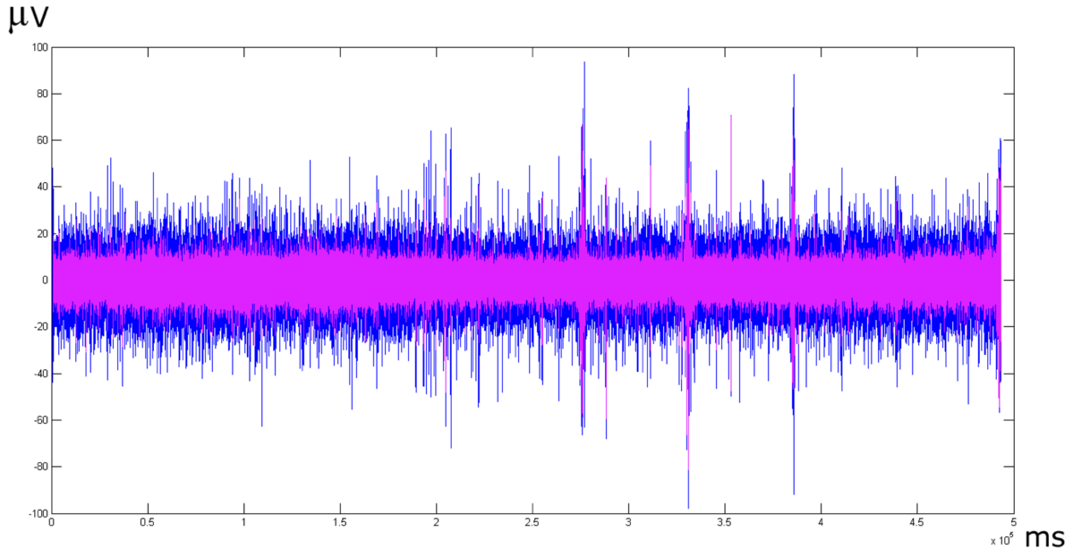


Figure 31. Grounded, AAR Kalman filtering example, from EEG data – overlay in pink is the filtered outcome.

The low-cost commercially available EEG amplifier, the Muse EEG Headband controller seen in **Figure 32** is used. The Muse is a flexible, adjustable, lightweight headband with 7 built-in sensors capable of reading 4 channels of data. The Muse has five dry sensors on the forehead and two 'SmartSense' conductive rubber ear sensors [296]. The standardized 10-20 electrode locations [297] are visible in **Figure 33**. The headband is configured to stream data in its research mode [298] (pre-set ID: 14), in this mode, the Muse streams 16-bit data at 500Hz, which is equal or comparable to medical-grade EEG specifications [227]. Five channels of EEG data are recorded, TP9, AF7, FPz, AF8 and TP10 [297] at a frequency of 500 Hz over a Bluetooth connection.

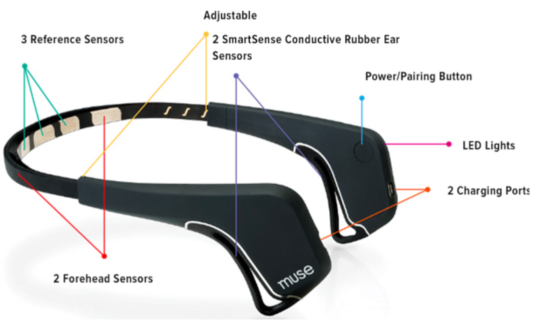


Figure 32. Muse Headband dry sensor locations.

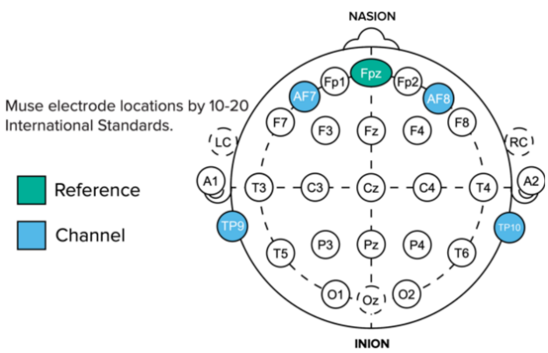


Figure 33. Muse Headband standardized 10-20 electrode locations.

By using an AAR Kalman filter on the data, we estimate the EEG wave during the EMG incident artifacts using surrounding neighboring EEG samples and correct those affected samples. This is done by evaluating a moving set of samples and checking for EMG contamination. The contamination is then removed by estimating a normal rate of progression for the signal to reach from point A to point B using a sliding window for the length of the recording.

EEG band power has been used as a feature to determine the active brain state [241]–[243]. Studies show [286], [299]–[301] that the EEG β rhythm (14–30 Hz) is activated when the brain is in a state of arousal. In other EEG studies, mental fatigue related features are associated with decreased α band (8-13 Hz) power at one or more parietal locations (e.g., P7 and P8). Ning-Han Liu *et al.* [282] connected these two factors in their study and showed that alertness can be measured by the signal power of α divided by the signal power of β. Timothy McMahan *et al.* [302] also demonstrated that the ratio is related to arousal.

Using the signal power of α divided by the signal power of β as the EEG feature, the EEG recordings are labeled with the CPT outcomes. A Butterworth bandpass filter was employed to extract the frequency response of the α and β bands from the EEG signal as demonstrated in [303]. Discrete Fourier Transform (DFT) was used to calculate the Power Spectral Density (PSD) of the α and β time series.

DFT periodogram methods for estimating the spectrum power density are prone to variation [304]. Periodogram estimate variation is correlated to the square of the value of the spectrum itself. Welch's method reduces this variance by averaging independent periodogram estimates. Each Welch window covers 50% of the next, which results in the smoothed-out average of independent periodogram spectrum estimations. We use a Hamming window as it produces the least amount of overshoot $\delta_{\mathrm{Hamming}} < \delta_{\mathrm{Hann}} < \delta_{\mathrm{Bartlett}}$ [304] with the most accurate results for EEG data [303], [305].

A Hamming window of M = 100 samples was chosen with a 50% overlap, and since the EEG frequency is 500 Hz, this Hamming window is equivalent to 200 ms of data. To help illustrate, an average data interval length is 2.3 seconds long and would have $2300 \div 200 \times 2 = 23$ overlapping Hamming windows. Let $\{xd(n)\}$ be the sequence, $d = 1, 2, 3 \cdots L$ signal intervals and M the interval length. Welch's method to estimate the power spectrum discrete-time sequence is shown in **Equation 22**. Where U is the normalization factor (**Equation 23**) and the Hamming window calculation is shown in (**Equation 24**). Using the Welch method, the ratio of the α band power f_α to the β band power f_β is simplified as (**Equation 25**).

Welch Method:

$$\hat{p}d(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} xd(n)w(n)e^{-j2\pi f}\right|^2$$ **Equation 22**

U is the normalization factor for the Welch Method:

$$U = \frac{1}{M}\sum_{n=0}^{M-1}|w(n)|$$ **Equation 23**

Hamming window:

$$w[n] = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{M}\right), & 0 \le n \le M, \\ 0 & otherwise. \end{cases}$$ **Equation 24**

The EEG Alertness feature:

$$\text{Alertness} = \frac{\hat{p}d(f_\alpha)}{\hat{p}d(f_\beta)}$$ **Equation 25**

Using the EEG alertness feature, the EEG recordings are labeled.

Table 9. EEG Alertness feature calculated for a sample of data that is collected during a Hit with fast reaction time.

Timestamp	Slide number	Slide image code	Slide expected response code	Response time	EEG Alertness feature $\frac{\hat{p}d(f_\alpha)}{\hat{p}d(f_\beta)}$
1.123	99	'y'	'w'	[]	8.035391
2.891	100	'x'	't'	0.415672	18.943130
3.659	101	'b'	'w'	[]	0.723423

In **Table 9,** a sample of three readings from the EEG data from a 'Hit' trial is displayed and the temporal EEG Alertness feature values are calculated and presented in the last column of the table.

4.1.2 Eye-gaze

Eye-gaze tracking has been used in various applications including understanding user intent [211], [306], behavior [307], procedural errors in complex tasks [308], communication devices for the disabled [309], [310], determining mental workload [311], [312] and understand user attention [30], [211], [306], [307], [313]–[316]. Eyelid movement in combination with facial features has been used to theorize an affective platform in [16]. This work focuses on gaze data obtained using an eye-gaze controller. Although there are other ways to obtain gaze data, some of them fall into the category of "intrusive collection" [316]. In addition, when setting up an infrared eye-gaze tracker, a relatively small effort is required, making it suitable for a Multimodal Multisensor study with PMLD/CP participants.

An eye-tracker measures the point of gaze and movement of the eye from one point of gaze to another. This measure serves as a marker of attention, the sustained emphasis of cognitive processing power on targeted information while ignoring distracting information [317]. Monitoring eye movements thus shows changes in attention. One way to explain changes in attention is the moving projector theory in which attention can be understood as a projector that moves when the focus is on the intended targets [318], [319]. When the projector illuminates the information, more efficient processing of the information may take place. However, during spatial attention shifts, this spotlight is turned off while attention moves to the next assisted location [318], [319]. This change of attention takes place in three mental stages; (a) a subject disengages the attention from the current focus, (b) attention is transitioned to the new location, and (c) attention is finally engaged at the new location [316], [320].

Eye-gaze data includes Cartesian information regarding the eye-gaze location relative to the bottom left corner of the screen. We track gaze, which is both on and off-screen. The combination of off-screen gaze tracking and eye detection provides information on eye-gaze behaviors that relate to user attention.

Three features were extracted from the eye-gaze data; 'eye scanning', 'eye dwelling' and 'eyes off-screen'. These features are commonly used in eye-gaze technologies to understand attention, interest and engagement [108], [240], [308], [311], [314], [321]–[327]. We introduce formulas to calculate eye-gaze scanning and dwelling, which are based on the definitions provided in [324]. All eye-gaze features are impartial to the elements on the screen and are calculated equally for signal and noise trials.

Scanning represents the eye-gaze behavior of when the gaze tracks across more than one image element. The image elements are defined as any target, Imitation or Contrast image in the Seek-X type CPT image matrix (seen in **Figure 21**), one of the three images from **Figure 20**.

Supervised machine learning methods require instances of labeled information to determine relationships between data and predetermined states. Pattern recognition algorithms use data matrixes to determine relationships between a series of data and the target classification. With a similar method, a series of eye-gaze data can be translated into eye movement behavior patterns, which will substitute our raw data in the classification process.

The scanning feature is calculated in **Equation 26** and represents the sum of the inverse distance from the center of each element. Where r_{in} is that distance; from the eye-gaze location to the center of image i out of $I = 16$ total image elements, for sample n, out of N total discrete sensor samples. The relevant diagram is shown in **Figure 34**.

$$\text{Scanning} = \sum_{n=1}^{N} \sum_{i=1}^{I} \frac{1}{r_{in}} \qquad \textbf{Equation 26}$$

Dwelling represents the eye-gaze behavior of when the gaze stays relatively in the same position for a duration of time. This behavior is independently calculated from the location of the image elements on the screen. The dwelling feature is calculated in **Equation 27**, which is the sum of the inverse distance from each eye-gaze position to the next. Where n is the sample number out of N total discrete sensor samples, and Δd is the distance the eyes have moved since the previous sample, as demonstrated in **Figure 35**.

$$\text{Dwelling} = \sum_{n=1}^{N} \frac{1}{\Delta d_n} \qquad \textbf{Equation 27}$$

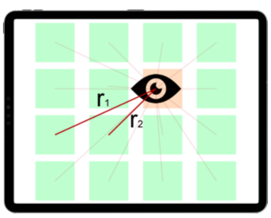


Figure 34. Scanning calculation with respect to the active elements on the screen

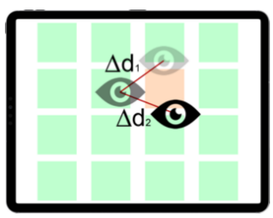


Figure 35. Dwelling calculation independent of active elements on the screen

The third feature extracted from the eye-gaze data is 'eyes off screen'. This continuous but binary feature determines if the participant is looking within the screen area, regardless of whether there was a slide or blank slide on the display. This feature is calculated as in **Equation 28**.

$$\text{Eyes off screen} = \begin{cases} 1 & eyes\ off\ screen \\ 0 & eyes\ on\ screen \end{cases} \qquad \textbf{Equation 28}$$

The Tobii EyeX eye-gaze tracking controller [228] displayed in Figure 36 is used to measure and track eye-gaze data for this experiment. Eye-gaze data is recorded at 70 Hz [228] over a USB 3.0 connection. This device uses near-infrared light to track the eye movements and gaze point of a student [229]. It works in variable light conditions and allows for student head movement while maintaining accuracy [328], which is crucial for our user target group. It tracks eye-gaze data regardless of facial features, ethnicity, eyewear or contact lenses [328], [329]. The Tobii EyeX sensor is an eye-tracking peripheral based on Tobii's IS3 hardware revision [329]. The Tobii EyeX uses backlight assisted near-infrared (NIR 850nm + red light (650nm)) to achieve a 95% tracking population [229], [230], [328], [330].

Figure 36. Tobii EyeX controller.

The eye-gaze data includes Cartesian information about the eye-gaze location for the bottom left corner of the screen. The screen dimensions were 1980 x 1080 pixels. The coordinates of the eye-gaze X values would typically be between [0-1980] and the Y values between [0-1080], see **Table 10**. The eye-gaze path (see **Figure 37**) would create scanning and dwelling patterns which are not only unique to the attention state but also unique to the participant. Participant specific eye-gaze patterns are described in section 4.6, Tools and empirical results.

Table 10. Sample eye-gaze data recording

Time	Slide Number	Slide Image Code	Gaze X-axes cord. (pixels)	Gaze Y-axes cord. (pixels)	Head distance	Left Eye Detected	Right Eye Detected
1.014	7	't'	811.252672	684.516768	726.225144	1	1
1.029	7	't'	831.771189	711.907167	727.572668	1	1
1.042	7	't'	726.168431	717.639933	737.710769	1	1
1.057	7	't'	789.94406	734.517148	773.65008	1	1

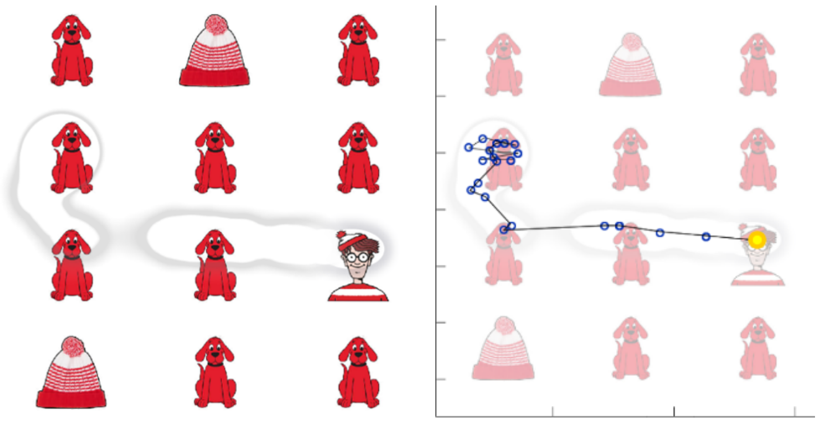


Figure 37. Real-time eye-gaze tracking over CPT slide (left) and a visual timeline of recorded eye-tracking data (right) is shown where each dot represents an eye-gaze data sample and the yellow dots represent the Target (Wally), was seen by the participant.

The eye-gaze controller, however, also makes good estimations of when the user is looking off-screen, so the X-Y gaze coordinates can exceed the screen resolution range. The controller also transmits information of when the left or right eye has not been detected by the sensor. The combination of gaze tracking outside the monitor and the eye detection provides information on when the user turns their head away from the screen.

4.1.3 Body pose

Humans interact with each other primarily through speech, but also through body gestures, to emphasize a specific part of speech and the display of emotions. Studies have shown that body posture and gesture can communicate affective modalities and also specific emotional categories [28], [29], [32], [54], [197], [331]–[334]. They have also been indicators of a firm or weak correlation of engagement during Human-Computer Interaction in gameplay [29], [32], [331], [333].

Tracking of the head, neck, mid-spine, right and left shoulders and right and left hands are recorded (see **Figure 38**). Lower joints are not included because of occlusion, as the table (part of the platform) prevents such recordings. In this study, the participant is positioned in front of a computer system and is challenged to press a button when they identify the target. This type of interaction setup restricts the range of body movements and gestures a participant can engage in. Due to this limitation, we are also restricted in the range of features we investigate in this mode of data. For this reason, we look at body fidgeting. Numerous studies [335]–[340] have investigated the importance of body fidgeting in detecting attention for students with PMLD. Fidgeting is an indicator of the onset of attention loss, boredom and engagement deterioration [333], [335], [341]–[343].

Figure 38. Body pose joints illustrated throughout the complete CPT task session in one graph.

We calculate rapid body movement from body pose to assess fidgeting levels. The equation to extract this feature is seen in **Equation 29**. Where Δd_j is the displacement vector of joint j out of N joints, and Δt is the time passing between the displacement samples.

$$\text{Body fidgeting} = \frac{1}{N}\sum_{j=1}^{N}\frac{|\Delta d_j|}{\Delta t} \qquad \textbf{Equation 29}$$

Body pose is acquired through the Kinect v2.0 SDK [231], which will provide joint tracking data at 30 Hz over USB 3.0. The Kinect 2 sensor [231], is a sensor platform (shown in **Figure 39**) for body tracking. Using structured light and machine learning it can infer body position [231].

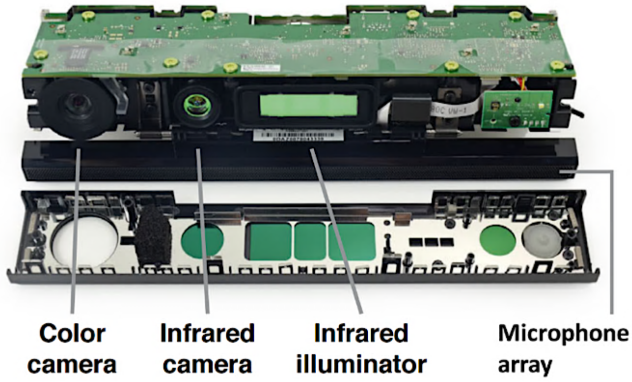


Figure 39. Kinect 2 sensor types and locations.

Table 11. Sample body pose recording segment; some joint data is not included to conserve space.

Slide number	Slide image code	Time (seconds)	Head X (meters)	Head Y (meters)	Head Z (meters)	Spine X (meters)	Spine Y (meters)	Spine Z (meters)
1	't'	1.033	-0.0048	0.3099	0.81217	0.05728	-0.0698	0.83084
1	't'	1.067	0.00119	0.3134	0.82115	0.01616	0.02168	0.83218
1	't'	1.100	0.0103	0.2616	0.77842	0.01625	0.02158	0.83184

Kinect 2 is reported with an average depth accuracy of under 2 mm in the central viewing angle and increases to 2-4 mm in the range of up to 3.5 m [232]. The furthest distance captured by Kinect 2 is 4.5 m where the average error typically increases beyond 4 mm. A sample of the body pose data recording is seen in **Table 11**.

4.1.4 Interaction data

Interaction data features were extracted from the participants' behavior activating a button press. The type of pressing, including quick presses or repetitive presses, was recorded as was other sensor data with a view to *behavior*, not just input, but as an independent sensor mode. This makes our approach unique as the input device is considered not only as an objective indicator of attention but also as a separate mode of interaction. We remain impartial to which slide is displayed and only consider the interaction behavior. How the button is pressed, specifically how fast the button is pressed, and how many times it is pressed is of interest. From button presses, we extract two features; single fast button presses and repetitive button presses. Single fast button presses are calculated using the formula described in **Equation 30**, with the caveat that they are only calculated if the participant presses the button once and only once during the response time duration. In other instances, the value for this feature is zero. The maximum press count is the second feature extracted from the button press data shown in **Equation 31**. This value is calculated for only the allowed response time interval and is zero when the button is not pressed. Fast response times in a CPT are a sign of participant attention [16] and repetitive button presses, are a sign of cumulative fatigue [344]. Therefore, by using these two features we can observe the changes in the participant's ability to maintain attention.

Single fast press = $\frac{1}{response\ time}$ **Equation 30**

Max press count = total press attempts **Equation 31**

4.1.5 Labeling and data fusion

The CPT provides an objective means of labeling the Multimodal Multisensor data [16]. The CPT has been used as an objective automated test to detect attention and label sensory data in prior studies [16], [173], [174]. The CPT outcome measures (Hits, False Alarms, Misses and Correct Omissions) are objective outcomes of the participant's attention to the CPT trials. Without these labels, there would be no objective measure or automated way of performing a supervised learning method on the data. An overview of how the data streams are collected and labeled against CPT outcome measures is shown in **Figure 40**. Each slide from the moment it is displayed until the moment of the participant's first button press, or until the moment of a new slide being shown (in case of no press), represents a sample of data (see **Figure 18**). These data frames were synchronized in time across all Multimodal Multisensor data streams. Overall, there were 2615 samples collected from the 59 data collection sessions across the four modes of data. The data from all four participants was collated and assessed per participant and in total as a group.

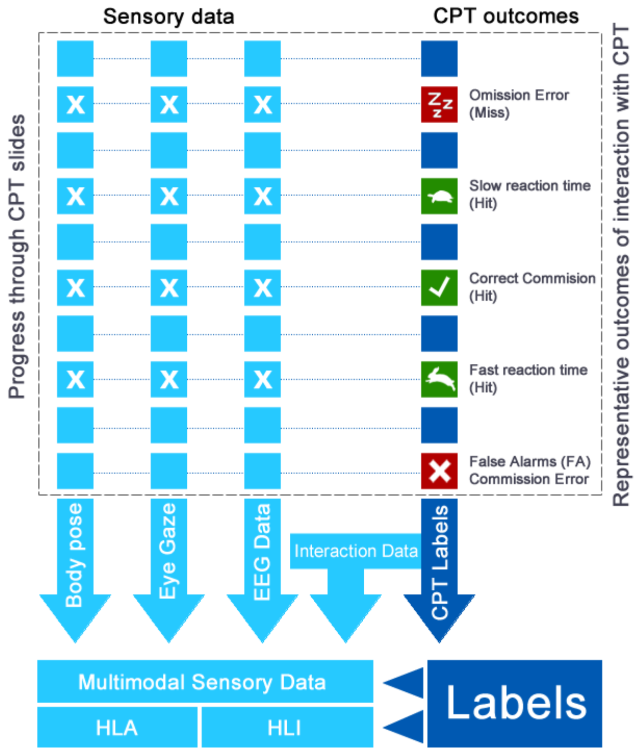


Figure 40. Multimodal fusion diagram shows the temporal connectivity between the samples and multi-level feature fusion.

To create a Multimodal Multisensor model of engagement using random forest, High-Level Compound Features (HLCF) were developed in [156] using the Mudra methodology [238]. HLCF creates a higher dimensionality in the feature space in a method described in the Mudra multimodal framework as 'high-level feature fusion' [238]. At the feature level, data has already been processed and feature functions have been extracted. The extracted features themselves can be fused, and not to the raw data. This is important when closely related modalities have independent representations, and the relationship between them is lost. For example, observing eye movement in the context of head movement can enhance a temporary understanding of user actions, which is the main goal of combining features at this stage. An exemplary combined frame can be represented with this semantic understanding of "the user is turning and looking above the right shoulder". This high-level understanding of data provides two possibilities, one for manually observing real-time data and checking for correctness, and secondly, it adds a contextual dimension that will be used at the decision-making level to build data interpretations.

We propose two HLCF in this work and later access their effectiveness. The first feature is a compound feature, which is a normalized mean of the features that traditionally serve as indicators of attention. The High-Level Attention feature is calculated as the mean of the normalized features of single fast presses, eye dwelling, eye scanning and EEG alertness, which are seen in **Equation 32**.

$$\text{HLA} = \frac{1}{4}(norm.Single\ fast\ press + norm.Dwelling + norm.Scanning + norm.Alertness) \qquad \textbf{Equation 32}$$

High-Level Inattention feature is calculated as the mean of normalized features of body fidgeting, eyes off-screen and press count as seen in **Equation 33**.

$$\text{HLI} = \frac{1}{3}(norm.Body\ fidgeting + norm.Eyes\ off\ screen + norm.Max\ press\ count) \qquad \textbf{Equation 33}$$

The process of detecting the user affective state through multiple input sources and their combination, known as "multimodal input fusion" [238], presents a number of challenges. Data managed by a Multimodal Multisensor system may originate from a variety of different sources; these sources may have different sampling frequencies. Temporal time syncing of this data and sampling it to a common denominator of the frequencies is a requirement.

The purpose of incorporating an objective measure of attention like the CPT in the experimental platform is to provide a reference label in the machine learning stage. Without these labels, there would be no objective measure or automated way of performing a supervised learning method on the data. An overview of how the data streams would be captured and labeled against CPT attention indicators is visualized in **Figure 40**.

The CPT specifications were firmly based on the standardized test described in a study that compared learning disabled student performance with mainstream students [1], [15], [150], [157], [160], [166], [170]. The number of Hits, False Alarms, Misses, Correct Omissions and response time for each button press is recorded for each participant (see **Table 3**). These indicators will be used to segment the Multimodal Multisensor data from the sensor inputs (eye-gaze, EEG, body pose and interaction data) to regions of high attention and low attention.

Swanson's application of the signal detection theory gives quantifiable information on the improvement or deterioration of attention in response to intervention, (CPT, [15], [140]). CPT is a reliable measure of attention, and there are significant correlations between its outcome measures (errors and response time) and teacher based ratings of inattention, hyperactivity and compulsivity in school children ([147]–[149]).

The CPT outcome measures will allow us to segment and label Multimodal Multisensor data from each student playing games (Eye-gaze, EEG, body pose and interaction data). This provides the labels by which we can supervise the learning method for the data. This is fundamental to our approach – where the CPT automates the labeling that allows us to perform a supervised labeling method on the data. The Multimodal Multisensor data relationship to participant attention and characteristics will be explored next in Chapter 4, Results and discussion.

4.2 Unbalanced datasets

Unbalanced datasets are prevalent in a multitude of fields and sectors, and especially machine learning science. An unbalanced data set is any data set that has an uneven distribution of samples. For example, if we had a data set of fruit, which included 5 oranges and 95 apples, the data set would be unbalanced. A common error is that if the unbalanced data set is not managed, a classification machine, looking at the aforementioned data set of fruit may report that all the fruit samples are 'apples'. By doing this it would achieve a 95% accuracy, however, this machine would never be able to detect a single orange (the minority class), and thus, does not generalize well. This is called the 'Accuracy paradox'[345], [346]. A better classification machine would be able to detect both apples and oranges, with accuracies that perform better than a stratified random and constant classifier (described in section 4.3.5, Stratified random classifier and Constant classifier).

Models trained on unbalanced datasets often have poor results when they have to generalize (predict a class or classify unseen observations). Regardless of the algorithm, some models will be more susceptible to unbalanced data than others. Ultimately, this results in a poor model. This is because of two reasons: Firstly, the algorithm receives significantly more examples from one class, prompting it to be biased towards that particular class. It does not learn what makes the other class "different" and fails to extract the underlying patterns that allow us to distinguish classes. Secondly, the algorithm learns that a given class is more common, making it "natural" for there to be a greater tendency towards it. The algorithm is then prone to overfitting the majority class. Just by predicting the majority class, models would score high on their loss-functions. In either of these instances, the accuracy paradox appears.

Both the inability to predict rare events, the minority class, and the misleading accuracy detracts from the predictive models built on unbalanced data sets. We will now discuss the accuracy paradox and the main techniques and methods available when dealing with this type of data.

4.2.1 Accuracy paradox

The accuracy paradox occurs when predictive models with a given level of accuracy have greater predictive power than models with higher accuracy. Despite optimizing classification error rate, high accuracy models may fail to capture crucial information transfer in the classification task. This is because 'accuracy' as a machine learning performance measure is not the best measure to assess machine learning algorithms, especially in the case of unbalanced datasets [345]. However, there are other machine learning performance measures, that are not sensitive to unbalanced data sets that should be used. Some of these performance measures are the AUC, log-likelihood, kappa, F1-score and True Positives (TP) and True Negatives (TN) from a confusion matrix. We will next discuss machine learning performance measures that are insensitive to unbalanced data sets.

4.3 Machine learning performance measures

We use three indexes to rank our machine learning methods. AUC, log-likelihood and kappa. All three are indifferent to unbalanced data sets. In the following section, we will define these and other performance measures.

4.3.1 AUC

Area Under the Receiver Operating Characteristic (AUROC) or shortened to AUC is a machine learning performance measure of the ability to avoid false classification. AUC combines TP and TN in a single measure [347]. The AUC or concordance statistic 'c' is the most commonly used measure for the diagnostic accuracy of quantitative tests [348]. It is a discrimination measure, which tells us how well we can classify trials in two groups: those with and those without the outcome of interest. Since the measure is based on ranks, it is not sensitive to systematic errors in the calibration of the quantitative tests.

Area under the ROC curve is a performance measurement for classification problems at various thresholds settings. ROC is a probability curve, and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. By analogy, the higher the AUC, the better the model is at predicting apples as 'apples' and oranges as 'oranges'. Likewise, the higher the AUC, the better the model is at distinguishing between 'attention' and 'inattention'. AUC is not sensitive to data distribution.

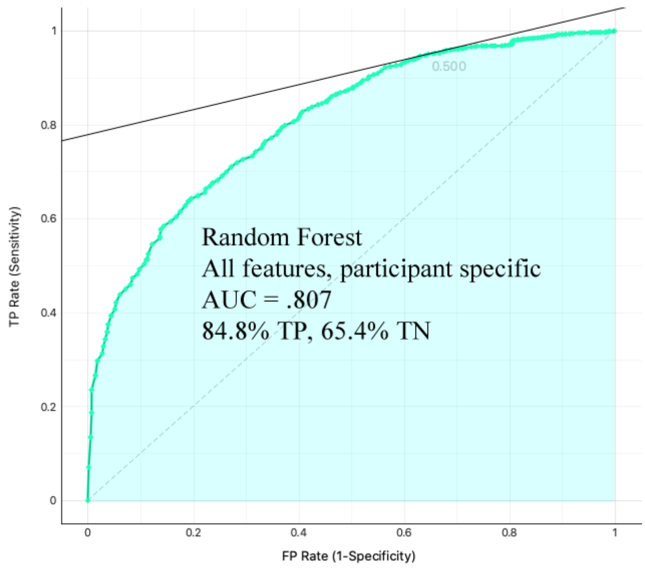


Figure 41. The ROC curve for our best model, with AUC shaded in light blue. The area under the ROC curve equals .807 in true classification units. A perfect classification would achieve 1. This model achieved 84.8% correct of the majority class ('attention') and 65.4% correct of the minority class ('inattention').

A classifier with no better accuracy than chance has an AUC of .50, meaning that there is a 50% chance in a two-outcome test to get either outcome. A test with perfect accuracy has an AUC of 1. AUC is equivalent to the Mann-Whitney U test statistic [349], [350]. The Mann-Whitney U test statistic (or Wilcoxon or Kruskall-Wallis test statistic) is equivalent to the AUC [351]. The ROC curve for the best performing classifier in this study is shown in **Figure 41**.

4.3.2 Cohen's kappa

Cohen's kappa (often simply called kappa) is a score for inter-rater agreement. It's a commonly used metric for evaluating the performance of machine learning algorithms and human annotators. When two binary variables are attempts by two individuals to measure the same thing, kappa can be used as a measure of agreement between the two individuals. kappa measures the percentage of data values in the main diagonal of the table and then adjusts these values for the amount of agreement that could be expected due to chance alone [352]. Because of this, it is often used for comparing or competing machine learning models against each other [353]. The benefit of using kappa, particularly in an unbalanced data set, is that the kappa statistic is describing how well the classifier performs above that baseline level of performance. kappa ranges from -1 to 1, with 0 indicating no agreement between the raters, 1 indicating a perfect agreement, and negative numbers indicating systematic disagreement.

Cohen's kappa is a measure of the agreement between two judges who determine which category a finite number of items belong to, whereby agreement due to chance is factored out. The two judges either agree in their rating, or they disagree, or there is no degree of disagreement. The formula to calculate kappa is shown in **Equation 34**. Where $P(a)$ is the classification accuracy and $P(e)$ is the probability classification success due to chance.

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)}$$ **Equation 34**

How to interpret Cohen's kappa

kappa is always less than or equal to 1. A value of 1 implies perfect agreement and a value of less than 1 implies less than perfect agreement. In rare situations, kappa can be negative. This is a sign that the two observers agreed less than would be expected just by chance. We rarely get perfect agreement. While the interpretation of kappa is somewhat arbitrary (and very task-dependent), Landis & Koch defined the following interpretation system for kappa [354] (see **Table 12**).

Table 12. Interpreting Cohen's kappa values. For negative values, the outcome is inversed.

kappa	Agreement
< .20	Poor agreement
.20 to .40	Fair agreement
.40 to .60	Moderate agreement
.60 to .80	Good agreement
.80 to 1.00	Very good agreement

4.3.3 Negative log-likelihood

Log-likelihood is a type of MLE (Maximum Likelihood Estimate) [355]. One of the most fundamental concepts of modern statistics is that of likelihood. Likelihood is a tool for summarizing the data's evidence about unknown parameters. It is the likelihood of a classification model predicting a data set. Log-likelihood is calculated by the product of the probabilities of each point. However, for both theoretical and practical reasons, working with the logarithm is easier than working with the product of the probabilities at every data point [356] as the likelihood would probably underflow. Log-likelihood takes advantage of the transformation of product to addition in exponentials of similar base ($\log(xy) = \log x + \log y$) to simplify the product of probabilities into a sum. Therefore, the math is a simpler for optimizing, as a sum instead of a product. For example, if we have a million data points, each with a probability of .7, the likelihood is $(.7)^{1{,}000{,}000}$, which is much smaller than computers can represent accurately using the standard floating-point notation, but the log-likelihood will still be a reasonable number. Negative log-likelihood can be calculated from the following formula seen in **Equation 35**, where $p_{y_i}(x_i)$, is the probability of all true predicted outcomes of class y being predicted at item i. x_i is a list of all true predicted outcomes with length of n.

Negative log-likelihood = $-\frac{1}{n}\sum_{i=1}^{n} \log\left(p_{y_i}(x_i)\right)$ **Equation 35**

4.3.4 Confusion matrices

The Confusion Matrix gives the number/proportion of instances between the predicted and actual class. TP and TN give insight into the ability of a machine learning model's performance to distinguish between the majority and minority classes. AUC combines these two. We compare the Confusion Matrices of the different machine learning methods explored in the literature review and extend their methods with other machine learning methods that have not been explored.

4.3.5 Machine learning approaches

The Python programming language has established itself as one of the most popular languages for scientific computing. Thanks to its high-level interactive nature and its mature ecosystem of scientific libraries, it is an attractive choice for algorithmic development and exploratory data analysis [357] [358]. However, as a general-purpose language, it is increasingly used not only in academic contexts but also in industry. Scikit-learn uses this rich environment to provide state-of-the-art implementations of many well-known machine learning algorithms while maintaining an easy-to-use interface that is tightly integrated with the Python language [359].

Scikit-learn is advantageous compared to other machine learning toolboxes in Python for the following reasons: Firstly, it is distributed under the BSD license (free for use). Secondly, it incorporates compiled code for efficiency, unlike MDP [360] and PyBrain [361], so it is more CPU and memory efficient. Thirdly, it depends only two other libraries NumPy [362] and SciPy [363], unlike PyMVPA [364] (a competing Python machine learning platform) that has numerous dependencies such as R [365], NiBabel [366] and Shogun [367], which can cause implementation incompatibilities. Lastly, it focuses on imperative programming [367], unlike PyBrain which uses a data-flow framework, this makes Scikit-learn model the mathematical algorithms closely, making code interpretation easy. While the package is mostly written in Python, it incorporates the C++ libraries LibSVM [368] and LibLinear [369] that provide reference implementations of SVMs and generalized linear models with compatible licenses.

In the following section, we describe the machine learning algorithms adopted to use in this work from Scikit-learn for Python.

Random forest

Random forest is an ensemble learning method used for classification, regression, and other machine learning tasks. Random forest tree is a machine learning algorithm based on decision trees. It was first proposed by Tin Kam Ho [370] and further developed by Leo Breiman and Adele Cutler [371], [372]. It is robust to overfitting [371]. A random forest builds a set of decision trees. Each tree is compiled from a bootstrap sample of training data. When developing individual trees, any subset of attributes is drawn (hence the term "random") from which the best split attribute is selected. The final model is based on the vote of most individually developed trees in the forest. In random forest, predictions are made about the class, not simply based on one decision trees, but by an (almost) unanimous prediction, made by 'K' decision trees. For this reason, the classification of random forest is truly ensemble i.e., for each decision tree, the class of instance is predicted, but the result is the class that was predicted the most often.

Random forest works for both classification and regression tasks. Studies suggest that random forest provides consistent pairwise similarity, which is crucial for multimodal data [373]. Pairwise-similarity facilitates the combination of features, adding higher dimensionality to the feature space whilst being less sensitive to data sample size [374].

The random forest library used in this work was from the Python library sci-kit-learn v0.23 [375], the 'RandomForestClassifier' [376]. The classifier parameters were set to create 200 trees, consider all 9 features at every split and the pruning setting was set so that nodes would split into 5 nodes or more. Random forest created the best fit model of attention and inattention in this work, which is presented in section 4.3.6, Classification results.

AdaBoost

This is another ensemble meta-algorithm that combines weak learners and adapts to the 'hardness' of each training sample. The AdaBoost (short for "Adaptive boosting") machine-learning algorithm was formulated by Yoav Freund and Robert Schapire [377]. It can be used with other learning algorithms to improve their performance. This is done by optimizing the weak learners. In this work we use the Python AdaBoost library from sci-kit-learn v0.23 [375], 'AdaBoostClassifier' [378]. The parameters were set to the following number of estimators: 100, learning rate: 1, boosting classification algorithm: SAMME.R [379] (updates base estimator's weight with probability estimates) and the regression loss function was linear.

Neural Network

An Artificial Neuron Network (ANN) is a computational model based on the structure and functions of biological Neural Networks (NN). The information that flows through the network affects the structure of the ANN because a neural network changes or learns, based on that input and output. ANNs are considered nonlinear statistical data modeling tools in which complex relationships between inputs and outputs are modeled or patterns are found.

ANN is also known as a neural network. We use a multi-layer perceptron (MLP) algorithm with backpropagation from the sci-kit-learn v0.23 [375] Python library 'MLPClassifier' [380]. Our model parameters were set to create 200 neurons in the hidden layer. We used ReLU (the Rectified Linear Unit function) as the activation function for the hidden layer [381] and the Adam optimizer for weight optimization [382].

Networks that use the rectifier function for hidden layers are often referred to as rectified networks. The adoption of ReLU can easily be considered one of the few milestones that lead to the development of very deep neural networks. Since ReLU is almost linear, they preserve many properties that make linear models easy to optimize with gradient-based methods. They also preserve many of the properties that make linear models generalizable. As a component of an artificial neuron in artificial neural networks (ANN), the activation function is responsible for processing weighted inputs and helping to deliver an output. When stacking more and more layers in an MLP, it has been empirically observed that an MLP with ReLU is much easier and faster to train than an MLP with *tanh*. The Adam optimizer was used as the solver for weight optimization. When introducing the Adam optimizer algorithm, the authors list the attractive benefits of using Adam on non-convex optimization problems [382]:

- Straightforward to implement
- Computationally efficient
- Little memory requirements
- Invariant to diagonal rescale of the gradients
- Well suited for problems that are large in terms of data and/or parameters
- Appropriate for non-stationary objectives
- Appropriate for problems with very noisy/or sparse gradients
- Hyper-parameters have intuitive interpretation and typically require little tuning

Non-convex problems are multivariate polynomials of a degree of 4 or higher [383]. Our work has 9 features (7 independent) used in the classification problem, making it a non-convex problem. Other MLP parameters; Alpha: L2 penalty (regularization term): .0001 and the maximum number of iterations: 200.

k-NN

A k-Nearest-Neighbor algorithm, often abbreviated k-NN, is an approach to data classification that estimates the probability that a data point is a member of one group or another according to the group in which the closest data points are located. The k-NN algorithm searches for *k* closest training examples in feature space and uses their average as a prediction. k-NN is an example of a "lazy learner" algorithm because it does not generate a data set model in advance. A k-NN is a data classification algorithm that attempts to determine which group a data point is in by looking at the data points around it. An algorithm, observing a point on a grid, trying to determine if a point is in group A or B, observes the states of neighboring points. The interval is determined arbitrarily, but the point is to take a sample of the data. If most of the points belong to group A, the data point in question is likely to be A instead of B, and vice versa. It is predicted according to the nearest training instances. We use the Python k-NN library from sci-kit-learn v0.23 [375], 'KNeighborsClassifier' [384].

The number of nearest neighbors: 20. The distance parameter (metric): Euclidean (“straight line,” distance between two points). Uniform weights as model criteria (all points in each neighborhood are weighted equally).

Naïve Bayes

Naïve Bayes is a quick and simple probabilistic classifier based on Bayes' theorem with the hypothesis of independence from characteristics. A naïve Bayes classifier uses probability theory to classify data. Naïve Bayes classification algorithms make use of Bayes' theorem. Bayes' theorem takes its name from Reverend Thomas Bayes (1702-1761), who studied how to calculate a distribution for the probability parameter of a binomial distribution. After Bayes' death, his friend Richard Price cured and presented this work in 1763, An Essay towards solving a Problem in the Doctrine of Chances [385]. The foundation of Bayes' theorem is that the probability of an event can be regulated by the introduction of new data. What makes a Bayes classifier naive is its assumption that all the attributes of a data point under consideration are independent of each other. A naive Bayes classifier is not a single algorithm, but a family of machine learning algorithms that take advantage of statistical independence. These algorithms are relatively easy to write and execute more efficiently than more complex Bayes algorithms. In this work we use the Python naïve Bayes library from Orange v3.25 [386], ‘OWBaseLearner’ [387].

Decision tree

A decision tree [388] is a graphical representation of specific decision situations that are used when complex branching occurs in a structured decision process. A decision tree is a predictive model based on a branched series of Boolean tests that use specific facts to draw more general conclusions. The main components of a decision tree involve decision points represented by specific nodes, actions and choices from a decision point. Each rule in a decision tree is represented by tracing a series of paths from the root to the node to the next node and so on until an action is reached. It is a tree with a forward pruning algorithm. This is a simple algorithm that divides data into nodes by class purity. It can also be used for classification and regression tasks. A binary tree (divided into two child nodes) was built, with 2 minimum instances in leaves; the algorithm will never create a split that would place less than 2 training examples in any of the branches. Subsets smaller than 5 instances were never split. The maximal tree depth was limited to 200 node levels. We use the Python decision tree library from sci-kit-learn v0.23 [375], ‘DecisionTreeClassifier’ [389].

SVM

Support for input vector machines (SVMs) maps to higher-dimensional feature spaces. SVM is a machine learning technique that separates the attribute space by a hyperplane, thus maximizing the margin between instances of different classes or class values. This technique often gives excellent results in terms of prediction. A popular SVM implementation from the LIBSVM package has been adopted. We use the Python SVM library from sci-kit-learn v0.23 [375] based on LIBSVM [390], ‘svm.LinearSVC’ [391]. Parameters for this library are the following. Cost is 1 (cost is the penalty term for loss and applies for classification and regression tasks). A Radial Basis Function (RBF) kernel is adopted with a gamma constant of ⅛ and the numerical deviation tolerance is .001 with an iteration limit of 100.

Stratified random classifier

The random classifier used in this work has been stratified to distribute the same proportion of classes that exist in the sample data [392]. For example, if a data set of 100 fruit includes 95 apples and 5 oranges, the random classifier would distribute classification outcomes proportionally at random with a 95% chance of 'apple' and a 5% chance of 'orange'. For this reason, a stratified random classifier achieves a balanced result to the sample data, meaning that every classification category, either in the majority or minority has an equal chance per capita to be classified correctly. This is why a stratified random classifier achieves a .50 AUC. We implemented this classifier to benchmark other classifiers.

Constant classifier

This learner, when predicting the class value with predictions, will return the relative frequencies of the classes in the learning set. When there are two or more majority classes, the classifier chooses the predicted class randomly, but always returns the same class for a particular example. It will be the reference for other models. Since a constant classifier completely disregards the minority class in the classification outcome, the constant classifier achieves an AUC of 0. The output of this classifier is always set to the majority class, in this case, 'attention'.

4.3.6 Classification results

Multimodal Multisensor data was collected during the length of 59 sessions. In total 2615 samples were gathered and labeled into two categories ('attention' and 'inattention') using nine features. The summary of the results is shown in **Table 13**. Overall, the random forest classification approach achieved the best classification results in all modes of data. This was both when including High-Level Compound Features, and when only using a sub-set of the data modes. This finding is supported by other studies [373], which suggests that random forest provides consistent pairwise similarity, crucial for Multimodal Multisensor data. Pairwise-similarity facilitates the combination of features, adding higher dimensionality to the feature space, whilst being less sensitive to data sample size [374].

We compared two different approaches to grouping our data. In the first approach, we consider the participants as independent, in the second approach we group the participant data together and model it as a whole. In the first approach, the models created have information about the Multimodal Multisensor ownership, in the second approach, the model sees a single owner, the group. The reason for grouping the data is to see if the data generalizes across the participants well.

In the first approach, the data was grouped by participant or person-specific. We found that the best classification model was achieved with random forest. The random forest model using both high- and low-level features achieved an 84.8% classification for 'attention' and a 65.4% accuracy for 'inattention'. This model has an 80.7% coverage of the AUC. The random forest performance improved NN performance by 5.5% better classifications of 'inattention' (TN, the minority class) and 0.3% better classifications of 'attention' (TP, the majority class). In the second approach, all data was put in a single group, and all participant data was considered as a whole. Random forest using both high- and low-level features achieved 93.3% classification for 'attention' and a 42.9% accuracy for 'inattention'. This was improved even more over the NN method, by 11.4% better classification of 'inattention' and 0.2% better 'attention'.

The improvement of classification of inattention, when we considered the data in participant groups, was expected as the characteristics of the participant (how they display their patterns of attention and inattention) are better represented individually than as a heterogeneous group. This however does not take away from the results found when we considered the data as a whole. The significance of looking at the results as a whole is to see how the patterns generalize across the participants. The random forest method performed 22.5% better at classifying 'inattention' when we considered the data per participant than when we considered the data as a group (TN = 65.4%). This means that a system that didn't have participant identification (like in a walk-up-and-use system [216]) would perform around 42.9% at detecting 'inattention' and 93.8% of 'attention'. However, a system that had participant information would perform around 65.4% at detecting 'inattention' and around 84.8% at detecting 'attention'.

The random forest method incorporated 100 trees and all nine features were included at each of the 255 nodes, with 128 leaves in total. The NN had 200 hidden layers and used the Adam optimizer [382] as a solver for weight optimization. AdaBoost, (another ensemble method), outperformed random forest for the single modality feature classification. However, in every example, using any machine learning method, Multimodal Multisensor data features delivered significantly better classification results than any single modality.

When compared to the second-best classification method, random forest outperforms neural network on the classification of 'inattention' classes with a margin of 16.5% and has an 11.7% better coverage in AUC (see **Table 13**). Besides neural networks, other machine learning methods were also assessed; AdaBoost, decision tree, k-Nearest Neighbor, naïve Bayes, and support vector machine. However, all had inferior performance when compared to random forest.

Including the two high-level handpicked features (HLA and HLI) in the feature space improved the classification in every sensor combination and every machine learning methodology. In the random forest model including HLCF increased the AUC by 1.5% more coverage, and the classification of True Positives (TP) by 0.9%, and True Negatives by 2.8%. On average, if only two modes of sensor input were available, including interaction data improves the outcome of AUC coverage by 16.8%, compared to any other two modes of data, making interaction data the single most important secondary feature. The single most important mode of data on its own however is eye-gaze, with 3.2% better AUC coverage compared to interaction data.

The system developed using these machine learning models would not be affected by both sensor fallout and occlusions. At best (all high- and low-level features using random forest) 80.7% AUC coverage is achieved. Using a sub-set of three sensor modes 78.1%-73% AUC coverage is achieved, whilst with a subset of two sensor modes (including interaction) 76.5%-69.2% AUC coverage is achieved. Using a subset of two sensor modes (not including interaction) 70.8%-61.0% AUC coverage is achieved, and with only a single mode of sensor data between 63.7%-48.8% AUC coverage is achieved.

Table 13. Best classification results achieved with random forest using multi-level feature fusion

Features	Best Classification Method Found	Negative log-likelihood for Attention (Less is better)	Negative log-likelihood For Inattention (Less is better)	kappa	AUC	TP	TN	F1	Precision	Recall
All Participant spec.	Random Forest	0.1342	0.9574	.3665	.807	84.8%	65.4%	.805	.808	.826
All Participant spec.	Neural Network	0.1826	0.9527	.3390	.799	84.5%	59.9%	.796	.794	.826
All features	Random Forest	0.1377	1.0149	.418	.803	93.8%	42.9%	.819	.817	.833
Low-Level		0.1440	0.9753	.374	.788	92.9%	40.1%	.806	.802	.820
High-Level		0.1250	0.9547	.237	.686	93.1%	26.4%	.768	.762	.794
All features	Neural Network	0.1237	0.8191	.300	.773	93.6%	31.5%	.786	.783	.808
Low-Level		0.1203	0.7910	.273	.767	95.3%	26.4%	.781	.783	.811
All features	AdaBoost	0.2775	2.1803	.388	.794	93.3%	40.7%	.810	.807	.824
Low-Level		0.2756	2.0422	.335	.765	90.7%	39.7%	.791	.785	.802
All features	Naïve Bays	0.1341	0.7574	.233	.712	91.3%	28.7%	.764	.755	.784
Low-Level		0.1097	0.7463	.095	.728	98.0%	8.50%	.732	.748	.796
All features	k-NN	0.1105	1.4139	.207	.746	96.4%	19.2%	.763	.771	.804
Low-Level		0.1115	1.5077	.169	.730	96.5%	15.9%	.752	.760	.799
All features	Tree	0.2545	1.6061	.309	.706	89.9%	38.4%	.782	.776	.793
Low-Level		0.1157	1.6414	.258	.686	89.4%	34.0%	.767	.760	.780
All features	SVM	0.1107	0.6620	.086	.454	76.2%	33.2%	.686	.701	.673
Low-Level		0.1026	0.6750	.059	.467	72.7%	34.0%	.667	.693	.647
Eye + EEG + Inter.	Random Forest	0.1429	1.0202	.349	.765	92.4%	38.4%	.793	.793	.812
Eye + Body + Inter.	Random Forest	0.1433	1.0784	.371	.781	91.6%	41.9%	.803	.798	.814
EEG + Body + Inter.	Random Forest	0.1619	1.5544	.318	.730	91.9%	36.0%	.788	.783	.804
Eye + EEG	Random Forest	0.1335	0.7164	.277	.679	95.0%	27.1%	.781	.783	.810
Eye + Body	Random Forest	0.1619	1.5544	.318	.708	93.7%	27.9%	.776	.772	.801
EEG + Body	AdaBoost	0.4902	2.2224	.122	.610	84.2%	27.4%	.719	.713	.725
Eye + Inter.	Random Forest	0.1380	1.1820	.308	.765	93.6%	32.2%	.788	.785	.810
Body + Inter.	AdaBoost	0.2579	2.0817	.327	.692	83.5%	52.1%	.776	.783	.770
EEG + Inter.	AdaBoost	0.3002	2.0323	.246	.708	85.7%	38.3%	.756	.753	.759
EEG	AdaBoost	0.2821	0.8682	.100	.559	84.6%	24.8%	.714	.706	.723
Eye-gaze	AdaBoost	0.2646	1.3051	.255	.637	89.2%	34.0 %	.766	.758	.778
Body	AdaBoost	0.3091	1.0059	.003	.488	93.2%	7.10%	.702	.674	.754
Interaction	AdaBoost	0.2491	0.6092	.035	.605	95.8%	6.70%	.713	.694	.774
Any	Stratified Random (P=78.43%)	1.5222	7.0000	-.001	.500	78.5%	21.5%	.788	.7914	.7845
Any	Constant Classifier	0.1004	0.6854	.000	.000	100%	0.00%	.702	.630	.794

4.4 Correlations between CPT, Sensor and Participant

In this section, we explore the correlations between CPT outcomes, Multimodal Multisensor data and participant characteristics.

4.4.1 Data points

Data collected in the Multimodal Multisensor study from the four participants is averaged by session, and by week. After the two weeklong pilot study. The data collection pilots lasted 11 weeks, up to 4 sessions per participant were collected weekly and a total of 59 sessions were obtained. Data collected in the sessions can be categorized by source, n two main categories, 'platform' and 'participant' seen in **Equation 36**. These categories include SDT outcomes, CPT outcomes, Multimodal Multisensor data features and participant characteristics. The data collected from the platform over the length of 59 sessions is then correlated against the participant characteristics. The normalized mean of each data point as per session and week were evaluated for correlations between other data points. In total 2615 frames of data were collected; 55 data points were calculated and over 2475 cross-correlation assessments between each of the data points were evaluated longitudinally for significance.

$$\text{Data point sources} = \begin{cases} Platform \begin{cases} SDT\ outcomes \\ CPT\ outcomes \\ MM\ features \end{cases} \\ Participant \begin{cases} Age \\ P\ scales \end{cases} \end{cases} \qquad \textbf{Equation 36}$$

The 55 separate data parameters were evaluated from each session in four categories. These included SDT and CPT outcomes measures, Multimodal Multisensor features and participant characteristics. We investigate the correlations between the data points in the category of 'platform' against themselves and the data points from 'platform' with 'participant'. While the participant was participating in the CPT, Multimodal Multisensor data from eye-gaze, EEG, body pose, and interaction data were collected. This data is synchronized with the SDT and CPT outcomes and an analysis of their significance and relationships is discussed in section 4.5, Correlation summary. A summary of the CPT and SDT data points was presented in section 2.7, a summary of the Multimodal Multisensor data points was presented in section 4.1 and participant characteristics were described in section 3.4. Some of the main platform data points are listed in **Table 14**. Participant characteristics such as age and their progress through the academic curriculum with respect to performance attainment targets 'P scales' were also collected [237]. These are listed in **Table 15**. The data points in regard to learner P scales are 'P scales mean', 'Speaking', 'Listening, 'Reading', 'Writing', 'Maths Space, Shape, Measure', 'Maths using and applying' and 'P scale Computing'.

4.5 Correlation summary

The data collection process took over 11 weeks, 2615 data samples were collected, which we then averaged to 59 sessions. Overall, 55 data points from the categories in (1) were assessed for cross-correlation significance against participant characteristics in the *Participants* section, a summary of some of these data points is shown in **Table 14** and **Table 15**. This leads to 2475 separate correlation tests using Pearson's correlation coefficient r [393]–[395]. Even though this sample count may seem low, 2615 SDT samples falls within the norms of prior studies conducted with a similar target group. Swanson's study collected 2240 SDT signal challenges from the participants with learning difficulty [15]. The study conducted by Goldberg *et al.* had a total of 240 SDT signal samples [166].

Table 14. Summary of data points and sources.

Data points	Attribution
Discriminability = d'	SDT
Bias = B''_D	SDT
Hit rate = H	SDT
False Alarm Rate = FAR	SDT
Correct Omissions	SDT
Wrong Omissions	SDT
Response time	CPT
Single fast press	CPT
Max press count	CPT
Eye scanning	Eye-gaze
Eye dwelling	Eye-gaze
Eyes off screen	Eye-gaze
Eye-gaze inverse entropy	Eye-gaze
EEG Alertness	EGG
Body fidgeting	Body pose
Body joint speed	Body pose
Body joint entropy	Body Pose
HLA	Compound feature
HLI	Compound feature

Table 15. Summary of participant characteristics.

Data points	Attribution
Age	Participant
P scales mean	Participant
P scale Speaking	Participant
P scale Listening	Participant
P scale Reading	Participant
P scale Writing	Participant
P scale Maths Space, Shape, Measure	Participant
P scale Maths using and applying	Participant

Some of the most interesting correlations have been presented with their relevant r values in **Table 16** and **Table 17**. The first table displays the cross-correlations between the SDT and CPT outcomes (like d' and B''_D) and Multimodal Multisensor data (like EEG, eye-gaze, and body pose) with participant characteristics (like individual P scales and age). The second table looks closer at the correlation between the SDT and CPT outcomes themselves. For both calculations, only the longitudinal correlations were considered, i.e. the correlations for the length of the 59 sessions, therefore N = 59. The results of these tests show a significant correlation between some of the data points, and a summary of the most interesting results is described in the following section.

Table 16. A summary of some of the correlation between the Multimodal Multisensor data and the CPT outcomes and participant characteristics. With correlation values in the designated cross-sections (N = 59).

	P scales	Speaking	Listening	Reading	Writing	Maths Space, Shape, Measure	Maths using and applying	Computing	Age	AHLF	DHLF	d'	B"$_D$	H	FAR	Miss %	Omission %	Gaze Inv-Entropy	Fidgeting	Joint entropy	EEG Attention	Eyes Off Screen	Eye Scanning	Eye Dwelling	Single Press	Press Count	Response Time
P scales	1.00	0.37	0.76	0.91	-0.31	0.55	0.88	0.88	0.66	1.00	0.67	0.99	-0.17	0.80	-0.99	-0.80	0.99	0.24	0.53	0.72	0.79	-0.71	0.49	0.61	0.52	0.11	0.46
Speaking	0.37	1.00	-0.24	0.30	-0.70	0.76	0.04	0.05	-0.45	0.38	-0.44	0.26	0.52	-0.17	-0.40	0.17	0.36	-0.79	-0.45	0.50	-0.05	0.23	-0.35	-0.22	0.76	-0.78	0.97
Listening	0.76	-0.24	1.00	0.56	-0.12	-0.12	0.71	0.72	0.87	0.74	0.93	0.85	-0.22	0.77	-0.77	-0.77	0.79	0.63	0.61	0.18	0.64	-0.99	0.91	0.94	-0.15	0.74	-0.06
Reading	0.91	0.30	0.56	1.00	0.03	0.72	0.96	0.96	0.68	0.92	0.63	0.84	-0.49	0.88	-0.85	-0.88	0.85	0.35	0.71	0.92	0.94	-0.47	0.19	0.30	0.70	-0.06	0.29
Writing	-0.31	-0.70	-0.12	0.03	1.00	-0.14	0.18	0.16	0.36	-0.31	0.25	-0.33	-0.88	0.30	0.42	-0.30	-0.39	0.68	0.62	0.07	0.33	0.21	-0.25	-0.34	-0.13	0.20	-0.85
Maths Space, Shape, Measure	0.55	0.76	-0.12	0.72	-0.14	1.00	0.51	0.51	-0.01	0.57	-0.08	0.40	-0.16	0.33	-0.50	-0.33	0.48	-0.30	0.16	0.92	0.49	0.20	-0.45	-0.32	1.00	-0.73	0.64
Maths using and applying	0.88	0.04	0.71	0.96	0.18	0.51	1.00	1.00	0.85	0.88	0.81	0.85	-0.62	0.98	-0.81	-0.98	0.83	0.59	0.86	0.79	0.99	-0.62	0.37	0.44	0.49	0.20	0.05
Computing	0.88	0.05	0.72	0.96	0.16	0.51	1.00	1.00	0.85	0.89	0.81	0.86	-0.61	0.98	-0.82	-0.98	0.84	0.58	0.85	0.79	0.99	-0.62	0.37	0.45	0.49	0.20	0.07
Age	0.66	-0.45	0.87	0.68	0.36	-0.01	0.85	0.85	1.00	0.65	0.99	0.71	-0.68	0.94	-0.61	-0.94	0.64	0.89	0.92	0.37	0.86	-0.80	0.66	0.66	-0.03	0.67	-0.37
AHLF	1.00	0.38	0.74	0.92	-0.31	0.57	0.88	0.89	0.65	1.00	0.66	0.98	-0.18	0.80	-0.99	-0.80	0.99	0.23	0.52	0.74	0.80	-0.69	0.47	0.59	0.55	0.09	0.47
IHLF	0.67	-0.44	0.93	0.63	0.25	-0.08	0.81	0.81	0.99	0.66	1.00	0.74	-0.57	0.90	-0.64	-0.90	0.67	0.86	0.86	0.29	0.80	-0.87	0.76	0.76	-0.11	0.73	-0.33
d'	0.99	0.26	0.85	0.84	-0.33	0.40	0.85	0.86	0.71	0.98	0.74	1.00	-0.14	0.80	-0.99	-0.80	0.99	0.31	0.53	0.60	0.76	-0.81	0.63	0.73	0.37	0.27	0.39
B"$_D$	-0.17	0.52	-0.22	-0.49	-0.88	-0.16	-0.62	-0.61	-0.68	-0.18	-0.57	-0.14	1.00	-0.70	0.05	0.70	-0.09	-0.81	-0.90	-0.45	-0.73	0.10	0.05	0.09	-0.15	-0.22	0.63
H	0.80	-0.17	0.77	0.88	0.30	0.33	0.98	0.98	0.94	0.80	0.90	0.80	-0.70	1.00	-0.73	-1.00	0.76	0.74	0.93	0.67	0.98	-0.68	0.46	0.51	0.31	0.38	-0.14
FAR	-0.99	-0.40	-0.77	-0.85	0.42	-0.50	-0.81	-0.82	-0.61	-0.99	-0.64	-0.99	0.05	-0.73	1.00	0.73	-1.00	-0.18	-0.43	-0.65	-0.71	0.74	-0.55	-0.66	-0.47	-0.14	-0.51
Miss %	-0.80	0.17	-0.77	-0.88	-0.30	-0.33	-0.98	-0.98	-0.94	-0.80	-0.90	-0.80	0.70	-1.00	0.73	1.00	-0.76	-0.74	-0.93	-0.67	-0.98	0.68	-0.46	-0.51	-0.31	-0.38	0.14
Omission %	0.99	0.36	0.79	0.85	-0.39	0.48	0.83	0.84	0.64	0.99	0.67	0.99	-0.09	0.76	-1.00	-0.76	1.00	0.22	0.47	0.64	0.74	-0.76	0.56	0.68	0.45	0.17	0.48
Gaze Inv-Entropy	0.24	-0.79	0.63	0.35	0.68	-0.30	0.59	0.58	0.89	0.23	0.86	0.31	-0.81	0.74	-0.18	-0.74	0.22	1.00	0.89	0.08	0.65	-0.57	0.51	0.44	-0.31	0.75	-0.75
Fidgeting	0.53	-0.45	0.61	0.71	0.62	0.16	0.86	0.85	0.92	0.52	0.86	0.53	-0.90	0.93	-0.43	-0.93	0.47	0.89	1.00	0.53	0.91	-0.51	0.33	0.33	0.15	0.44	-0.47
Joint entropy	0.72	0.50	0.18	0.92	0.07	0.92	0.79	0.79	0.37	0.74	0.29	0.60	-0.45	0.67	-0.65	-0.67	0.64	0.08	0.53	1.00	0.79	-0.08	-0.21	-0.09	0.91	-0.44	0.40
EEG Attention	0.79	-0.05	0.64	0.94	0.33	0.49	0.99	0.99	0.86	0.80	0.80	0.76	-0.73	0.98	-0.71	-0.98	0.74	0.65	0.91	0.79	1.00	-0.54	0.29	0.35	0.47	0.19	-0.06
Eyes Off Screen	-0.71	0.23	-0.99	-0.47	0.21	0.20	-0.62	-0.62	-0.80	-0.69	-0.87	-0.81	0.10	-0.68	0.74	0.68	-0.76	-0.57	-0.51	-0.08	-0.54	1.00	-0.96	-0.98	0.23	-0.76	0.03
Eye Scanning	0.49	-0.35	0.91	0.19	-0.25	-0.45	0.37	0.37	0.66	0.47	0.76	0.63	0.05	0.46	-0.55	-0.46	0.56	0.51	0.33	-0.21	0.29	-0.96	1.00	0.99	-0.48	0.86	-0.12
Eye Dwelling	0.61	-0.22	0.94	0.30	-0.34	-0.32	0.44	0.45	0.66	0.59	0.76	0.73	0.09	0.51	-0.66	-0.51	0.68	0.44	0.33	-0.09	0.35	-0.98	0.99	1.00	-0.35	0.78	0.02
Single Press	0.52	0.76	-0.15	0.70	-0.13	1.00	0.49	0.49	-0.03	0.55	-0.11	0.37	-0.15	0.31	-0.47	-0.31	0.45	-0.31	0.15	0.91	0.47	0.23	-0.48	-0.35	1.00	-0.75	0.63
Press Count	0.11	-0.78	0.74	-0.06	0.20	-0.73	0.20	0.20	0.67	0.09	0.73	0.27	-0.22	0.38	-0.14	-0.38	0.17	0.75	0.44	-0.44	0.19	-0.76	0.86	0.78	-0.75	1.00	-0.60
Response Time	0.46	0.97	-0.06	0.29	-0.85	0.64	0.05	0.07	-0.37	0.47	-0.33	0.39	0.63	-0.14	-0.51	0.14	0.48	-0.75	-0.47	0.40	-0.06	0.03	-0.12	0.02	0.63	-0.60	1.00

Table 17. A summary of the correlations between the SDT and CPT outcomes measures of the 59 CPT sessions.

	Session Number	Week Number	d' (D Prime)	BiasD	Hit Rate	FA (False Alarm rate)	Mean Press Count	mean First Response Time	mean Response Times	Hit %	Miss %	Wrong Commission %	Omission %
Session Number	1.00	0.82	-0.07	-0.11	0.07	-0.01	0.17	-0.46	-0.44	0.07	-0.07	-0.01	0.06
Week Number	0.82	1.00	-0.03	-0.08	-0.06	0.02	0.11	-0.18	-0.14	-0.06	0.06	0.02	1.56×10^{-4}
d' (D Prime)	-0.07	-0.03	1.00	0.22	0.51	-0.68	-0.11	-0.34	-0.37	0.51	-0.51	-0.68	0.45
BiasD	-0.11	-0.08	0.22	1.00	-0.46	-0.64	-0.20	0.31	0.23	-0.46	0.46	-0.64	0.58
Hit Rate	0.07	-0.06	0.51	-0.46	1.00	-0.06	0.12	-0.70	-0.69	1.00	-1.00	-0.06	-0.02
FA (False Alarm rate)	-0.01	0.02	-0.68	-0.64	-0.06	1.00	0.22	0.14	0.23	-0.06	0.06	1.00	-0.85
Mean Press Count	0.17	0.11	-0.11	-0.20	0.12	0.22	1.00	-0.22	-0.08	0.12	-0.12	0.22	-0.19
mean First Response Time	-0.46	-0.18	-0.34	0.31	-0.70	0.14	-0.22	1.00	0.96	-0.70	0.70	0.14	-0.02
mean Response Times	-0.44	-0.14	-0.37	0.23	-0.69	0.23	-0.08	0.96	1.00	-0.69	0.69	0.23	-0.13
Hit %	0.07	-0.06	0.51	-0.46	1.00	-0.06	0.12	-0.70	-0.69	1.00	-1.00	-0.06	-0.02
Miss %	-0.07	0.06	-0.51	0.46	-1.00	0.06	-0.12	0.70	0.69	-1.00	1.00	0.06	0.02
Wrong Commission %	-0.01	0.02	-0.68	-0.64	-0.06	1.00	0.22	0.14	0.23	-0.06	0.06	1.00	-0.85
Omission %	0.06	1.56×10^{-4}	0.45	0.58	-0.02	-0.85	-0.19	-0.02	-0.13	-0.02	0.02	-0.85	1.00

1.0 0.5 0 -0.5 -1.0

Participant progress in their academic curriculum (P scales mean) had a strong positive correlation with participant ability to maintain attention in the CPT (d'), r(57) = .986, P = .0068. This agrees with other studies that found that sustained attention related to greater academic outcomes [73], [84], [88], [90], [92]. Participant progress in their academic curriculum however has a strong negative correlation with their impulsivity (FAR), r(57) = -.991, P = .004. Participants who can maintain selective and sustained attention for longer, and do not give in to impulsive responses when fatigued, had progressed further in their academic studies, as assessed with P scales. This echoes other longitudinal studies that found that self-control and inhibition, increase academic performance [73], [77].

Participants that were selective in their responses and could hold back their actions until they confirmed the signal (through higher Correct Omissions) had progressed further in their P scales r(57) = .995, P = .0025. This selectivity and ability to be less impulsive (with higher Correct Omissions), also has a strong positive correlation with the participant's ability to maintain attention for longer (d'), r(57) = .994, P = .0029.

Despite studies [1], [140], [143], [151] suggesting that children with learning difficulties are traditionally more conservative, we found that out of the 59 sessions, 24 sessions had a positive bias, 33 had negative bias and 2 were neutral. Participant bias was overall liberal (negative) $\overline{B''_D}$ = -.1105, $\sigma(B''_D) = .6323$, as seen in **Figure 42**. Participants showed no significant bias change during the sessions r(57) = -.078, P = .7205. While there are varied reports on bias (B''_D) and sensitivity (d') being independent variables in SDT [151], we found no significant correlation between them r(57) = -.138, P = .8615.

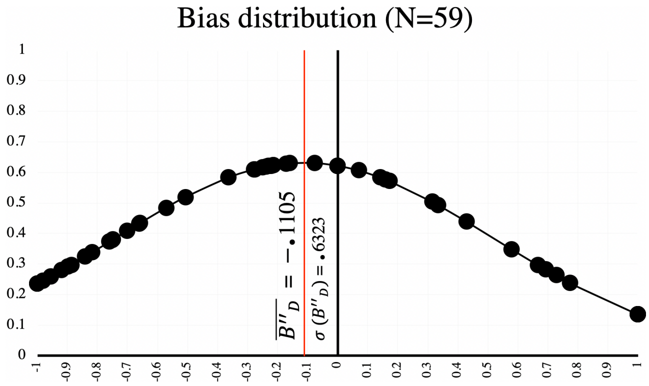


Figure 42. Bias distribution for participants with learning difficulty shows a slight negative outcome.

When the participants were more *alert* (EEG Alertness) they made more Hits r(57) = .981, P = .0097 and had faster response times r(57) = -.958, P = .0211. Participants that had higher EEG Alertness during their CPT, found more Hard Target (signals with more noise) r(57) = .971, P = .0144. These participants also had progressed further in both P scales 'Maths using and applying' and 'Computing' r(57) = .988, P = .0060, r(57) = .985, P = .0073. These two P scales indeed had a direct correlation with faster responses r(57) = -.918, P = .0408, r(57) = .913, P = .0436 and also greater Hit rate r(57) = .978, P = .0108, r(57) = .976, P = .0117. Participants with higher EEG alertness values had progressed further in their 'Reading' P scales r(57) = .9356, P = .0322.

Participants that have greater body movement range progressed further in their 'Reading' scales r(57) = .916, P = .0420 and participants who had showed use of their eye movements in structured form (eye scanning and eye dwelling) had progressed further in their 'Listening' P scales r(57) = .915, P = .0426, r(57) = .03.

Participants that did better in only pressing the button once (not having multiple button presses on a single attempt) had progressed further in their 'Maths Space, Shape, Measure' P scale r(57) = .999, P = .0002 and did better in their 'Speaking' P scale r(57) = .966, P = .0173. Older participants had faster response times r(57) = -.933, P = .0333.

Participants that had higher HLA feature values had progressed further in their 'Reading' P scale, r(57) = .915, P = .0423, had more control and inhibition in the CPT activity (less False Alarms) r(57) = -.989, P = .0054 and were able to hold selective and sustained attention for longer periods of time (d').

Despite participants having quicker response times after every session r(57) = .908, P = 4E-8, their responses did not become any more precise in the SDT challenge. Actually, participant CPT performances (d') worsened after every session slightly, but not significantly r(57) = -.141, P = .717.

Participant press count was significantly increased after every session r(57) = .644, P = .0015. Button press increased for all participants after each session, which could be a sign of cumulative fatigue [344].

4.6 Tools and empirical results

This section will describe the tools developed to assist the data collection process and some findings that were collected during the results analysis process.

4.6.1 Memory sharing over LAN between Matlab instances

The data collection platform developed in Matlab 2016a [221], would label all the data collected in real-time and save all session data to the disk. This was necessary to avoid any information loss if the systems crashed. As the system would not be able to label the data after it crashed, real-time labeling was developed as part of the data collection platform. Real-time labeling and saving of the data to disk meant that there was a significant workload on the system processors. To overcome this, four Matlab 2016a instances were run simultaneously on two independent computers that shared memory through LAN. An image of this platform can be seen in **Figure 17**. One Matlab instance administered the CPT and the other three initiated sensory connectivity, checked for human presence, a working data connection and lastly saved the sensor data to disk.

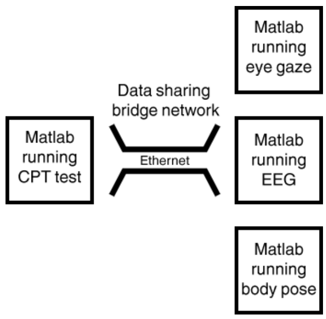


Figure 43. Memory sharing was developed between two computers, running four instances of Matlab 2016a to reduce single CPU workload.

This data bridge is visualized in **Figure 43**. The two systems communicated readiness and the CPT would automatically start, only after all sensors were ready, collecting human data and the CPT has been initiated with a participant profile (that set the CPT parameters).

4.6.2 Feature visualizer tool

A graphical user interface was developed to assist with the data collection strategy. A feature visualization tool was developed to quickly monitor the data results after each session, and monitor for data fall-out, sensors troubleshooting or to spot any errors in the setup or the system see **Figure 44**.

Figure 44. GUI for data collection validation.

The visualization tool has a filter for quick session lookup. Sessions can be selected by choosing the participant name, the session week number and the session name. All 48 slides are shown in the bottom timeline and their labels are displayed over the session number. A bracket adjustment tool limits the window of the slides count shown and can be used to zoom in to specific regions of the session. Each bar graph displays the magnitude of each of the features measured in the Multimodal Multisensor platform, within the constraints of the bracketed slides window.

4.6.3 Will’s data

Will often looks out of the window (this is to the left of this image), and he also looks downward between slides (at the piece of plastic he plays with in his hand). He is a great performer and, as **Figure 45** shows, he can still manage to find the targets despite all distractions from the outside. Will’s average correct response rate is 74.89% (Hits and Correct Omissions) and his average incorrect response rate is 25.11% (False Alarms and Misses). His mean EEG attention feature value is .4537 and his mean body fidgeting value is 47.6988 mm/s.

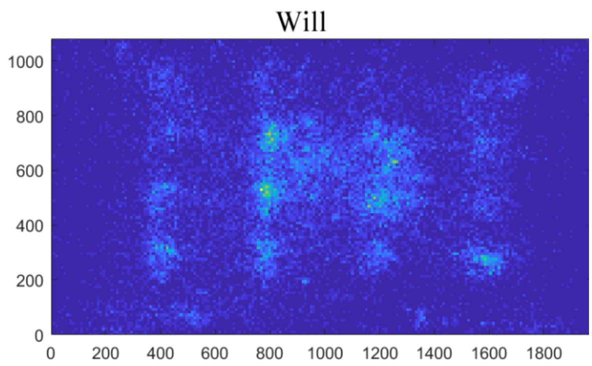


Figure 45. Will’s eye-gaze pattern heat map plotted in pixels.

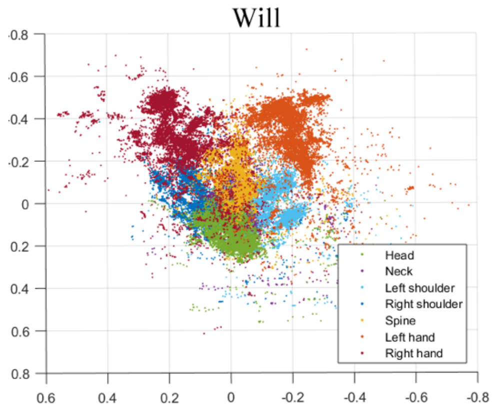


Figure 46. Will’s body joint position clusters plotted in meters.

See **Figure 46** for his body joint positions over all sessions. Will’s mean press count is 1.6027 presses per trial. His mean first response time is 1.9267 seconds. Will’s mean d' sensitivity level is 1.9506 and his mean B''_D bias is -.2608 (liberal).

4.6.4 Jen's data

Jen spends most of the activity looking at the center of the screen, and she is extremely good at keeping sustained attention. Sometimes her eyes wander off the screen but by keeping the screen in her peripheral vision she hardly ever Misses the targets (seen in **Figure 47**). Jen's average correct response rate is 96.04% (Hits and Correct Omissions) and her average incorrect response rate is 3.96% (False Alarms and Misses). Her mean EEG attention feature value is .4809 and her mean body fidgeting value is 48.9619 mm/s.

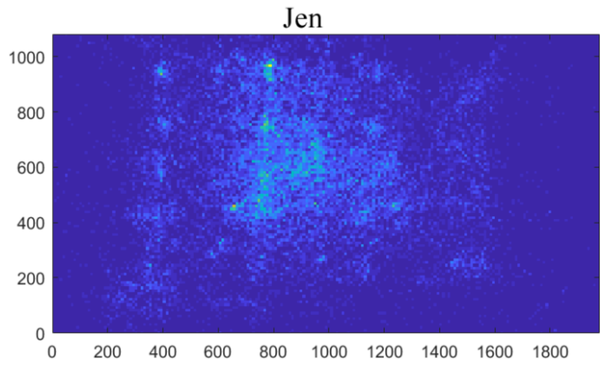


Figure 47. Jen's eye-gaze pattern heat map plotted in pixels.

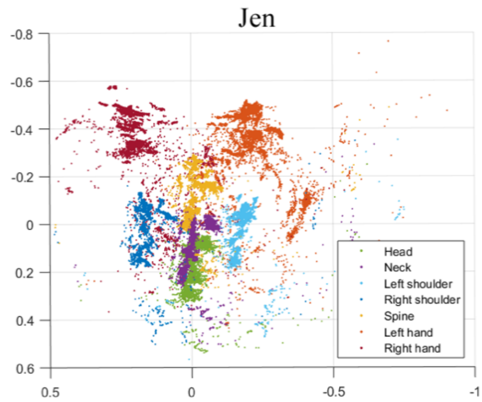


Figure 48. Jen's body joint position clusters plotted in meters.

See **Figure 48** for her body joint positions over all sessions. Jen's mean press count is 2.1524 presses per trial. Her mean first response time is 1.3927 seconds. Jen's mean d' sensitivity level is 6.889 and her mean B''_D bias is -.0729 (liberal).

4.6.5 Mark's data

Mark uses the switch and his voice to communicate the actions in the study. His eye movement is mainly focused on the top and center areas of the screen. At school he uses both eye-gaze technology to communicate and switches to interact with computer interfaces, so he is not a stranger to this type of technology. Mark has the most restrictions in his head movement and struggles to maintain concentrated eye-gaze. In spite of this, he persevered in the study and made his best effort to perform in the activity (see **Figure 49**). Mark's average correct response rate is 57.63% (Hits and Correct Omissions) and his average incorrect response rate is 42.37% (False Alarms and Misses). His mean EEG attention feature value is .3757 and his mean body fidgeting value is 22.4214 mm/s.

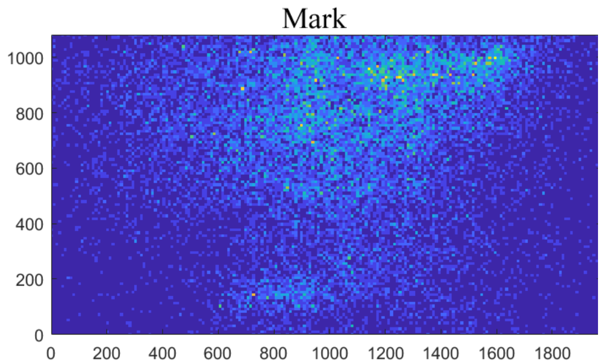


Figure 49. Mark's eye-gaze patterns plotted in pixels.

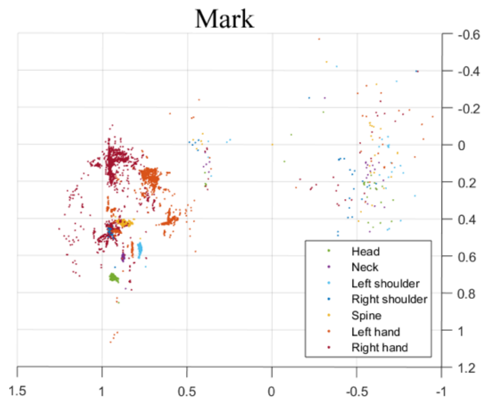


Figure 50. Mark's body joint position clusters plotted in meters.

See **Figure 50** for his body joint positions over all sessions. Mark's mean press count is 2.2360 presses per trial. His mean first response time is 4.8864 seconds. Mark's mean d' sensitivity level is 1.3208 and his mean B''_D bias is .1782 (conservative).

4.6.6 Rick's data

Rick is excellent at keeping focused on the task, and his distractions look like they are either coming from his left side (where I am seated) or downwards towards the floor. He seems to scan upwards and downwards more than horizontally (**Figure 51**). He also seems to like the animated reward because his vision is dwelling on the location of that animation even after it has finished. Rick is a great performer and sustains attention in this task. Will's average correct response rate is 8.11% (Hits and Correct Omissions) and his average incorrect response rate is 19.89% (False Alarms and Misses). His mean EEG attention feature value is .4398 and his mean body fidgeting value is 5.1245 mm/s.

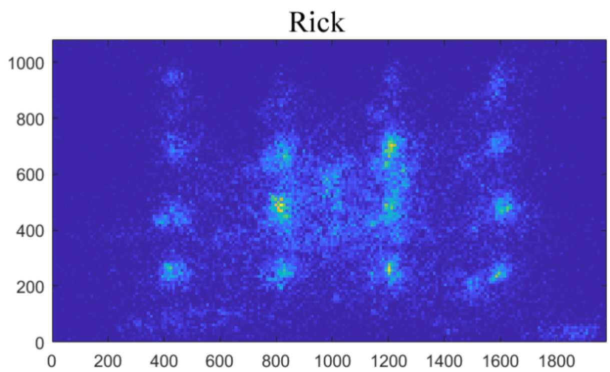


Figure 51. Rick eye-gaze pattern heat map plotted in pixels.

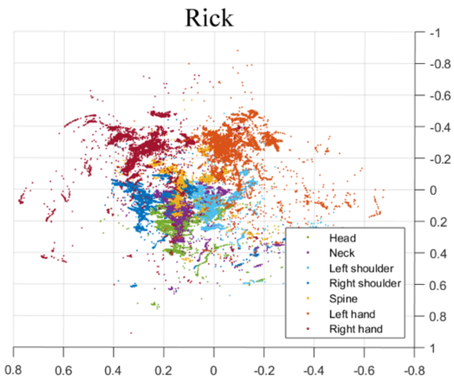


Figure 52. Rick's body joint position clusters plotted in meters.

See **Figure 52** for his body movements over all sessions. Rick's mean press count is 8.3150 presses per trial. His mean first response time is 1.7676 seconds. Rick's mean d' sensitivity level is 2.8982 and his mean B''_D bias is -.1866 (liberal).

4.6.7 Empirical lessons learned from the pilot study

Negative feedback loop

We found that the PMLD/CP participants can be easily encouraged with both negative and positive feedback. With consideration in retaining the message and purpose of feedback, the negative feedback volume level was reduced to the least interesting form of itself. The loud feedback sound for incorrect responses was found to encourage a negative feedback reward loop, so the feedback for incorrect responses was altered to a dull sounding thud. Having negative feedback sound as uninviting as possible was critical in avoiding a negative feedback loop where the negative feedback becomes the reward itself to incorrect responses.

However, removing the negative feedback entirely was seen to encourage three of our four participants to press the button in succession and sometimes rapidly. While these button presses are recorded in the background, the user is under the impression that the button has no effect in the interactive CPT game. This interaction-to-feedback relationship does not fairly or correctly model the user's world model and causes numerous excess presses, in response to them thinking that the system had not recorded their button press.

Paper-based SDT signal and noise ground truth test

CPT theory relies on signal detection and signal detection criteria require the user to be thoroughly familiar with the signal and noise (non-signal) objective. Therefore, there should be little or no learning involved in understanding and identifying the signal and response mechanism. This can add learning or skill bias to the data. In relation to the PMLD user base, the learning challenge of the experiment was tested and re-established before each session because users with PMLD may not have long-term learning ability or may have memory recall difficulty.

We learned that for the participants, there is a difference between recognizing the signal and acting on it. There is a three-step learning process; first defining the signal, second knowing that the correct response is the detection of the signal and third demonstrating 'pressing the button'. Vice versa for the noise trials, the user needs to recognize that there is no signal, knowing that the correct response here is to hold back on their response (pressing the button). In total that makes 5 individual learning tasks, in groups of 3 and 2 that need to chain together for the user to achieve learning independence in the CPT task.

Figure 53. A paper-based trial was used to pre-train participants in finding the signal. (1) introduces Wally to the participant, (2) Wally as a character says "Hi", (3) Bernard the red dog is introduced, (4) the red wool hat is introduced (5) Cookie monster is introduced, (6) practice SDT test is performed with the paper-based trial.

To remove the independent variable of skill and learning curve effect from our data paper-based trials, we establish the signal and activity framework before every week of the recording. In the paper-based trials seen in **Figure 53**, we have graduated the learning process that is required as part of the CPT. This was also conducted before every new session to ground truth the participant's understanding of the CPT signal, noise and the correct response.

Making the game familiar and fun

Traditional CPTs use letters as their signal noise trials [15], [143], [148]–[150], [157], [160], [165], [166], [168], [170]–[172], [189], [396], [397]. For example, on a signal trial, the letter X would appear on the screen while on a noise trial other letters of the alphabet are displayed. In this work, to encourage our age group to participate in the test and to continue participating over the numerous trials (11-19 sessions per participant, totaling 59 sessions), cartoon characters were used as signal representatives to reinforce a sense of familiarity.

It is essential for the SDT criteria for the signal to be familiar and clearly distinguishable for our participants. Critically, for the CPT to be valid, there should be no skill or learning effect in the signal detection aspect of the CPT. Studies have shown that object recognition is assisted when objects are named [398], [399]. For this reason, the main cartoon character representing the signal (in our case 'Wally') was given a name and character. The relationship between the signal 'Wally' and the user is strengthened by having 'Wally' as a character that relates to the participant by a greeting at the start of each session. Wally says "Hi" after he introduces himself to the user using a friendly sounding audio recording that is played at the beginning of every session. The introductory slides to each session precede the experiment itself and have the role of strengthening the participant's recollection with the game after the paper-based trial (**Figure 54**).

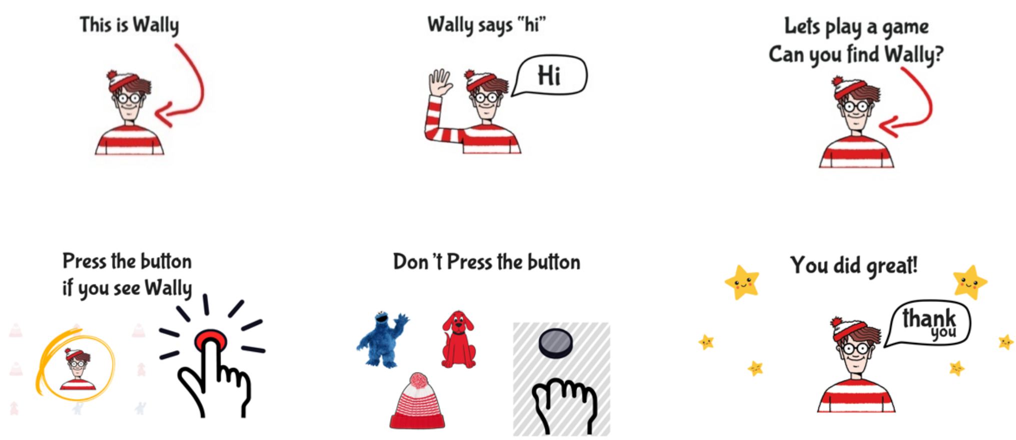


Figure 54. Introductory slides to the Seek-X type CPT create a fun and recognizable game environment for the participant.

By the 3rd week of the study we found that, occasionally, participants would start calling the other characters "Not Wally." If this was a task where the user was required to distinguish between 4 different signals, that would be counterproductive. However, in a signal detection test there is only one signal 'Wally', and everything else is non-signal "not Wally". This response reinforces the idea that the participants categorized the other characters as non-Targets by calling the other characters, not by their actual names, but simply naming them all "not Wally". This is important in establishing signal detection validity; our participants understand and recognize the correct signal when presented with it, and can distinguish that from the noise trials.

The following procedure was taken before every session using the paper-based test:

1. Introduce signal ("Wally").
1. Establish a personal connection with the signal ("Wally says Hi").
2. Introduce non-signal (introduce other 3 cartoon characters).
3. Establish signal difference (ask the participant if other characters are Wally).
4. Give positive affirmation that they are not the signal ("No they are not Wally").
5. Establish basics of signal detection activity with the user (tell the user to point to Wally in a 3 x 3 grid of characters – importantly this is reduced from the 4 x 4 grid that they will be performing in the actual computerized CPT). The positive feedback helps the participant establish that recognizing Wally is of significance.
6. Establish that absence of the signal with the participant.
7. Establish that the participant needs to act on this recognition with a button press.
8. Establish that the participant needs to control their impulsivity and response in the absence of the signal.
9. Validate that the participant understands the challenge, the response mechanism and can demonstrate a signal detection and distinction between Target and not Target images using the paper-based trials.

Unexpected participant performance

We found that the teacher's expectation, guided by the participant's capacity outside the study did not always have a direct relationship with the participant's capacity and performance in the study. Specifically, in one case (the participant Rick), it was expected that he would have difficulty responding to the CPT signal after the allocated display time of 2.8 seconds, so the time was increased to 10 seconds. By accident, and by using an alternative CPT profile for this user, we discovered that 2.8 seconds was actually sufficient for his processing time. This case is a significant argument for how expectations and assumptions are necessary, however, they do not necessarily map to the CPT and should be challenged during the initial pilot study.

Data retention during pre-maturely ended trials

In many instances the participant may have to leave the study prematurely, due to health reasons, schedule restrictions or a restroom visit. In this case, valuable data can be salvaged despite the experiment not being concluded. The experimental platform was adapted to be able to conclude before the 48 slides have been displayed so that no data loss occurred and all data was instantly saved to disk.

Chapter 5. CONCLUSIONS

The importance of learner engagement to obtain meaningful learning outcomes and deep learning was explored. We showed that teachers and carers face challenges in attributing learner affect, including engagement for PMLD/CP. This results in negative results and low-quality care for the PMLD/CP learner. To overcome this, there are many observer and paper-based methods to track engagement, however, they are not automated or rely on the subjective ratings of the observer. This leads to inter-rater reliability issues, which can cause unreliability and validity doubts over these approaches.

To address this, we conducted a literature review to establish the relationships between the affect states of learning and learner performance. Learner performance is related to learner affect in the ZPF diagram of learner affect states, through skill and challenge levels. In the ZPF theory, flow is the optimal state of engagement in a learning experience. Flow itself is a state where the learner's skill level is matched almost equally with the difficulty of the challenge they are engaged with. When the learner is under-challenged, they are in boredom and when they are over-challenged, they are in frustration. Only when the challenge level is delicately balanced with learner skill level, is the learner in flow. These are the affect states of learning, frustration, flow and boredom.

Studies have used participant performance in a game with discrete levels of difficulty to track the participant affect states and related it to academic achievement. In a way, these studies demonstrate that continuous monitoring of performance can be used to understand learner affect. Other studies have theorized the use of the Continuous Performance Test (CPT) as an objective way to label Multimodal Multisensor data in regard to modeling attention and affect state. CPT are established tests to track sustained and selective attention in studies.

In order to investigate these approaches further, **RQ1** asks:

RQ1: Can we create a model of attention for PMLD/CP students using the CPT?

In this work, we expand on the prior aforementioned research and develop a system that uses CPT outcomes to label Multimodal Multisensor data with labels of attention for PMLD/CP. We later develop machine learning models of attention using the Multimodal Multisensor data and the CPT outcomes. We then investigate each models' classification performance. Lastly, the correlations in the data points collected in the research are reviewed.

An approach to labeling Multimodal Multisensor data to train machine-learning algorithms to infer the attention of students with profound and multiple disabilities has been presented. We posit that this approach can overcome the variation in observer inter-rater reliability when using standardized scales in tracking the emotional expression of students with such profound disabilities.

In this work, 2615 samples, over the length of 59 sessions, were collected and classified into two categories ('attention' and 'inattention') using nine features. In total there were 2051 'Attention' samples and 564 'inattention'. Weighted class parameters were used in the machine learning algorithms to counter the unbalanced data set and machine learning performance measures insensitive to unbalanced data sets were used to compare the different models.

We found that the best classification model was achieved with random forest. The random forest model using both high- and low-level features achieved an 84.8% classification for 'attention' and a 65.4% accuracy for 'inattention'. This model has an 8.7% coverage of the AUC. The random forest performance improved NN performance by 5.5% better classifications of 'inattention' (TN, the minority class) and .3% better classifications of 'attention' (TP, the majority class). When the data was looked at per participant, TN classification (of 'inattention') improves by 16.87% compared to when all the data was put in a single group. The accuracy of our approach increases with multiple modes of sensor input. The sensory modes of eye-gaze, EEG and interaction data show the highest classification advantages in every sub-set of modes. We theorize that a system based on these models would be robust to sensor occlusion and fall-out as it could realign to any sub-set of data modes.

In **RQ2** we look in greater detail at these exact correlations, between the CPT outcomes and learner characteristics, including participant Multimodal Multisensor data, as well as their P scales.

RQ2: What are the main correlations found in the CPT outcomes and the Multimodal Multisensor data?

To address **RQ2** we examined the data collected longitudinally. We had 2615 samples collected over 59 sessions. Even though the sample count of 2615 may seem low, it is within the norms of prior studies, where Swanson collected 2240 SDT samples [15], and Goldberg *et al.* had a total of 240 SDT signal samples [166].

In total 55 separate data parameters were evaluated. These included SDT and CPT outcomes measures, Multimodal Multisensor features and participant characteristics. We investigate the correlations between the data points in the category of 'platform' against themselves and the data points from 'platform' with 'participant'. Over 2475 separate correlation tests were carried using Pearson's correlation coefficient r, N = 59. These results show significant correlations between SDT, CPT and participant characteristics. Importantly, a strong positive correlation was found between participant ability to maintain sustained and selective attention in the CPT to their academic progress in school (d'), $r(57) = .986$, $P = .0068$. Participants who were less impulsive and more selective in the test also did better in their academic performance $r(57) = .995$, $P = .0025$. The Seek-X type CPT also showed specific physiological characteristics, including body movement range and eye-gaze that were significant in P scales such as 'Reading' and 'Listening' $P < .05$.

Despite studies [1], [140], [143], [151] suggesting that children with learning difficulties are traditionally more conservative, we found that participant bias was overall liberal $\overline{B''_D} = -.1105$, $\sigma(B''_D) = .6323$. Participants showed no significant bias change during the sessions $r(57) = -.078$, $P = .7205$. While studies having varied reporting regarding the correlation between bias (B''_D) and sensitivity (d') [151], we found no significant correlation $r(57) = -.138$, $P = .8615$.

5.1 Summary and discussion

We demonstrated that the CPT can be used as part of a platform to objectively label Multimodal Multisensor data for students with PMLD/CP to infer 'attention' and 'inattention'. We also showed that a random forest model can reliably track the attention of students with profound disabilities and that our approach is robust to both sensor fallout and occlusions **(RQ1)**.

Correlations between SDT and CPT outcomes, participant physiological characteristics and academic scales were investigated. It seems possible that participants' ability to sustain attention has enabled them to progress at school. At the same time, a higher level of cognitive functioning may enable better-sustained attention as well as the ability to progress at school. Given these correlations, the CPT would appear an appropriate method to label Multimodal Multisensor data of students with PMLD/CP **(RQ2)** to assess sustained attention. The CPT could be utilized, in the future, to help to assess and monitor participant performance, and also to help understand the specific challenges (such as physiological and cognitive impairments and limitations) that may hinder their academic performance. A system incorporating these models can help teachers track attention in students using the most appropriate set of sensors for that individual student.

To summarize the outcomes of this research:

- Understanding learner affect is important to improve learner outcomes, especially for PMLD and CP students, where teachers and carers face challenges to obtain objective affect attribution.
- Performance outcomes are an indicator of affect state in the ZPF diagram of learning.
- CPT has been theorized as an objective method to label sensor data with performance outcomes.
- A methodology to create a model of performance outcomes using Multimodal Multisensor data (eye-gaze, body pose, EEG and interaction data) and CPT outcomes as labels is proposed.
- New forms of CPT, the Seek-X type CPT was developed.
- A novel Multimodal Multisensor data collection methodology for people with PMLD/CP has been developed.
- The CPT was used to collect 2615 samples from 59 CPT sessions with PMLD and CP participants.
- Random forest outperformed other methods of machine learning classification, including NN, SVM, Naïve Bayes, AdaBoost and k-NN, possibly because random forest provides consistent pairwise similarity by adding higher dimensionality to the feature space whilst being less sensitive to smaller sample sizes.
- Our random forest model achieved 84.8% correct classification of 'attention' and 65.4% correct classification of 'inattention', with an overall AUC coverage of 8.7%.
- Multimodal Multisensor data had superior classification results compared to any sub-set of sensors, for example using all four modes of sensors improves the AUC coverage by an average of 5.15% compared to when only three modes are used, 7.85% improvement when only 2 modes are used, and 24.45% improvement compared to when a single mode is used.

- In case of occlusion or loss of a single data mode, a subset of data can still give way to acceptable classification results, with eye-gaze, body pose, and interaction data achieving 78.1% coverage of the AUC.
- Some modes of data offer better value for classification purposes than other modes; eye-gaze, EEG, and interaction data carrying the most relevant data for attention classification, in that order.
- Hand-picked compound features of Attention and Inattention improved classification outcomes.
- Correlational data between participant characteristics, Multimodal Multisensor feature data and the CPT outcome measures were evaluated.
- Paper published in ICDRVAT 2018:
  M. Taheri *et al.*, "State Diagram for Affective Learning in an Educational Platform," pp. 4–6, 2018.
- Paper published in ICACII 2020:
  M. Taheri, D. Brown, and N. Sherkat, "Modeling Engagement with Multimodal Multisensor Data: The Continuous Performance Test as an Objective Tool to Track Flow," in ICACII 2020: 14. International Conference on Affective Computing and Intelligent Interaction, 2020, pp. 1–13.
- Paper submitted to ACM ICMI 2020:
  M. Taheri, D. Brown, and N. Sherkat, "Multimodal Multisensor Data Relationship to Learner Potential, Physiological Data Relationships with Learner Capacity" in ICMI 2020: 22. International

5.2 Future work and funded projects

The CPT provides a measure to continuously monitor the learner's attention state over a duration of time. This measure can be realized in an interactive game format, where the student's gaming performance (how quickly they respond to goals and how effectively they negotiate challenges) can be the new metric to track learner performance using sensor data. In this way, the gameplay itself can be the learning stage of the Multimodal Multisensor interactive system to create a user profile. The architecture for this system is presented in **Figure 55**. This user profile can then be deployed as an intervention in future levels of the game to provide a novel immersive experience to the game player. This model is a bridge between sensor data and participant attention that could later be used to assess learner attention, independent of the CPT, by using sensor data alone.

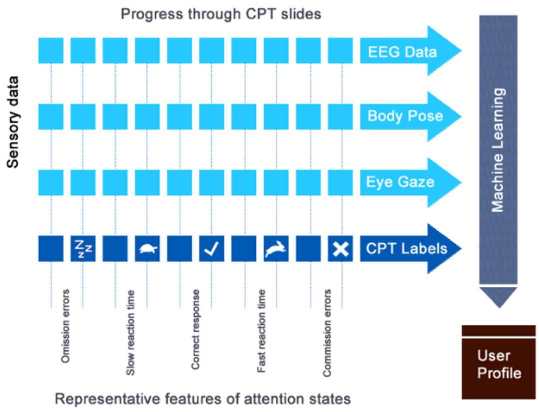


Figure 55. A learner profile constructed from a machine learning model trained on Multimodal Multisensor data.

The 'User Profile' is stored in Cloud Architecture – a personalized model of attention for that user. When the user is next exposed to a learning activity on a mobile platform, the sensor data is collected in real-time, and temporal frames of sensor information are fused at the data, feature and decision level and are used as the input of the probabilistic machine learning layer. The machine learning layer uses the user's temporal sensor data frame to conduct probabilistic attention level recognition based on the prior attention model (which is stored in the user profile). Temporal user attention levels with confidence ratings are returned in real-time.

With the increase of smartwatch use and the peripheral expansion using Bluetooth, a more extensive range of external sensors is now available to us (heart rate, heart rate variability, body movement). In the absence of a Kinect platform, a combination of smartphone and smartwatch sensors could be used to determine body pose. The data from the webcam could be used to determine facial cues for emotion and eye-gaze. This data can then be compressed and sent to the cloud for machine learning and profiling. The results can be sent back to the mobile interaction platform at the site to determine a suitable proactive intervention.

Assisted learning [61] [62] [63] is of particular interest to students with varying degrees of learning difficulty. An example of using affective computing in a special needs educational setting includes the application of an intelligent agent in assisted learning by monitoring the user's response to different learning routes by finding their optimal learning pathway. This has an impact because customizing the learning pathway could allow the system to compensate for user impairments and challenges and adapt the learning experience to suit the user's more receptive learning pathways (focusing on their abilities, not limitations). Affective personalized learning is shown to not only encourage participation and engagement in the classroom but also in extra-curricular clubs and work-related learning in the local community [49].

Research from this Ph.D. has formed the basis of two projects. These projects aim to use the methods and models developed in this work to create models of attention tracking for students.

- Pathway+ (Erasmus+ project ref: 2017-1-UK01-KA201-036761)
  https://pedagogics-pathway.eu/
- NTU Smart Campus (Internal project)

Appendix A. Information pack (Complete version)

Attention tracking using a combination of eye-gaze, head-posture and EEG

A research study for students who have learning difficulty is taking place in Nottingham. We would like to invite you to take part in this study. In order to help you decide whether to take part in this study, this sheet explains why the research is being done and what it would involve for you. Please take the time to read the following information carefully.

What is the purpose of this study?

It is often a challenge to keep students interested long enough in an activity. Especially if the activity requires the student to keep focus and active participation for a continuous-time. Research has shown that head-posture and eye-gaze can be a good measure for gauging the attention of a person. In this study, we will use the commercially available "Muse headband" EEG recording device to see if we can measure the relationship between eye-gaze, head-posture, and EEG in attention tracking.

What will I be expected to do if I take part?

If you choose to take part, we require you to sign a consent form. We will also check that it is appropriate for you to take part in the study.

The study involves participation in a continuous performance task (CPT) while the eye-gaze position, head-posture position, and EEG measurements are made in the background. The CPT involves two stages. This activity is designed to assess the average duration of sustained and selective attention of the participants.

The first CPT stage will involve a popular cartoon character appearing on the screen at different intervals. The participant is asked to press a button whenever they see that character and disregard any other activity on the screen or around them. The second task substitutes the character with the participant's name. The participant is asked to press the button whenever they see their name and disregard the other names or symbols appearing on the screen.

When performing the exercise, the participant will be watching a computer screen that will display the cartoon character or names in intervals. The participant will have a button to press when the specific cartoon character or name appears. The test should take no longer than 10 minutes and with preparation no longer than 25 minutes. Although in some circumstances it may take longer. This will consist of breaks or rests as needed.

We anticipate up to four sessions each week for three months. A maximum of 20 sessions will be recorded.

What are the potential benefits of taking part?

There may be no direct benefit to you. However, we will monitor the performance of your participation in the CPT throughout the experiment and give you feedback on the results. The information we get from the study should help us to determine if features like eye-gaze, head-posture and EEG are good tools for measuring attention. The data will also allow us to discover any correlations between these features. This is a new broadening science, knowledge regarding how we can access sustained and selection attention from external features gives us a fantastic opportunity. One possibility could be how one or all of these tools and knowledge could be incorporated into an educational scope for developing dynamic and personalized lessons. This lesson could mitigate attention dropping moments with a dynamic intervention designed to revive the students' attention level.

Will it cost me anything to take part?

It will not cost you anything to take part. All costs will be paid for by the research.

Will my taking part in the study be kept confidential?

Yes. We will follow established ethical and legal practices, and all information about you will be handled confidentially. All information that is collected about you during the study will be kept strictly confidential, and any information about you will have your name and address removed so you cannot be identified.

What data will be collected?

No long term personally identifiable information will be collected for this study. Hardware and software designed only to read eye-gaze and head-posture will be used (Microsoft Kinect, Tobii EyeX). A webcam video record will be temporarily used to validate this data. The eye-gaze and head-posture data will be recorded as numeric data (eye direction, head tilt, and rotation). This will be obtained from automated computer algorithms designed to recognize these features. The video recording will be used to manually validate the information of the head and eye-gaze parameters computed by the automated algorithm. As soon as the eye-gaze and head-posture data are validated, the video data will be destroyed. This will be no longer than one month after it was initially recorded. Your consent will be requested for all the data collected in the experiment. For publication no video imagery will be used, all other data will be anonymized, and no records of name or data leading to the identification of the participants will ever be made public. At no point will any imagery be stored for long-term or public publication.

What will happen if I don't want to carry on with the study?

Your participation is voluntary, and you are free to withdraw at any time, without giving any reason and without your legal rights being affected. If you withdraw the information collected by the computer system will still be used in the project analysis and all information will follow the normal anonymization process regardless.

What will happen to the results of the research?

It is intended that the results of the research will be published formally in scientific journals and published in newsletters. You will not be identified in any report or publication.

Our genuine thanks for your time in reading this information pack.

Contact Information
Mohammad Taheri
Tel: 07411227118
E-mail: mohammad.taheri@ntu.ac.uk
Nottingham Trent University
Ph.D. Candidate in Brain-Computer Interaction

NOTTINGHAM TRENT UNIVERSITY

Appendix B. Information pack (Easy read version)

Attention tracking using a combination of eye-gaze, head-posture and EEG.

What is the purpose of this study?

You may have noticed it is often a challenge to keep students interested long enough in an activity. This is especially true if this activity requires the student to keep focus and active participation for a continuous-time. Previous research has shown that head-posture and eye-gaze can be used as determents for the attention of a person. We want to see if we can understand this better.

This study will use a commercially available meditating tool called the "Muse headband" to wirelessly monitor brain waves and see if we can measure the relationship between eye-gaze, head-posture, and EEG in an attention tracking activity.

What will I be expected to do if I take part?

If you choose to take part, we require you to sign a consent form.

The participant is asked to press a button whenever they see a specific object of interest on the computer screen — this shows that they are attentive — some time passes and the object appears on the screen again, this measures their sustained attention. The object that appears on the screen is personalized to their participant's likes and interests (for example pizza, skating, a Formula-1 car or a TV show personality…).

The test should take no longer than 10 minutes and with preparation no longer than 25 minutes. This will consist of breaks or rests as needed. We anticipate up to four sessions each week for three months. A total of 20 sessions will be recorded.

What are the potential benefits of taking part?

We will monitor the performance of your participation and give you feedback on the results. Knowledge regarding how we can understand sustained and selection attention from external features like eye-gaze, head-posture and EEG data from commercially available headbands gives us an amazing opportunity. One possibility could be how one or all of these tools and knowledge could be incorporated into an educational scope for developing dynamic and personalized lessons. This lesson could mitigate attention dropping moments with a dynamic intervention designed to revive the students' attention level.

Will it cost me anything to take part?

All costs will be paid for by the research. It will not cost you anything to take part.

Will my taking part in the study be kept confidential? What data will be collected?

Yes. Your participation is made confidential and your consent will be requested for all the data collected in the experiment. At no point will any imagery be stored for long term or public publication. All information is anonymized and your name and address are removed so you cannot be identified. The eye-gaze and head-posture data will be recorded as numeric data (eye direction, head tilt and rotation) using the Microsoft Kinect and Tobii EyeX. A webcam video record will be temporarily used to validate the information of the head and eye-gaze information; it will be destroyed from one month after it was recorded. For publication no video imagery will be used, no records of name or data leading to the identification of the participants will ever be made public.

What will happen if I don't want to carry on with the study?

Your participation is voluntary, and you are free to withdraw at any time, without giving any reason and without your legal rights being affected.

What will happen to the results of the research?

It is intended that the results of the research will be published formally in scientific journals and published in patient newsletters. You will not be identified in any report or publication.

Our genuine thanks for your time in reading this information pack.

<u>Contact Information</u>

Mohammad Taheri
Tel: 07411227118
E-mail: mohammad.taheri@ntu.ac.uk
Nottingham Trent University
Ph.D. Candidate in Brain-Computer Interaction

NOTTINGHAM
TRENT UNIVERSITY

Appendix C. Consent form

I, the undersigned, confirm that (please tick the box as appropriate):

1.	I have read and understood the information about the project, as provided in the Information Sheet.	☐
2.	I have been given the opportunity to ask questions about the project and my participation.	☐
3.	I voluntarily agree to participate in the project.	☐
4.	I understand I can withdraw at any time without giving reasons and that I will not be penalized for withdrawing nor will I be questioned on why I have withdrawn.	☐
5.	The procedures regarding confidentiality have been clearly explained (e.g., use of names, pseudonyms, anonymization of data, etc.) to me.	☐
6.	If applicable, separate terms of consent for interviews, audio, video or other forms of data collection have been explained and provided to me.	☐
7.	The use of the data in research, publications, sharing and archiving has been explained to me.	☐
8.	I understand that other researchers will have access to this data only if they agree to preserve the confidentiality of the data and if they agree to the terms I have specified in this form.	☐
9.	I, along with the Researcher, agree to sign and date this informed consent form.	☐

Participant / Parent (if under 16 years of age):

Name of Participant ________________ Signature ________________ Date ________________

Appendix D. Ethical issues

Will the participant feel overly exerted from the continuous performance task?

Participants will be informed at each stage that they can withdraw from the study at any time without needing to give a reason. The participants will be monitored and asked if they want to postpone the exercise to a later time. The continuous performance task requires the participant to keep selective and sustained attention for 5 minutes at a time. This can be tiring for the participant, as this task is naturally a repetitive one. Even though repetition is a requirement of the CPT task, steps can be taken to minimize the tiring aspect of the task. First, a genuine connection with the participant showed us to be beneficial in encouraging the engagement of the participant. In our previous study, we observed that the participants are more interested in a dialog where the experiment organizer also takes a genuine interest in their well-being and current events in their daily life. Secondly, each day the task is performed with a new cartoon character, and so the participant will focus their attention on a new subject. This 'reveal' is designed to be exciting and create anticipation for the next experiment day. Choosing cartoon characters instead of the conventional single character which is used in the classic CPT experiment is also designed to engage the participants more and make the task more fun. The tiring aspect of CPT experiments not only affects the quality of experience of the participants but can also negatively affect the data. All sessions will be performed under the knowledge or observation of a students' care-worker.

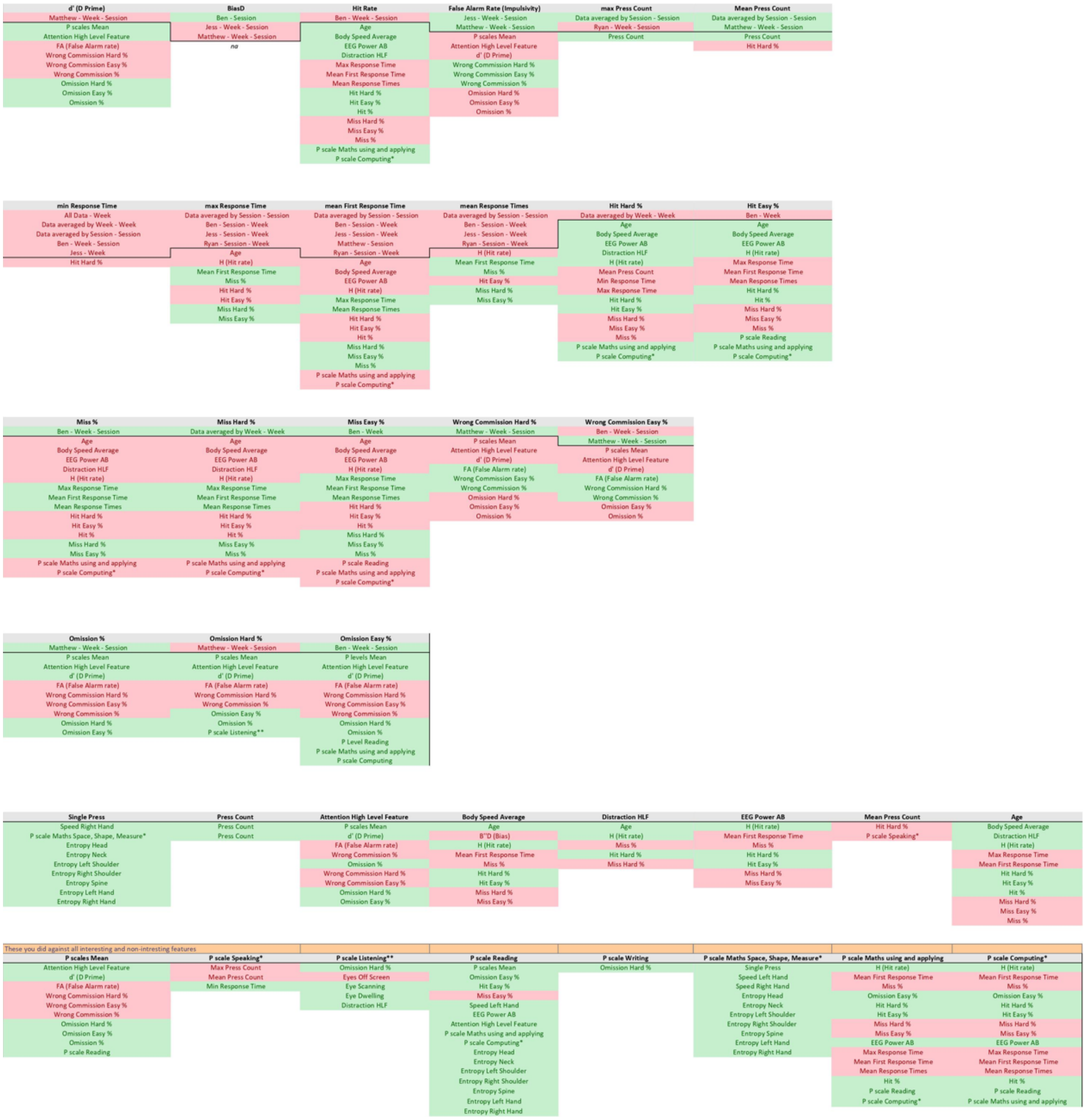
What data will be collected and how will this be managed?

No long term personally identifiable information will be collected for this study. Hardware and software designed only to read eye-gaze and head-posture will be used (Microsoft Kinect, Tobii EyeX). A webcam video record will be temporarily used to validate this data. The eye-gaze and head-posture data will be recorded as numeric data (eye direction, head tilt, and rotation). This will be obtained from automated computer algorithms designed to recognize these features. The video recording will be used to manually validate the information of the head and eye-gaze parameters computed by the automated algorithm. As soon as the eye-gaze and head-posture data are validated, the video data will be destroyed. This will be no longer than one month after it was initially recorded. Your consent will be requested for all the data collected in the experiment. For publication no video imagery will be used, all other data will be anonymized, and no records of name or data leading to the identification of the participants will ever be made public. At no point will any imagery be stored for long-term or public publication.

Mohammad Taheri
E-mail: mohammad.taheri@my.ntu.ac.uk
Nottingham Trent University
Ph.D. Candidate in Brain-Computer Interaction

Appendix E. Correlations breakdown

An instant overview of all the significant cross-correlations found in the MM system is shown in **Figure 56**.



d' (D Prime)	BiasD	Hit Rate	False Alarm Rate (Impulsivity)	max Press Count	Mean Press Count
Matthew - Week - Session	Ben - Session	Ben - Week - Session	Jess - Week - Session	Data averaged by Session - Session	Data averaged by Session - Session
P scales Mean	Jess - Week - Session	Age	Matthew - Week - Session	Ryan - Week - Session	Matthew - Week - Session
Attention High Level Feature	Matthew - Week - Session	Body Speed Average	P scales Mean	Press Count	Press Count
FA (False Alarm rate)	na	EEG Power AB	Attention High Level Feature		Hit Hard %
Wrong Commission Hard %		Distraction HLF	d' (D Prime)		
Wrong Commission Easy %		Max Response Time	Wrong Commission Hard %		
Wrong Commission %		Mean First Response Time	Wrong Commission Easy %		
Omission Hard %		Mean Response Times	Wrong Commission %		
Omission Easy %		Hit Hard %	Omission Hard %		
Omission %		Hit Easy %	Omission Easy %		
		Hit %	Omission %		
		Miss Hard %			
		Miss Easy %			
		Miss %			
		P scale Maths using and applying			
		P scale Computing*			

min Response Time	max Response Time	mean First Response Time	mean Response Times	Hit Hard %	Hit Easy %
All Data - Week	Data averaged by Session - Session	Data averaged by Session - Session	Data averaged by Session - Session	Data averaged by Week - Week	Ben - Week
Data averaged by Week - Week	Ben - Session - Week	Ben - Session - Week	Ben - Session - Week	Age	Age
Data averaged by Session - Session	Jess - Session - Week	Jess - Session - Week	Jess - Session - Week	Body Speed Average	Body Speed Average
Ben - Week - Session	Ryan - Session - Week	Matthew - Session	Ryan - Session - Week	EEG Power AB	EEG Power AB
Jess - Week	Age	Ryan - Session - Week	H (Hit rate)	Distraction HLF	H (Hit rate)
Hit Hard %	H (Hit rate)	Age	Mean First Response Time	H (Hit rate)	Max Response Time
	Mean First Response Time	Body Speed Average	Miss %	Mean Press Count	Mean First Response Time
	Miss %	EEG Power AB	Hit Easy %	Min Response Time	Mean Response Times
	Hit Hard %	H (Hit rate)	Miss Hard %	Max Response Time	Hit Hard %
	Hit Easy %	Max Response Time	Miss Easy %	Hit Hard %	Hit %
	Miss Hard %	Mean Response Times		Hit Easy %	Miss Hard %
	Miss Easy %	Hit Hard %		Miss Hard %	Miss Easy %
		Hit Easy %		Miss Easy %	Miss %
		Hit %		Miss %	P scale Reading
		Miss Hard %		P scale Maths using and applying	P scale Maths using and applying
		Miss Easy %		P scale Computing*	P scale Computing*
		Miss %			
		P scale Maths using and applying			
		P scale Computing*			

Miss %	Miss Hard %	Miss Easy %	Wrong Commission Hard %	Wrong Commission Easy %
Ben - Week - Session	Data averaged by Week - Week	Ben - Week	Matthew - Week - Session	Ben - Week - Session
Age	Age	Age	P scales Mean	Matthew - Week - Session
Body Speed Average	Body Speed Average	Body Speed Average	Attention High Level Feature	P scales Mean
EEG Power AB	EEG Power AB	EEG Power AB	d' (D Prime)	Attention High Level Feature
Distraction HLF	Distraction HLF	H (Hit rate)	FA (False Alarm rate)	d' (D Prime)
H (Hit rate)	H (Hit rate)	Max Response Time	Wrong Commission Easy %	FA (False Alarm rate)
Max Response Time	Max Response Time	Mean First Response Time	Wrong Commission %	Wrong Commission Hard %
Mean First Response Time	Mean First Response Time	Mean Response Times	Omission Hard %	Wrong Commission %
Mean Response Times	Mean Response Times	Hit Hard %	Omission Easy %	Omission Easy %
Hit Hard %	Hit Hard %	Hit Easy %	Omission %	Omission %
Hit Easy %	Hit Easy %	Hit %		
Hit %	Hit %	Miss Hard %		
Miss Hard %	Miss Easy %	Miss Easy %		
Miss Easy %	Miss %	Miss %		
P scale Maths using and applying	P scale Maths using and applying	P scale Reading		
P scale Computing*	P scale Computing*	P scale Maths using and applying		
		P scale Computing*		

Omission %	Omission Hard %	Omission Easy %
Matthew - Week - Session	Matthew - Week - Session	Ben - Week - Session
P scales Mean	P scales Mean	P levels Mean
Attention High Level Feature	Attention High Level Feature	Attention High Level Feature
d' (D Prime)	d' (D Prime)	d' (D Prime)
FA (False Alarm rate)	FA (False Alarm rate)	FA (False Alarm rate)
Wrong Commission Hard %	Wrong Commission Hard %	Wrong Commission Hard %
Wrong Commission Easy %	Wrong Commission %	Wrong Commission Easy %
Wrong Commission %	Omission Easy %	Wrong Commission %
Omission Hard %	Omission %	Omission Hard %
Omission Easy %	P scale Listening**	Omission %
		P Level Reading
		P scale Maths using and applying
		P scale Computing

Single Press	Press Count	Attention High Level Feature	Body Speed Average	Distraction HLF	EEG Power AB	Mean Press Count	Age
Speed Right Hand	Press Count	P scales Mean	Age	Age	H (Hit rate)	Hit Hard %	Body Speed Average
P scale Maths Space, Shape, Measure*	Press Count	d' (D Prime)	B''D (Bias)	H (Hit rate)	Mean First Response Time	P scale Speaking*	Distraction HLF
Entropy Head		FA (False Alarm rate)	H (Hit rate)	Miss %	Miss %		H (Hit rate)
Entropy Neck		Wrong Commission %	Mean First Response Time	Hit Hard %	Hit Hard %		Max Response Time
Entropy Left Shoulder		Omission %	Miss %	Miss Hard %	Hit Easy %		Mean First Response Time
Entropy Right Shoulder		Wrong Commission Hard %	Hit Hard %		Miss Hard %		Hit Hard %
Entropy Spine		Wrong Commission Easy %	Hit Easy %		Miss Easy %		Hit Easy %
Entropy Left Hand		Omission Hard %	Miss Hard %				Hit %
Entropy Right Hand		Omission Easy %	Miss Easy %				Miss Hard %
							Miss Easy %
							Miss %

These you did against all interesting and non-intresting features

P scales Mean	P scale Speaking*	P scale Listening**	P scale Reading	P scale Writing	P scale Maths Space, Shape, Measure*	P scale Maths using and applying	P scale Computing*
Attention High Level Feature	Max Press Count	Omission Hard %	P scales Mean	Omission Hard %	Single Press	H (Hit rate)	H (Hit rate)
d' (D Prime)	Mean Press Count	Eyes Off Screen	Omission Easy %		Speed Left Hand	Mean First Response Time	Mean First Response Time
FA (False Alarm rate)	Min Response Time	Eye Scanning	Hit Easy %		Speed Right Hand	Miss %	Miss %
Wrong Commission Hard %		Eye Dwelling	Miss Easy %		Entropy Head	Omission Easy %	Omission Easy %
Wrong Commission Easy %		Distraction HLF	Speed Left Hand		Entropy Neck	Hit Hard %	Hit Hard %
Wrong Commission %			EEG Power AB		Entropy Left Shoulder	Hit Easy %	Hit Easy %
Omission Hard %			Attention High Level Feature		Entropy Right Shoulder	Miss Hard %	Miss Hard %
Omission Easy %			P scale Maths using and applying		Entropy Spine	Miss Easy %	Miss Easy %
Omission %			P scale Computing*		Entropy Left Hand	EEG Power AB	EEG Power AB
P scale Reading			Entropy Head		Entropy Right Hand	Max Response Time	Max Response Time
			Entropy Neck			Mean First Response Time	Mean First Response Time
			Entropy Left Shoulder			Mean Response Times	Mean Response Times
			Entropy Right Shoulder			Hit %	Hit %
			Entropy Spine			P scale Reading	P scale Reading
			Entropy Left Hand			P scale Computing*	P scale Maths using and applying
			Entropy Right Hand				

Figure 56. The correlation outcomes are shown in 36 independent tables. Green color shows a significant positive correlation and red color shows a significant negative correlation. All correlations are r(57) and at least P < .05.

d' (D Prime)
Mark - Week - Session
P scales Mean
High-Level Attention feature
FAR (False Alarm Rate)
Wrong Commission Hard %
Wrong Commission Easy %
Wrong Commission %
Omission Hard %
Omission Easy %
Omission %

Mark is the only participant with a reduced d' (D Prime) sensitivity level. His reduced every week and with every session. A student's relative progress in the school (P scales) had a direct correlation with their average sensitivity amount. The hand-crafted High-Level Attention feature (HLA) had a direct correlation with sensitivity. Sensitivity had an inverse correlation with False Alarm Rate (FAR) and Wrong Commissions which are a sign of impulsivity. Higher Correct Omissions are directly correlated with sensitivity level, so participants with higher sensitivity levels were better at holding back button presses on Imitation targets.

Bias B''_D
Will - Session
Jen - Week - Session
Mark - Week - Session

Note:
Negative bias is liberal, more likely to call a signal a Target .
Positive bias is more conservative and less likely to call a signal a Target .

Despite Will having a negative bias overall on average (liberal), Will became more conservative as sessions progressed, he was less likely to press the button with every session. This was despite both Jen and Mark becoming more liberal per week and session with their responses, they both pressed the button more frequently. Rick had no change in his behavior, in regards to his bias.

Hit rate
Will - Week - Session
Age
Body Speed Average
EEG Power AB
High-Level Inattention feature
Max Response Time
Mean First Response Time
Mean Response times
Hit Hard %
Hit Easy %
Hit %
Miss Hard %
Miss Easy %
Miss %
P scales Maths using and applying
P scales Computing

The Hit rate is the number of targets hit divided by all targets. Will's Hit rate (H) reduced weekly and with every session, but his False Alarm Rate did not reduce significantly so he did not become more precise. Participant Hit rate had a direct correlation with Age, average body speed, EEG Power and also the High-Level Inattention (HLI) feature. Participants had a better Hit rate when they responded to challenges quicker. Participants who had more progress in 'P scales Maths using and applying' or 'P scales Computing' had a higher Hit rate.

False Alarm Rate FAR (Impulsivity)
Jen - Week - Session
Mark - Week - Session
P scales Mean
High-Level Attention feature
d' (D Prime)
Wrong Commission Hard %
Wrong Commission Easy %
Wrong Commission %
Omission Hard %
Omission Easy %
Omission %

False Alarms or Wrong Commissions are a measure of impulsivity. Jen and Mark both increased their False Alarm Rates with every session and every week. Participants that had progressed higher in their education had lower False Alarm Rates. A participant's sensitivity value *d'* and the High-Level Attention Feature had an inverse relationship with False Alarm Rate. Expectedly, False Alarm Rate, had an inverse relationship with Correct Omissions.

Wrong Commission Hard %
Mark - Week - Session
P scales Mean
High-Level Attention feature
d' (D Prime)
FAR (False Alarm Rate)
Wrong Commission Easy %
Wrong Commission %
Omission Hard %
Omission Easy %
Omission %

Mark was the only participant that had more wrong Hard Target Commissions and overall Target Commissions with every session and every week. Despite both Mark and Jen having more Hard and Easy Wrong Commissions overall, with every session and every week. Just like overall False Alarm Rates, participants that had progressed higher in their education had lower Hard Target False Alarm Rates. Also, just like overall False Alarm Rates, for Hard Target False Alarms, High-Level Attention feature and the participant's sensitivity value *d'* had an inverse relationship with Hard Target False Alarms. As expected again, just like overall False Alarm Rates, Wrong Commission for Hard Target was also inversely related to Correct Omissions.

Wrong Commission Easy %
Will - Week - Session
Mark - Week - Session
P scales Mean
High-Level Attention feature
d' (D Prime)
FAR (False Alarm Rate)
Wrong Commission Hard %
Wrong Commission %
Omission Easy %
Omission %

Mark was the only participant that had more overall, Hard and Easy wrong Hard Target Commissions with every session and every week. Will improved his impulsivity and was the only participant that had less Easy Target False Alarms with every session and every week. Just like overall False Alarm Rate and Hard Target False Alarm Rates, participants that had progressed higher in their education had lower Easy Target False Alarm Rates. Also, just like overall and Hard Target False Alarm Rates, for Easy Target False Alarms, participants with a lower High-Level Attention feature or a lower sensitivity value *d'*, had more overall, Hard and Easy False Alarms. As expected again, just like overall False Alarm Rates, and Wrong Commissions for Hard Target, Wrong Commission for Easy Target was also inversely related to Easy Target Correct Omissions and overall Correct Omissions.

max Press Count
Data averaged by Session - Session
Rick - Week - Session
Press Count

The average max press count of all participants increased significantly with every session. However, Rick was the only participant that his max press count decreased with every session. Participants' average press count had a direct correlation with their maximum press count.

Mean Press Count
Data averaged by Session - Session
Mark - Week - Session
Press Count
Hit Hard %

The mean press count for all participants' data had increased with every session. However, Mark was the only participant that his average press count had individually increased as well. The higher the average press count the lower the Hit rate of the participants.

min Response Time
All Data - Week
Data averaged by Week - Week
Data averaged by Session - Session
Will - Week - Session
Jen - Week
Hit Hard %

Fastest response time became faster with every week and every session for all participants. However, Will and Jen were the only two participants whom their fastest response time reduced with every week, Will's fastest response time also reduced with every session. Fast response times had an inverse correlation with the Hit rate.

max Response Time
Data averaged by Session - Session
Will - Session - Week
Jen - Session - Week
Rick - Session - Week
Age
H (Hit rate)
Mean First Response Time
Miss %
Hit Hard %
Hit Easy %
Miss Hard %
Miss Easy %

Longest response time reduced for all participants on average. All participants bar Mark had a significant reduction in their max response time with every session and week. Older participants had a lower max response time. Longer max response times had a direct correlation with missing targets.

mean First Response Time
Data averaged by Session - Session
Will - Session - Week
Jen - Session - Week
Mark - Session
Rick - Session - Week
Age
Body Speed Average
EEG Power AB
H (Hit rate)
Max Response Time
Mean Response times
Hit Hard %
Hit Easy %
Hit %
Miss Hard %
Miss Easy %
Miss %
P scales Maths using and applying
P scales Computing

The mean first response time decreased with every session for all participants. The older the participant the faster their mean first response times. Also, faster mean first response times were negatively correlated with Body Speed, EEG Power AB, Hit rate, 'P scales Maths using and applying' and 'P scales Computing'. Participants that were quicker to respond on average, had more Misses (Wrong Omissions) for all types of targets.

mean Response Times
Data averaged by Session - Session
Will - Session - Week
Jen - Session - Week
Rick - Session - Week
H (Hit rate)
Mean First Response Time
Miss %
Hit Easy %
Miss Hard %
Miss Easy %

Average participant data shows overall that mean response times reduced every session. Will, Jen and Rick's data mean response times became faster with every session and with every week. Taking longer to respond did not correlate with errors in harder targets but it correlated with more Misses on Easy Target.

Hit Hard %
Data averaged by Week - Week
Age
Body Speed Average
EEG Power AB
High-Level Inattention feature
H (Hit rate)
Mean Press Count
Min Response Time
Max Response Time
Hit Hard %
Hit Easy %
Miss Hard %
Miss Easy %
Miss %
P scales Maths using and applying
P scales Computing

Correct detection of Hard Target reduced week by week. Since the signal detection value (*d′*) had not decreased week by week, and participant bias did not show more conservative behavior (for most participants except for Will), this decrease did not influence participant signal detection *d′* and was not influenced by a change in their bias. Participants with a greater

Age, Average Body Speed, EEG Power AB, had a higher Hit rate. The High-Level Inattention feature correlated directly with more hits however as it doesn't correlate with a higher sensitivity level, we can conclude that higher High-Level Inattention feature leads to higher impulsivity. More press counts correlate to a lower Hard Hit rate, meaning that participants that repetitively pressed the button did not actually improve their sensitivity to harder targets. However, higher press counts don't significantly reduce their ability to distinguish easier targets. The fastest responses do not lead to better Hard Target detections, also long responses do not lead to better Hard Target distinction. Participants who had progressed more in 'P scales Maths using and applying' or 'P scales Computing' had a higher Hard Target Hit rate.

Hit Easy %
Will - Week
Age
Body Speed Average
EEG Power AB
H (Hit rate)
Max Response Time
Mean First Response Time
Mean Response times
Hit Hard %
Hit %
Miss Hard %
Miss Easy %
Miss %
P scales Reading
P scales Maths using and applying
P scales Computing

Will was the only participant who had a significantly lower Hit rate for Easy Target with every week. As with Hard Target, a higher Hit rate for Easy Target was correlated with Age, Body Speed Average, EEG band Power AB. Like Hard Target, Easy Target also correlate inversely with Max Response times, meaning that the sessions which had the longest response times had fewer Target hits, easy or hard. However, unlike Hard Target, Easy Target were not missed more if the participant makes more button presses or if they respond quickly. Participants that had a higher Hit rate for Hard Target were able to get more Easy Target. Like Hard Target, detecting Easy Target was correlated with the participant's progress in 'P scales Maths using and applying' and 'P scales Computing', but unlike Hard Target, it also correlates with 'P scales Reading'.

Miss %
Will - Week - Session
Age
Body Speed Average
EEG Power AB
High-Level Inattention feature
H (Hit rate)
Max Response Time
Mean First Response Time
Mean Response times
Hit Hard %
Hit Easy %
Hit %
Miss Hard %
Miss Easy %
P scales Maths using and applying
P scales Computing

Misses or Omission errors are an indicator of inattention. With every session and week, Will's ability to detect Easy Target reduced, he was also the only participant that had a significant decrease in his Hit rate. This in turn, also significantly increased Will's overall Misses with every session and week, meaning that after every session and every week, Will was the only participant who had reduced attention ability. Will was also the only participant that after 15 sessions and 6 weeks of data collection, withdrew from further participation on the last data collection day. This could be an indicator that Will's interest in the game reduced, which in turn made him more inattentive and triggered his request to exit the process. Older participants had fewer overall Misses. Participants who had more average body speed had a lower Miss percentage and a higher Hit rate. Participants with lower EEG Power AB values had a higher Miss %. When comparing *d'*, Hit rate and Miss % against the two HLCFs (High-Level Compound Features) we can see that *d'* had a direct relationship with HLA, and the HLI feature had a direct relationship with higher Commissions which led to a higher Hit rate and Miss % and HLA had a direct relationship with higher *d'* sensitivity values. Miss % reduced with greater Age, Higher EEG Power AB and Higher High-Level Inattention feature value. Just like overall Miss %, participants with higher Max Response Times, longer Mean First Response Times or over higher Mean Response Times, made greater Hard Target Misses. Participants who made more progress in their 'P scales Maths and applying' or 'Computing' had a lower overall Miss %.

Miss Hard %
Data averaged by Week - Week
Age
Body Speed Average
EEG Power AB
High-Level Inattention feature
H (Hit rate)
Max Response Time
Mean First Response Time
Mean Response times
Hit Hard %
Hit Easy %
Hit %
Miss Easy %
Miss %
P scales Maths using and applying
P scales Computing

Even though only Will's overall Miss % was increased with every session and week, all participants had increased Miss % for Hard Target every week. Just like the overall Miss %, the Hard Miss % was also less with greater Age, higher body speeds, higher EEG Power AB and higher High-Level Inattention feature value. Just like overall Miss %, participants with higher Max Response Times, longer Mean First Response Times or over higher Mean Response Times, made greater Hard Target Misses. Just like the overall Miss %, participants who made more progress in their 'P scales Maths and applying' or 'Computing' had a lower overall Hard Target Miss %.

Miss Easy %
Will – Week
Age
Body Speed Average
EEG Power AB
H (Hit rate)
Max Response Time
Mean First Response Time
Mean Response times
Hit Hard %
Hit Easy %
Hit %
Miss Hard %
Miss Easy %
Miss %
P scales Reading
P scales Maths using and applying
P scales Computing

Like overall Miss %, Will's Easy Target Miss % had also increased with every week but not by every session. Just like overall Miss % and Hard Miss %, Easy Miss % also reduces with greater Age, higher body speeds, and higher EEG Power AB. Unlike Hard Target Misses and overall Miss %, a higher High-Level Inattention feature does not lead to less Easy Target Misses, this means that when the High-Level Inattention feature was high, participants were more likely to be more impulsive (or less exact) when it came to Hard Target, not Easy ones. This means that High-Level Inattention feature was a good indicator for early indications of impulsivity, fatigue and less exact responses. Just like the overall Miss % and Hard Target Miss %, participants who made more progress in their 'P scales Maths and applying' and 'Computing' had a lower overall Easy Target Miss %, however in the case of Easy Target Miss %, participants with greater 'P scales Reading' progress had lower Easy Target Miss %.

Omission %
Mark - Week - Session
P scales Mean
High-Level Attention feature
d′ (D Prime)
FAR (False Alarm Rate)
Wrong Commission Hard %
Wrong Commission Easy %
Wrong Commission %
Omission Hard %
Omission Easy %

Mark was the only participant who consistently had higher Correct Omissions with every session and after every week. Higher P scales, participants with higher High-Level Attention feature values or better sensitivity *d′* outcomes consistently had higher Correct Omission percentages. This could mean that participants that had the ability to make better distinctions, were significantly better at determining that the Target was not there. High Correct Omissions also had a negative relationship with False Alarm Rate which was an indicator of impulsivity. In conclusion, participants with higher Correct Omissions were less impulsive and also had fewer Wrong Commissions. This also means that they were far less impulsive, hence had a significantly lower False Alarm Rate. To conclude attention was inversely correlated to impulsivity.

Omission Hard %
Mark - Week - Session
P scales Mean
High-Level Attention feature
d′ (D Prime)
FAR (False Alarm Rate)
Wrong Commission Hard %
Wrong Commission %
Omission Easy %
Omission %
P scales Listening

Despite Mark's overall Correct Omissions significantly increasing with every session and week, his Hard Target Correct Omissions decreased significantly after every session and after every week. This could be because Hard Imitation slides became more and more straining after every session and week, while the Easy Imitation slides did not become significantly more challenging over time. This shows that Correct Omissions in Hard and Easy Target do not necessarily follow suit. Just as in overall Correct Omissions, student P scales and their sensitivity *d′* ability had a direct correlation with Hard Target Omissions, as discussed in the overall Omissions' analysis, this could be because participants with higher sensitivity *d′* ability or P scales, were more able to distinguish between Target and Imitation slides. Also, it had been shown that the High-Level Attention feature had a direct correlation with Hard Target Correct Omissions. Unlike overall Omissions, Hard Target Omissions were easier for participants with higher 'P scales Listening'.

Omission Easy %
Will - Week - Session
P scales Mean
High-Level Attention feature
d′ (D Prime)
FAR (False Alarm Rate)
Wrong Commission Hard %
Wrong Commission Easy %
Wrong Commission %
Omission Hard %
Omission %
P scales Reading
P scales Maths using and applying
P scales Computing

Will's Easy Target Correct Omissions improved with every session and every week; however, this had not resulted in a higher sensitivity value *d′* because his Misses had increased with every week as well. Also, considering that his Hits had reduced with every week and every session this could mean he had just lost interest in the activity and was engaging less with the response button overall. Participants with higher P scales performed better at Omissions, both Easy and Hard Omissions. However, P scales did not have a significant correlation to Hit rate (for Hard or Easy Target), while having a significant relationship with lower False Alarm Rates and higher *d′* sensitivity level. This means that higher P scales correlated with lower impulsivity (holding back on Imitation slides), hence lower False Alarms and higher Correct Omissions, which in the absence of any relationship between P scales and Hit rate, led to a higher *d′* sensitivity value as well. Expectedly there was an inverse relationship between Correct Easy Omissions and False Alarms and Wrong Commissions of both Easy and Hard slides. Participants that were better at Correct easy Omissions were also better at Correct Hard Omissions. Participants with higher 'P scales Reading', 'Maths using and applying' and 'Computing' were better at holding back button presses on Easy Imitation slides.

Press Count
Max Press Count
Mean Press Count

The direct relationship between Press Count and Max Press Count and Mean Press Count shows that participants that had the highest press counts also had consistent higher presses throughout the session.

High-Level Attention feature
P scales Mean
d′ (D Prime)
FAR (False Alarm Rate)
Wrong Commission %
Omission %
Wrong Commission Hard %
Wrong Commission Easy %
Omission Hard %
Omission Easy %

The hand-picked compound High-Level Attention feature had a direct correlation with *d′* sensitivity which means that the High-Level Attention feature had a direct relationship with higher outcomes in both SDT and the CPT, which is also related to higher attention and performance over sustained and selective periods.

Body Speed Average
Age
B''D (Bias)
H (Hit rate)
Mean First Response Time
Miss %
Hit Hard %
Hit Easy %
Miss Hard %
Miss Easy %

Participant Age had a direct correlation with mean Body Speed. Meaning that the older participants had more body movement (while sat down). This, however, could simply be because our youngest participant had less movement as he uses a wheelchair. Participants with higher body speed, also had negative bias, meaning that they were more liberal with their button presses. This is also shown in the relationship with higher Hit rates and lower Miss rates. Participants that had higher body speed (fidgeting values) also had faster response times.

High-Level Inattention feature
Age
H (Hit rate)
Miss %
Hit Hard %
Miss Hard %

Older Participants had higher High-Level Inattention feature values. The High-Level Inattention feature had a direct correlation with higher Hit rates and fewer Misses, without leading to greater sensitivity d' values. Meaning that the feature relates to participants taking more chances, being more impulsive and having a lower threshold for pressing the button on Target or non-Target slides. In conclusion, the High-Level Compound Feature which was calculated as the mean of the normalized feature values of body fidgeting, eyes off-screen and press count, directly correlated with more impulsive behavior and less calculated button presses. In other words, participants with higher High-Level Inattention feature values were more liberal in their button presses.

EEG Power AB
H (Hit rate)
Mean First Response Time
Miss %
Hit Hard %
Hit Easy %
Miss Hard %
Miss Easy %

EEG Power AB is the power spectrum density of the of the α bands over the power spectrum density of the β band. These readings were taken from the frontal lobe at locations TP9, AF7, FPz, AF8 and TP1. As discussed in the feature extraction section 4.1.1, EEG, this ratio is noted to be directly correlated with alertness. In our study this EEG feature (the ratio of the power spectrum density of α to β) was directly related to faster response times. Also, very much like the HLI, Age, Body Speed Average, ‘P scales Reading’, ‘P scales Maths using and applying’ and ‘P scales Computing’, EEG Power AB had a direct correlation with higher Hit rates and lower Misses. In other words, participants with greater values in these indexes had a more liberal approach to button presses and a lower threshold to when they pressed the button.

Mean Press Count
Hit Hard %

The more repetitive presses a participant had, the worse they were at discovering Hard Target.

Age
Body Speed Average
High-Level Inattention feature
H (Hit rate)
Max Response Time
Mean First Response Time
Hit Hard %
Hit Easy %
Hit %
Miss Hard %
Miss Easy %
Miss %

Participant Age had a direct correlation with Body Speed (fidgeting) however as discussed in the relationships with Body Speed, this could be the outcome of the youngest participant Mark using a wheelchair. Older participants also had faster mean response times. Participant age had a direct correlation with Hit rate and lower Misses, however, this did not result in higher sensitivity levels, as there was no relationship age and lower False Alarms, so the relationship between Age and *d'* (sensitivity) wasn't as strong. This was similar to the relationships of High-Level Inattention feature, which shows that the older participants were more liberal with their button presses and had a lower threshold (criteria) to when they pressed the button on either Target or Imitation slides. This relationship was further reinforced by the direct relationship between Age and High-Level Inattention feature.

P scales Mean
High-Level Attention feature
d' (D Prime)
FAR (False Alarm Rate)
Wrong Commission Hard %
Wrong Commission Easy %
Wrong Commission %
Omission Hard %
Omission Easy %
Omission %
P scales Reading

The mean of participant P scales (participant average progress in learning) is of importance to us, as it denotes participant progress through the academic curriculum. Interestingly, despite the age of our participants being varied (16, 18, 19 and 19) and age having a significant correlation to Hit rate and response times, participant Age does not correlate to participant P scales. P scales had a direct correlation to sensitivity levels *d'* (D Prime) which is the main success outcome of the SDT in the CPT. Meaning that participants that did better in the CPT had more intellectual capacity to also perform better in their academic studies. Participants with higher P scales were also less likely to be impulsive as their False Alarm Rate was lower, and they were more selective by holding back their responses on non-Target signals for both Easy and Hard Targets. This was further strengthened by the correlation between higher P scales and higher Correct Omissions for both Easy and Hard Targets. Lastly, participants who did better in their 'P scales Reading' had a higher average P scale overall. This could be due to the nature of 'reading' having a role in achievements in other P scale categories.

P scales Speaking
Max Press Count
Mean Press Count
Min Response Time

Participants with less repetitive press counts (less impulsivity) performed better in their 'P scales Speaking'. Participants that also take longer to respond (longer response times) perform better in their 'P scales Speaking'.

P scales Listening
Omission Hard %
Eyes Off-Screen
Eye Scanning
Eye Dwelling
High-Level Inattention feature

Participants that had better 'P scales Listening' did better at keeping their eyes on the screen, they were also better in scanning the screen and focusing on the Targets when they found them. Participants with higher 'P scales Listening' also had higher average High-Level Inattention feature levels, which as discussed in the analysis of High-Level Inattention feature relates to being more liberal with and having a lower threshold (criteria) for pressing the button.

P scales Reading
P scales Mean
Omission Easy %
Hit Easy %
Miss Easy %
Speed Left Hand
EEG Power AB
High-Level Attention feature
P scales Maths using and applying
P scales Computing
Entropy Head
Entropy Neck
Entropy Left Shoulder
Entropy Right Shoulder
Entropy Spine
Entropy Left Hand
Entropy Right Hand

As discussed in the correlations for P scales, the progress a participant had made in their Reading P scales had a direct correlation with their progress in other P scales and in turn, improved their academic progress overall. Participants that had advanced further in their 'P scales Reading' had less Misses on Easy Target. Participants that had advanced further in their 'P scales Reading' also had more Easy Target Hits and fewer Misses. Participants with higher 'P scales Reading' had higher EEG Power AB feature (the ratio of the power spectrum density of α to β) values, greater HLA values, and greater range in body movement. This greater body range movement however did not result in a higher body speed, but just in greater range in body movement. There was also a direct relationship between 'P scales Reading' and 'P scales Computing'.

P scales Writing
Omission Hard %

Participants who had greater progress in their 'P scales Writing' were better at not pressing the button on Hard Imitation slides, meaning that they were better at determining that the Target was not there in a hard to assess non-Target slide. Meaning that they were less impulsive; this could be due to the similarity between the skills required in writing and distinguishing and differentiating different but similar shapes.

P scales Maths Space, Shape, Measure
Single Press
Speed Left Hand
Speed Right Hand
Entropy Head
Entropy Neck
Entropy Left Shoulder
Entropy Right Shoulder
Entropy Spine
Entropy Left Hand
Entropy Right Hand

Participants that showed greater speed in their hand movements or had a greater range in their body movements overall had higher 'P scales Maths, Space, Shape and Measure'. Participants that had greater confident Single Presses (as compared to numerous presses on each slide) had progressed further in their 'P scales Maths, Space, Shape and Measure'.

P scales Maths using and applying
H (Hit rate)
Mean First Response Time
Miss %
Omission Easy %
Hit Hard %
Hit Easy %
Miss Hard %
Miss Easy %
EEG Power AB
Max Response Time
Mean First Response Time
Mean Response times
Hit %
P scales Reading
P scales Computing

Just as in the case of Body Speed, High-Level Inattention feature, EEG power AB and Age, participants who were more liberal with their button presses (had a lower threshold to when they press the button) had progressed further in their 'P scales Maths using an applying'. Participants who had faster Response Times in total, also did better in their 'P scales for Maths using and applying'. This P scale ('Maths using and applying') also related directly to 'P scales Reading' and 'Computing'. Meaning that the further the participant was in 'Maths using and applying' the further they also were in 'Reading' and 'Computing'. Despite being more liberal with their button presses, participants that had progressed further in their 'P scales Maths using and applying', were able to perform better in holding back on Easy Imitation Targets.

P scales Computing
H (Hit rate)
Mean First Response Time
Miss %
Omission Easy %
Hit Hard %
Hit Easy %
Miss Hard %
Miss Easy %
EEG Power AB
Max Response Time
Mean First Response Time
Mean Response times
Hit %
P scales Reading
P scales Maths using and applying

Participants who were more liberal with their button presses (they had a lower threshold or criteria) for pressing the button, had progressed further in their 'P scales Computing'. Similarly, participants with a lower button press threshold, had faster Body Speed movements, higher High-Level Inattention feature values, higher EEG power AB values and were older. Participants that had faster button presses progressed further with their 'P scales Computing'. Despite being more liberal with their button presses, participants that had progressed further in their 'P scales Computing' were able to perform better in holding back on Easy Imitation Targets. There is also a direct correlation between progressing in 'P scales Computing' and Progressing in 'P scales Reading' and 'P scales Maths using and applying'.

Chapter 7. REFERENCES

[1] M. J. Eliason and L. C. Richman, "The Continuous Performance Test in learning disabled and nondisabled children.," *J. Learn. Disabil.*, vol. 20, no. 10, pp. 614–9, Dec. 1987.

[2] E. Bloom and N. Heath, "Recognition, expression, and understanding facial expressions of emotion in adolescents with nonverbal and general learning disabilities," *J. Learn. Disabil.*, vol. 43, no. 2, pp. 180–192, Mar. 2010.

[3] V. L. Petti, S. L. Voelker, D. L. Shore, and S. E. Hayman-Abello, "Perception of Nonverbal Emotion Cues by Children with Nonverbal Learning Disabilities," *J. Dev. Phys. Disabil.*, vol. 15, no. 1, pp. 23–36, 2003.

[4] H. B. Holder and S. W. Kirkpatrick, "Interpretation of emotion from facial expressions in children with and without learning disabilities," *J. Learn. Disabil.*, vol. 24, no. 3, pp. 170–177, Mar. 1991.

[5] L. Weigel, P. E. Langdon, S. Collins, and Y. O'Brien, "Challenging behaviour and learning disabilities: The relationship between expressed emotion and staff attributions," *Br. J. Clin. Psychol.*, vol. 45, no. 2, pp. 205–216, 2006.

[6] F. Heider, *The psychology of interpersonal relations.* Hoboken: John Wiley & Sons Inc, 1958.

[7] D. Dagnan, P. Trower, and R. Smith, "Care staff responses to people with learning disabilities and challenging behaviour: A cognitive-emotional analysis," *Br. J. Clin. Psychol.*, vol. 37, no. 1, pp. 59–68, Feb. 1998.

[8] R. Sharrock, A. Day, F. Qazi, and C. R. Brewin, "Explanations by professional care staff, optimism and helping behaviour: an application of attribution theory.," *Psychol. Med.*, vol. 20, no. 4, pp. 849–55, Nov. 1990.

[9] "The COMPLEX LEARNING DIFFICULTIES AND DISABILITIES Research Project: Developing meaningful pathways to personalised learning Engagement Profile and Scale Research Team Contacts."

[10] B. Carpenter, "Engaging young people with complex learning difficulties and disabilities in learning," no. Cldd.

[11] J. P. Roscoe, "'Can the use of a humanoid robot enhance the learning of children with learning disabilities?,'" University of Nottingham, 2014.

[12] P. Standen *et al.*, "Engaging Students with Profound and Multiple Disabilities Using Humanoid Robots," Springer, Cham, 2014, pp. 419–430.

[13] T. Hughes-Roberts *et al.*, "Examining engagement and achievement in learners with individual needs through robotic-based teaching sessions," *Br. J. Educ. Technol.*, vol. 50, no. 5, pp. 2736–2750, Sep. 2019.

[14] J. I. Marcum, "A statistical theory of target detection by pulsed radar," Dec. 1947.

[15] H. L. Swanson, “Vigilance Deficit in Learning Disabled Children: A Signal Detection Analysis,” *J. Child Psychol. Psychiatry.*, vol. 22, no. 4, pp. 393–9, Oct. 1981.

[16] X. Li and Q. Ji, “Active affective state detection and user assistance with dynamic Bayesian networks,” *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans.*, vol. 35, no. 1, pp. 93–105, 2005.

[17] A. Lanatà, G. Valenza, and E. P. Scilingo, “Eye gaze patterns in emotional pictures,” *J. Ambient Intell. Humaniz. Comput.*, vol. 4, no. 6, pp. 705–715, Dec. 2013.

[18] R. N. Khushaba, C. Wise, S. Kodagoda, J. Louviere, B. E. Kahn, and C. Townsend, “Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking,” *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3803–3812, Jul. 2013.

[19] R. W. Picard, “Affective Computing,” *MIT Press*, no. 321, pp. 1–16, 1995.

[20] M. Pantic, A. Nijholt, A. Pentland, and T. S. Huang, “Human-Centred Intelligent Human-Computer Interaction (HCI^2): how far are we from attaining it?,” *Int. J. Auton. Adapt. Commun. Syst.*, vol. 1, no. 2, pp. 168–187, 2008.

[21] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[22] M. Pantic, A. Pentland, A. Nijholt, and T. Huang, “Human Computing and Machine Understanding of Human Behavior: A Survey,” 2006.

[23] Beckman Institute, “Multimodal Human Computer Interaction Towards a Proactive Computer,” 2001.

[24] R. Ishii, Y. Shinohara, I. Nakano, and T. Nishida, “Combining Multiple Types of Eye-gaze Information to Predict User’s Conversational Engagement,” *Hum. Factors*, pp. 1–8, 2011.

[25] A. Jaimes and N. Sebe, “Multimodal human–computer interaction: A survey,” *Comput. Vis. Image Underst.*, vol. 108, no. 1–2, pp. 116–134, Oct. 2007.

[26] M. Pantic, N. Sebe, J. F. Cohn, and T. Huang, “Affective Multimodal Human-Computer Interaction,” *Proc. 13th Annu. ACM Int. Conf. Multimed.*, pp. 669–676, 2005.

[27] R. Sharma, V. I. Pavlovic, and T. S. Huang, “Toward multimodal human-computer interface,” *Proc. IEEE*, vol. 86, no. 5, pp. 853–869, May 1998.

[28] A. Kleinsmith and N. Bianchi-Berthouze, “Recognizing Affective Dimensions from Body Posture,” in *Affective Computing and Intelligent Interaction*, 2007, pp. 48–58.

[29] N. Savva, A. Scarinzi, and N. Bianchi-Berthouze, “Continuous recognition of player’s affective body expression as dynamic quality of aesthetic experience,” *IEEE Trans. Comput. Intell. AI Games*, vol. 4, no. 3, pp. 199–212, 2012.

[30] S. Asteriadis, K. Karpouzis, and S. Kollias, “The Importance of Eye Gaze and Head Pose to Estimating Levels of Attention,” *2011 Third Int. Conf. Games Virtual Worlds Serious Appl.*, no. September 2015, pp. 186–191, 2011.

[31] Z. Zeng, J. Tu, B. M. Pianfetti, and T. S. Huang, "Audio-visual affective expression recognition through multistream fused HMM," *IEEE Trans. Multimed.*, vol. 10, no. 4, pp. 570–577, 2008.

[32] N. Bianchi-Berthouze, W. W. Kim, and D. Patel, "Does Body Movement Engage You More in Digital Game Play? and Why?," in *Affective Computing and Intelligent Interaction*, 2007, pp. 102–113.

[33] "scipy.stats.norm — SciPy v0.16.1 Reference Guide." [Online]. Available: https://docs.scipy.org/doc/scipy-0.16.1/reference/generated/scipy.stats.norm.html. [Accessed: 17-May-2020].

[34] S. Aslan *et al.*, "Investigating the Impact of a Real-time, Multimodal Student Engagement Analytics Technology in Authentic Classrooms," in *Conference on Human Factors in Computing Systems - Proceedings*, 2019.

[35] G. A. Berry, V. I. Pavlovic, and T. S. Huang, "A Multimodal Human-Computer Interface for the Control of a Virtual Environment," pp. 141–144, 1998.

[36] C. H. Wu, Y. M. Huang, and J. P. Hwang, "Review of affective computing in education/learning: Trends and challenges," *Br. J. Educ. Technol.*, vol. 47, no. 6, pp. 1304–1323, Nov. 2016.

[37] C. Djeraba, A. Lablack, and Y. Benabbas, *Multi-Modal User Interactions in Controlled Environments*. 2010.

[38] S. K. D'Mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Comput. Surv.*, vol. 47, no. 3, 2015.

[39] G. R. (2007). VandenBos, *APA dictionary of psychology (1st ed.)*. 2007.

[40] A. R. Basawapatna, A. Repenning, K. H. Koh, and H. Nickerson, "The zones of proximal flow," *Proc. ninth Annu. Int. ACM Conf. Int. Comput. Educ. Res. - ICER '13*, p. 67, 2013.

[41] D. Watson and L. A. Clark, "Self- Versus Peer Ratings of Specific Emotional Traits: Evidence of Convergent and Discriminant Validity," *J. Pers. Soc. Psychol.*, vol. 60, no. 6, pp. 927–940, Jun. 1991.

[42] American Psychological Association, *APA College Dictionary of Psychology*, Second. Washington, DC, 2009.

[43] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, *The Classification of Educational Goals*. 1956.

[44] L. W. Anderson *et al.*, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. 2001.

[45] C. Peters, G. Castellano, and S. de Freitas, "An exploration of user engagement in HCI," *Proc. Int. Work. Affect. Virtual Agents Soc. Robot. - Affin. '09*, no. January 2014, pp. 1–3, 2009.

[46] R. Iovannone, G. Dunlap, H. Huber, and D. Kincaid, "Effective Educational Practices for Students with Autism Spectrum Disorders.," *Focus Autism Other Dev. Disabil.*, vol.

18, pp. 150–165, 2003.

[47] B. Carpenter *et al.*, *Engaging Learners with Complex Learning Difficulties and Disabilities: A resource book for teachers and teaching assistants*. Taylor & Francis, 2015.

[48] C. R. Henrie, "Measuring Student Engagement in Technology-Mediated Learning Environments," *ProQuest Diss. Theses*, p. 172, 2016.

[49] J. Sebba, N. Brown, S. Steward, M. Galton, and M. James, "An Investigation of Personalised Learning Approaches Used by Schools," 2007.

[50] R. Dolan and T. Hall, "Universal Design for Learning: Implications for large-scale assessment," *IDA Perspect.*, vol. 27, pp. 22–25, 2001.

[51] "Engaging children with complex learning difficulties and disabilities in the Primary Classroom. Barry Carpenter," no. 2002, 2011.

[52] R. W. Picard, *Affective computing*. MIT Press, 2000.

[53] H. L. O'Brien and E. G. Toms, "What is user enagement? A Conceptual Framework for defining user engagement with technology," *J. Am. Soc. Inf. Sci. Technol.*, vol. 59, no. 6, pp. 938–955, Apr. 2008.

[54] N. Savva and N. Bianchi-Berthouze, "Automatic recognition of affective body movement in a video game scenario," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng.*, vol. 78 LNICST, pp. 149–159, 2012.

[55] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artif. Intell.*, vol. 166, no. 1–2, pp. 140–164, Aug. 2005.

[56] M. P. Michalowski, S. Sabanovic, and R. Simmons, "A Spatial Model of Engagement for a Social Robot."

[57] C. Peters, S. Asteriadis, K. Karpouzis, and E. de Sevin, "Towards a Real-Time Gaze-Based Shared Attention for a Virtual Agent," *Work. Affect. Interact. Nat. Environ. (AFFINE), ACM Int. Conf. Multimodal Interfaces*, no. June 2014, 2008.

[58] R. E. Kraut and J. D. Price, "Machiavellianism in parents and their children.," *J. Pers. Soc. Psychol.*, vol. 33, no. 6, pp. 782–786, 1976.

[59] A. Siwy-hudowska and I. Pilch, "Machiavellianism in families: Analysis of relationships between the Machiavellianism level of grown-up children and their parents," *Psychol. Stud. (Mysore).*, vol. 52, no. January, pp. 16–23, 2014.

[60] A. Barlow, P. Qualter, and M. Stylianou, "Relationships between Machiavellianism, emotional intelligence and theory of mind in children," *Pers. Individ. Dif.*, vol. 48, no. 1, pp. 78–82, Jan. .

[61] T. Brooks, B. Cockbill, J. Fotheringham, and H. Rawson, "The Complex Learning Difficulties and Disabilities Research Project Developing pathways to personalised learning."

[62] B. Carpenter, S. Rose, H. Rawson, and J. Egerton, "The rules of engagement," *SEN Mag.*, no. 54, pp. 34–54, 2011.

[63] D. J. Shernoff, M. Csikszentmihalyi, B. Shneider, E. S. Shernoff, B. Schneider, and E. S. Shernoff, "Student engagement in high school classrooms from the perspective of flow theory," *Sch. Psychol. Q.*, vol. 18, no. 2, pp. 158–176, 2003.

[64] M. Taheri *et al.*, "State Diagram for Affective Learning in an Educational Platform," pp. 4–6, 2018.

[65] P. H. Mirvis and M. Csikszentmihalyi, "Flow: The Psychology of Optimal Experience," *Acad. Manag. Rev.*, vol. 16, no. 3, pp. 636–640, 1991.

[66] M. Csikszentmihalyi, "Creativity: Invention, Flow and the psychology of discovery and Invention," in *Personnel Psychology*, vol. 51, no. 3, 1997, p. 16.

[67] M. Csikszentmihalyi and M. Csikzentmihaly, "Flow: The psychology of optimal experience," vol. 41, no. May, 1991.

[68] M. Chen, B. Kolko, E. Cuddihy, and E. Medina, "Modelling and measuring engagement in computer games," in *The annual conference for the Digital Games Research Association (DiGRA)*, 2005.

[69] M. D. Chen, "Writings | Mark Danger Chen." [Online]. Available: https://web.archive.org/web/20180308115459/http://markdangerchen.net/papers/#comment-14. [Accessed: 26-May-2020].

[70] M. Chen, B. E. Kolko, E. Cuddihy, and E. Medina, "Modeling but NOT measuring engagement in computer games," *Proc. 7th Int. Conf. Games + Learn. + Soc. Conf.*, vol. Madison, W, pp. 55–63, 2011.

[71] S. D'Mello and A. Graesser, "Dynamics of affective states during complex learning," *Learn. Instr.*, vol. 22, no. 2, pp. 145–157, 2012.

[72] M. Csikszentmihalyi, "The Flow of Creativity," *Creat. Flow Psychol. Discov. Invent.*, pp. 107–126, 1996.

[73] M. Csikszentmihalyi, J. Nakamura, and S. Abuhamdeh, "Flow A General Context," *Handb. Competence Motiv. (pp. 598-608)*, pp. 598–608, 2005.

[74] W. Friedman, *About time: Inventing the fourth dimension.* 1990.

[75] R. Conti, "Time Flies: Investigating the Connection Between Intrinsic Motivation and the Experience of Time," *J. Pers.*, vol. 69, no. 1, pp. 1–26, 2001.

[76] W. James, "The principles of psychology, Vol I.," in *The principles of psychology, Vol I.*, 2004.

[77] A. Possamai and A. J. Blasi, "The Protestant Ethic and the Spirit of Capitalism, by Max Weber," *The SAGE Encyclopedia of the Sociology of Religion*. 2020.

[78] D. E. Berlyne, *Conflict, arousal, and curiosity.* McGraw-Hill Book Company, 2006.

[79] J. M. Hunt, "Intrinsic motivation and its role in psychological development," in *Nebraska symposium on motivation*, 1965, vol. 13, pp. 189–282.

[80] A. Basawapatna, A. Repenning, and M. Savignano, "The Zones of Proximal Flow

Tutorial Designing Computational Thinking Cliffhangers," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 2019, no. February, pp. 428–434.

[81] E. A. Linnenbrink and P. R. Pintrich, "Achievement goal theory and affect: An asymmetrical bidirectional model," *Educational Psychologist*, vol. 37, no. 2. Lawrence Erlbaum Associates Inc., pp. 69–78, 2002.

[82] M. Csikszentmihalyi, "Applications of flow in human development and education: The collected works of Mihaly Csikszentmihalyi," *Appl. Flow Hum. Dev. Educ. Collect. Work. Mihaly Csikszentmihalyi*, pp. 1–494, 2014.

[83] M. Csikszentmihalyi and I. Csikszentmihalyi, *Optimal experience: Psychological studies of flow in consciousness*. 1992.

[84] E. A. Locke, K. N. Shaw, L. M. Saari, and G. P. Latham, "Goal setting and task performance: 1969-1980," *Psychol. Bull.*, vol. 90, no. 1, pp. 125–152, 1981.

[85] S. Craig, A. Graesser, J. Sullins, and B. Gholson, "Affect and learning: An exploratory look into the role of affect in learning with AutoTutor," *J. Educ. Media*, vol. 29, no. 3, pp. 241–250, 2004.

[86] S. K. Perry, *Writing in flow: Keys to enhance creativity*, 1st ed. Cincinnati Ohio: Writer's Digest Books, 1999.

[87] K. Sawyer, "Improvisational Creativity: An Analysis of Jazz Performance," *Creat. Res. J.*, vol. 5, no. 3, pp. 253–263, 1992.

[88] M. Csikszentmihalyi, K. R. Rathunde, S. Whalen, and M. S. M. Wong, "Talented teenagers : the roots of success and failure," p. 307, 1993.

[89] S. A. Jackson, P. R. Thomas, H. W. Marsh, and C. J. Smethurst, "Relationships between Flow, Self-Concept, Psychological Skills, and Performance," *J. Appl. Sport Psychol.*, vol. 13, no. 2, pp. 129–153, 2001.

[90] G. L. Stein, J. C. Kimiecik, J. Daniels, and S. A. Jackson, "Psychological antecedents of flow in recreational sport. Personality and Social Psychology Bulletin," *Socirty Personal. Soc. Pscycology, Inc.*, vol. 21, no. 2, pp. 125–135, 1995.

[91] L. L. S. Vygotsky, "Mind in society: The development of higher psychological processes," *Mind Soc. Dev. High. Psychol. Process.*, vol. Mind in So, p. 159, 1978.

[92] C. Heine, "Flow and Achievement in Mathematics," University of Chicago, Chicago, IL, 1996.

[93] J. A. Schmidt, "Overcoming challenges: Exploring the role of action, experience, and opportunity in fostering resilience among adolescents.," *Diss. Abstr. Int. Sect. B Sci. Eng.*, 1999.

[94] C. R. Henrie, L. R. Halverson, and C. R. Graham, "Measuring student engagement in technology-mediated learning: A review," *Comput. Educ.*, vol. 90, pp. 36–53, 2015.

[95] M. J. Lee and B. Ferwerda, "Personalizing online educational tools," *Humaniz. 2017 - Proc. 2017 ACM Work. Theory-Informed User Model. Tailoring Pers. Interfaces, co-located*

with IUI 2017, pp. 27–30, 2017.

[96] A. K. |Wiggins. J. B. |Grafsgaard. J. F. |Boyer. K. E. E. N. |Lester. J. C. Vail, "The Affective Impact of Tutor Questions: Predicting Frustration and Engagement," *Int. Educ. Data Min. Soc.*, 2016.

[97] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," *MindTrek - 12th Int. MindTrek Conf. Entertain. Media Ubiquitous Era*, no. January 2008, p. 13, 2008.

[98] K. Karpouzis, "Affective and gameful interfaces," in *International Conference on Intelligent User Interfaces, Proceedings IUI*, 2017, pp. 33–34.

[99] G. Hookham and K. Nesbitt, "A Systematic Review of the Definition and Measurement of Engagement in Serious Games," in *ACM International Conference Proceeding Series*, 2019.

[100] E. Gottardo and A. R. Pimentel, "Affective human-computer interaction in educational software: A hybrid model for emotion inference," in *ACM International Conference Proceeding Series*, 2017.

[101] J. Hamari, D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke, and T. Edwards, "Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning," *Comput. Human Behav.*, vol. 54, no. JANUARY, pp. 170–179, 2016.

[102] S. K. D'Mello, N. Bosch, and H. Chen, "Multimodal-multisensor affect detection," *Handb. Multimodal-Multisensor Interfaces Found. User Model. Common Modality Comb. - Vol. 2*, pp. 167–202, 2018.

[103] M. W. Eysenck, N. Derakshan, R. Santos, and M. G. Calvo, "Anxiety and cognitive performance: Attentional control theory," *Emotion*, vol. 7, no. 2. pp. 336–353, May-2007.

[104] N. Thompson and T. J. McGill, "Genetics with Jean: the design, development and evaluation of an affective tutoring system," *Educ. Technol. Res. Dev.*, vol. 65, no. 2, pp. 279–299, Apr. 2017.

[105] D. Goleman, *Emotional intelligence*. Bantam Books, 1995.

[106] P. Vos, P. De Cock, V. Munde, K. Petry, W. Van Den Noortgate, and B. Maes, "The tell-tale: What do heart rate; skin temperature and skin conductance reveal about emotions of people with severe and profound intellectual disabilities?," *Res. Dev. Disabil.*, vol. 33, no. 4, pp. 1117–1127, Jul. 2012.

[107] S. D. Mello, R. Picard, and A. Graesser, "Towards An Affect-Sensitive AutoTutor."

[108] M. Csíkszentmihályi, *Flow*. 2012.

[109] B. Carpenter, "A vision for the 21st century special school.," *London Spec. Sch. Acad. Trust*, 2010.

[110] C. Blackburn, "Engaging young children with complex learning dificulties and disabilities," vol. 14, no. 2, pp. 38–44, 2012.

[111] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, 2001.

[112] M. Hendrix, T. Bellamy-Wood, S. McKay, V. Bloom, and I. Dunwell, "Implementing Adaptive Game Difficulty Balancing in Serious Games," *IEEE Trans. Games*, vol. 1502, no. c, pp. 1–1, 2018.

[113] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, 2012.

[114] A. Omurtag, H. Aghajani, and H. O. Keles, "Decoding human mental states by whole-head EEG+fNIRS during category fluency task performance," *J. Neural Eng.*, vol. 14, no. 6, 2017.

[115] A. Gelan *et al.*, "Affordances and limitations of learning analytics for computer-assisted language learning: a case study of the VITAL project," *Comput. Assist. Lang. Learn.*, vol. 31, no. 3, pp. 294–319, 2018.

[116] I. Pages, L. Theorists, and V. L. Vygotsky, "Learning and Teaching with New Tech : Educational Theorists," pp. 1–3, 2018.

[117] H. Boulton *et al.*, "Multi-modalities in classroom learning environments," in *INTED2018 Proceedings*, 2018, vol. 1, no. February, p. 8.

[118] D. Cazzato, S. M. Castro, O. Agamennoni, G. Fernández, and H. Voos, "A non-invasive tool for attention-deficit disorder analysis based on gaze tracks," 2019, pp. 1–6.

[119] S. Athi Narayanan, M. R. Kaimal, K. Bijlani, M. Prasanth, and K. Sunil Kumar, "Computer vision based attentiveness detection methods in E-Learning," in *ACM International Conference Proceeding Series*, 2014, vol. 10-11-Octo.

[120] K. S. Saluja, J. Dv, S. Arjun, P. Biswas, and T. Paul, "Analyzing Eye Gaze of Users with Learning Disability," 2019, pp. 95–99.

[121] C. Conati, "Probabilistic assessment of user's emotions in educational games," *Appl. Artif. Intell.*, vol. 16, no. 7–8, pp. 555–575, 2002.

[122] S. Craig, A. Graesser, J. Sullins, and B. Gholson, "Affect and learning: An exploratory look into the role of affect in learning with AutoTutor," *J. Educ. Media*, vol. 29, no. 3, pp. 241–250, 2004.

[123] R. Pekrun and L. Linnenbrink-Garcia, *International Handbook of Psychology in Education*. 2010.

[124] A. Tellegen and G. Atkinson, "Openness to absorbing and self-altering experiences ('absorption'), a trait related to hypnotic susceptibility," *J. Abnorm. Psychol.*, vol. 83, no. 3, pp. 268–277, Jun. 1974.

[125] C. M. Mueller and C. S. Dweck, "Praise for intelligence can undermine children's motivation and performance.," *J. Pers. Soc. Psychol.*, vol. 75, no. 1, pp. 33–52, 1998.

[126] L. Vygotsky, "The interaction between learning and development." pp. 34–40, 1978.

[127] S. Chaiklin, “The zone of proximal development in Vygotsky’s analysis of learning and instruction,” *Vygotsky’s Educ. theory Pract. Cult. Context*, pp. 1–21, 2003.

[128] J. Radford, P. Bosanquet, R. Webster, and P. Blatchford, “Scaffolding learning for independence: Clarifying teacher and teaching assistant roles for children with special educational needs,” *Learn. Instr.*, vol. 36, pp. 1–10, 2015.

[129] C. Read, “Scaffolding children’s talk and learning,” *Child Psychol. Psychiatry*, pp. 1–18, 2006.

[130] I. Verenikina, “Understanding Scaffolding and the ZPD in Educational Research,” *Development*, p. 8, 2003.

[131] J. Hermida, *Facilitating Deep Learning: Pathways to Success for University and College Teachers*. 2015.

[132] L. Robertson, *International handbook of emotions in education*, vol. 31, no. 1. 2015.

[133] D. H. Rose and A. Meyer, “Teaching Every Student in the Digital Age: Universal Design for Learning.,” Nov. 2001.

[134] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle, “A systematic literature review of empirical evidence on computer games and serious games,” *Comput. Educ.*, vol. 59, no. 2, pp. 661–686, 2012.

[135] J. C. Abrams, “On learning disabilities: Affective considerations,” *J. Reading, Writing, Learn. Disabil. Int.*, vol. 2, no. 3, p. 190, Jan. 1986.

[136] B. Hiebert, B. Y. Wong, and M. Hunter, “Affective influences on learning disabled adolescents.,” *Learn. Disabil. Q.*, vol. 5, no. 4, pp. 334–343, 1982.

[137] D. H. Hargreaves, “A new shape for schooling?,” 2006.

[138] “SSATrust - Project information,” 2015. [Online]. Available: http://complexld.ssatrust.org.uk/project-information.html. [Accessed: 13-Mar-2015].

[139] R. R. McCrae, “Consensual validation of personality traits: Evidence from self-reports and ratings,” *J. Pers. Soc. Psychol.*, vol. 43, no. 2, pp. 293–303, 1982.

[140] H. L. Swanson, “A Developmental Study of Vigilance in Learning-disabled and Nondisabled Children,” *J. Abnorm. Child Psychol.*, vol. 11, no. 3, pp. 415–29, Sep. 1983.

[141] J. S. Warm, R. Parasuraman, and G. Matthews, “Vigilance requires hard mental work and is stressful.,” *Hum. Factors*, vol. 50, no. 3, pp. 433–41, Jun. 2008.

[142] G. J. DuPaul, A. D. Anastopoulos, T. L. Shelton, D. C. Guevremont, and L. Metevia, “Multimethod Assessment of Attention-Deficit Hyperactivity Disorder: The Diagnostic Utility of Clinic-Based Tests,” *J. Clin. Child Psychol.*, vol. 21, no. 4, pp. 394–402, Dec. 1992.

[143] C. A. Riccio, J. J. M. Waldrop, C. R. Reynolds, and P. Lowe, “Effects of stimulants on the continuous performance test (CPT): implications for CPT use and interpretation.,” *J. Neuropsychiatry Clin. Neurosci.*, vol. 13, no. 3, pp. 326–335, Aug. 2001.

[144] C. K. Conners, “The computerized continuous performance test.,” *Psychopharmacol. Bull.*, vol. 21, no. 4, pp. 891–2, Jan. 1985.

[145] H. Stanislaw and N. Todorov, “Calculation of signal detection theory measures,” *Behav. Res. Methods, Instruments, Comput.*, vol. 31, no. 1, pp. 137–149, Mar. 1999.

[146] A. J. Fallgatter, D. Brandeis, and W. K. Strik, “A robust assessment of the NoGo-anteriorisation of P300 microstates in a cued Continuous Performance Test,” *Brain Topogr.*, vol. 9, no. 4, pp. 295–302, Jun. 1997.

[147] S. Raz, Y. Bar-Haim, A. Sadeh, and O. Dan, “Reliability and validity of the online continuous performance test among young adults.,” *Assessment*, vol. 21, no. 1, pp. 108–18, 2014.

[148] J. M. Halperin, V. Sharma, E. Greenblatt, and S. T. Schwartz, “Assessment of the Continuous Performance Test: Reliability and validity in a nonreferred sample.,” *Psychol. Assess.*, vol. 3, no. 4, pp. 603–608, 1991.

[149] S. H. Klee and B. D. Garfinkel, “The computerized continuous performance task: a new measure of inattention.,” *J. Abnorm. Child Psychol.*, vol. 11, no. 4, pp. 487–495, 1983.

[150] J. N. Epstein, A. Erkanli, C. K. Conners, J. Klaric, J. E. Costello, and A. Angold, “Relations between Continuous Performance Test performance measures and ADHD behaviors.,” *J. Abnorm. Child Psychol.*, vol. 31, no. 5, pp. 543–54, Oct. 2003.

[151] A. J. Sostek, M. S. Buchsbaum, and J. L. Rapoport, “Effects of amphetamine on vigilance performance in normal and hyperactive children,” *J. Abnorm. Child Psychol.*, vol. 8, no. 4, pp. 491–500, 1980.

[152] W. Peterson, T. Birdsall, and W. Fox, “The theory of signal detectability,” *Trans. IRE Prof. Gr. Inf. Theory*, vol. 4, no. 4, pp. 171–212, Sep. 1954.

[153] W. P. Tanner and J. A. Swets, “A decision-making theory of visual detection,” *Psychol. Rev.*, vol. 61, no. 6, pp. 401–409, 1954.

[154] N. A. Macmillan and C. D. Creelman, *Detection Theory: A User’s Guide*. Psychology Press, 2005.

[155] J. A. Swets and R. M. Pickett, *Evaluation of diagnostic systems : methods from signal detection theory*. Academic Press, 1982.

[156] M. Taheri, D. Brown, and N. Sherkat, “Modeling Engagement with Multimodal Multisensor Data: The Continuous Performance Test as an Objective Tool to Track Flow,” in *ICACII 2020: 14. International Conference on Affective Computing and Intelligent Interaction*, 2020, pp. 1–13.

[157] J Loong, “The manual for the continuous performance test,” San Luis Obispo (CA), 1991.

[158] K. S. Lassiter, J. C. Whitten, A. N. Bardos, R. C. D’Amato, and D. J. Raggio, “The construct specificity of the continuous performance test: Does inattention relate to behavior and achievement?,” *Dev. Neuropsychol.*, vol. 10, no. 2, pp. 179–188, 1994.

[159] J. Shi, T. Tao, W. Chen, L. Cheng, L. Wang, and X. Zhang, “Sustained attention in

intellectually gifted children assessed using a continuous performance test.," *PLoS One*, vol. 8, no. 2, p. e57417, Jan. 2013.

[160] C. A. Riccio, Cecil R. Reynolds, P. Lowe, and J. J. Moore, "The continuous performance test: a window on the neural substrates for attention?," *Arch. Clin. Neuropsychol.*, vol. 17, no. 3, pp. 235–272, Apr. 2002.

[161] National Association of Special Education Teachers, "Characteristics of Children with Learning Disabilities," 2005.

[162] R. P. Anderson, C. G. Halcomb, and R. B. Doyle, "The Measurement of Attentional Deficits," *Except. Child.*, vol. 39, no. 7, pp. 534–539, Apr. 1973.

[163] R. A. Barkley, "The ecological validity of laboratory and analogue assessment methods of ADHD symptoms," *J. Abnorm. Child Psychol.*, vol. 19, no. 2, pp. 149–178, 1991.

[164] J. K. Torgesen, "Studies of children with learning disabilities who perform poorly on memory span tasks.," *Journal of Learning Disabilities*, vol. 21, no. 10. SAGE PublicationsSage UK: London, England, pp. 605–612, 18-Dec-1988.

[165] P. V. Corkum and L. S. Siegel, "Is the Continuous Performance Task a Valuable Research Tool for use with Children with Attention-Deficit-Hyperactivity Disorder?," *J. Child Psychol. Psychiatry*, vol. 34, no. 7, pp. 1217–1239, 1993.

[166] J. O. Goldberg and M. M. Konstantareas, "Vigilance in Hyperactive and Normal Children on a Self-Paced Operant Task," *J. Child Psychol. Psychiatry*, vol. 22, no. 1, pp. 55–63, Jan. 1981.

[167] K. S. Katz *et al.*, "Effect of cerebral lesions on continuous performance test responses of school age children born prematurely," *J. Pediatr. Psychol.*, vol. 21, no. 6, pp. 841–855, 1996.

[168] S. G. Tarver and D. P. Hallahan, "Attention Deficits In Children With Learning Disabilities," *J. Learn. Disabil.*, vol. 7, no. 9, pp. 560–569, Nov. 1974.

[169] R. P. Anderson, C. G. Halcomb, R. B. Doyle, C. Holcomb, R. D.-E. Children, and U. 1973, "Measurement of attention in hyperactive children," *Except. Child.*, vol. 39, no. 7, pp. 534–539, Apr. 1973.

[170] L. Swanson, "Modification of Comprehension Deficits in Learning Disabled Children," *Learn. Disabil. Q.*, vol. 4, no. 2, pp. 189–202, May 1981.

[171] F. Vrana and R. O. O. Pihl, "Selective attention deficit in learning disabled children: a cognitive interpretation.," *J. Learn. Disabil.*, vol. 13, no. 7, pp. 387–391, Aug. 1980.

[172] L. Mussgay and R. Hertwig, "Signal detection indices in schizophrenics on a visual, auditory, and bimodal Continuous Performance Test," *Schizophr. Res.*, vol. 3, no. 5–6, pp. 303–310, 1990.

[173] H. Edebol, L. Helldin, and T. Norlander, "Measuring adult Attention Deficit Hyperactivity Disorder using the Quantified Behavior Test Plus," *PsyCh J.*, vol. 2, no. 1, pp. 48–62, 2013.

[174] T. S. Emser, B. A. Johnston, J. D. Steele, S. Kooij, L. Thorell, and H. Christiansen,

"Assessing ADHD symptoms in children and adults: Evaluating the role of objective measures," *Behav. Brain Funct.*, vol. 14, no. 1, pp. 1–14, 2018.

[175] "Qbtech... Rethinking ADHD." [Online]. Available: https://www.qbtech.com/. [Accessed: 19-May-2020].

[176] "The Test of Variables of Attention (T.O.V.A.®)." [Online]. Available: https://www.tovatest.com/. [Accessed: 19-May-2020].

[177] J. A. Swets and D. M. Green, *Signal detection theory and psychophysics*. Los Altos, Calif. : Peninsula Publishing, 1988.

[178] A. L. Brophy and P. B. Elliott, "Alternatives to a table of criterion values in signal detection theory," *Behav. Res. Methods, Instruments, Comput.*, vol. 18, no. 3, pp. 651–684, May 1986.

[179] "Welcome to Python.org." [Online]. Available: https://www.python.org/. [Accessed: 25-May-2020].

[180] A. L. Brophy, "Approximation of the inverse normal distribution function," *Behav. Res. Methods, Instruments, Comput.*, vol. 17, no. 3, pp. 415–417, May 1985.

[181] W. Donaldson, "Measuring Recognition Memory," *J. Exp. Psychol. Gen.*, vol. 121, no. 3, pp. 275–277, 1992.

[182] "SSATrust - Engagement profile & scale."

[183] S. A. Jackson and H. W. Marsh, "Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale," *J. Sport Exerc. Psycology*, vol. 18, 1996.

[184] F. Levy, "The Development of Sustained Attention (Vigilance) and Inhibition in Children: Some Normative Data," *J. Child Psychol. Psychiatry*, vol. 21, no. 1, pp. 77–84, 1980.

[185] M. Rosenberg, S. Noonan, J. Degutis, and M. Esterman, "Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task.," *Atten. Percept. Psychophys.*, vol. 75, no. 3, pp. 426–39, 2013.

[186] G. P. Richards, "Sustained and Selective Attention in Children with Learning Disabilities," pp. 1–2, 2015.

[187] C. Overtoom *et al.*, "Inhibition in children with attention-deficit/hyperactivity disorder: a psychophysiological study of the stop task.," *Biol. Psychiatry*, vol. 51, no. 8, pp. 668–76, 2002.

[188] C. C. E. OVERTOOM *et al.*, "Associations Between Event-Related Potentials and Measures of Attention and Inhibition in the Continuous Performance Task in Children With ADHD and Normal Controls," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 37, no. 9, pp. 977–985, Sep. 1998.

[189] Qbtech, "Clinical Documentation for QbTest in Patients With ADHD."

[190] "Conners Continuous Performance Test 3rd Edition." [Online]. Available: https://storefront.mhs.com/collections/conners-cpt-3. [Accessed: 27-May-2020].

[191] C. M. Lewandowski, “TOVA Manual,” *Eff. Br. mindfulness Interv. acute pain Exp. An Exam. Individ. Differ.*, vol. 1, no. 844, 2015.

[192] T. H. van Leeuwen *et al.*, “The continuous performance test revisited with neuroelectric mapping: impaired orienting in children with attention deficits,” *Behav. Brain Res.*, vol. 94, no. 1, pp. 97–110, Jul. 1998.

[193] H. E. Rosvold, A. F. Mirsky, I. Sarason, E. D. Bransome, and L. H. Beck, “A continuous performance test of brain damage.,” *J. Consult. Psychol.*, vol. 20, no. 5, pp. 343–350, 1956.

[194] R. L. Lindsay, T. Tomazic, M. D. Levine, and P. J. Accardo, “Attentional function as measured by a Continuous Performance Task in children with dyscalculia.,” *J. Dev. Behav. Pediatr.*, vol. 22, no. 5, pp. 287–92, 2001.

[195] D. Delgado-Gomez *et al.*, “Microsoft kinect-based continuous performance test: An objective attention deficit hyperactivity disorder assessment,” *J. Med. Internet Res.*, vol. 19, no. 3, pp. 1–8, 2017.

[196] X. Li and Q. Ji, “Active affective state detection and user assistance with dynamic Bayesian networks,” *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans.*, vol. 35, no. 1, pp. 93–105, 2005.

[197] H. Gunes and M. Piccardi, “Affect recognition from face and body: early fusion vs. late fusion,” *Syst. Man Cybern. 2005 IEEE Int. Conf.*, vol. 4, pp. 3437–3443, 2005.

[198] B. Schuller, *Multimodal Affect Databases—Collection, Challenges, and Chances*. Oxford University Press, 2014.

[199] J. A. Russell, J.-A. Bachorowski, and J.-M. Fernández-Dols, “Facial and Vocal Expressions of Emotion,” *Annu. Rev. Psychol.*, vol. 54, no. 1, pp. 329–349, Feb. 2003.

[200] J. A. Russell, “Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies.,” *Psychol. Bull.*, vol. 115, no. 1, pp. 102–141, 1994.

[201] H. A. Elfenbein and N. Ambady, “On the universality and cultural specificity of emotion recognition: A meta-analysis,” *Psychological Bulletin*, vol. 128, no. 2. American Psychological Association Inc., pp. 203–235, 2002.

[202] F. Karray, M. Alemzadeh, J. a Saleh, and M. N. Arab, “Human-Computer Interaction: Overview on State of the Art,” *Int. J. Smart Sens. Intell. Syst.*, vol. 1, no. 1, pp. 137–159, 2008.

[203] L. Wei and H. Hu, “Towards Multimodal Human-Machine Interface for Hands-free Control : A survey School of Computer Science & Electronic Engineering,” *Tech. Rep.*, no. January, pp. 1–26, 2011.

[204] N. Sebe, I. Cohen, T. Gevers, and T. S. Huang, “Multimodal Approaches for Emotion Recognition: A Survey,” in *Internet Imaging VI*, 2005, vol. 5670, pp. 56–67.

[205] A. B. Ashraf, S. Lucey, T. Chen, Z. Ambadar, K. M. Prkachin, and P. E. Solomon, “The painful face – Pain expression recognition using active appearance models,” *Image Vis. Comput.*, vol. 27, no. 12, pp. 1788–1796, Nov. 2009.

[206] Z. Duric *et al.*, “Integrating perceptual and cognitive modeling for adaptive and

intelligent human-computer interaction," *Proc. IEEE*, vol. 90, no. 7, pp. 1272–1289, Jul. 2002.

[207] L. Maat and M. Pantic, "Gaze-X: Adaptive, affective, multimodal interface for single-user office scenarios," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007, vol. 4451 LNAI, pp. 251–271.

[208] A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration," *Int. J. Human-Computer Stud.*, vol. 65, pp. 724–736, 2007.

[209] S. D. Mello, R. W. Picard, A. Graesser, S. D'Mello, R. W. Picard, and A. Graesser, "Toward an Affect-Sensitive AutoTutor," *IEEE Intell. Syst.*, vol. 22, no. 4, pp. 53–61, Jul. 2007.

[210] M. Garcia Valdez, J.-J. Merelo, A. Hernandez Aguila, and A. Mancilla Soto, *Mining of Keystroke and Mouse Dynamics to Increase the Engagement of Students with Programming Assignments*, vol. 56, no. 7. 2019.

[211] A. Doshi and M. M. Trivedi, "On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 453–462, 2009.

[212] S. K. D'Mello and A. Graesser, "Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features," *User Model. User-adapt. Interact.*, vol. 20, no. 2, pp. 147–187, 2010.

[213] N. Alyuz, E. Okur, U. Genc, S. Aslan, C. Tanriover, and A. A. Esme, "An unobtrusive and multimodal approach for behavioral engagement detection of students," *MIE 2017 - Proc. 1st ACM SIGCHI Int. Work. Multimodal Interact. Educ. Co-located with ICMI 2017*, vol. 2017-Novem, no. 1, pp. 26–32, 2017.

[214] E. Okur, N. Alyuz, S. Aslan, U. Genc, C. Tanriover, and A. A. Esme, "Behavioral engagement detection of students in the wild," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10331 LNAI, pp. 250–261, 2017.

[215] N. Alyuz *et al.*, "Semi-supervised model personalization for improved detection of learner's emotional engagement," *ICMI 2016 - Proc. 18th ACM Int. Conf. Multimodal Interact.*, pp. 100–107, 2016.

[216] C. Lewis, P. Poison, C. Wharton, and J. Rieman, "Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces," in *Conference on Human Factors in Computing Systems - Proceedings*, 1990, pp. 235–242.

[217] A. Waibel, M. T. Vo, P. Duchnowski, and S. Manke, "Multimodal interfaces," *Artif. Intell. Rev.*, vol. 10, no. 3–4, pp. 299–319, 1996.

[218] S. Oviatt and R. Lunsford, "Multimodal Interfaces for Cell Phones and Mobile Technology," *J. Sol-Gel Sci. Technol.*, vol. 8, no. 2, pp. 127–132, Jun. 1997.

[219] D. H. Sykes, V. I. Douglas, G. Weiss, and K. K. Minde, "Attention in hyperactive children and the effects methylphenidaten (Ritalin)," *J. child Psychol. psychiatry*, vol. 12,

no. 2, pp. 129–139, 1971.

[220] R. C. Wilson, A. Shenhav, M. Straccia, and J. D. Cohen, "The Eighty Five Percent Rule for optimal learning," *Nat. Commun.*, vol. 10, no. 1, pp. 1–9, 2019.

[221] "R2016a - Updates to the MATLAB and Simulink product families - MATLAB & Simulink." [Online]. Available: https://uk.mathworks.com/products/new_products/release2016a.html. [Accessed: 18-May-2020].

[222] "Where's Wally? - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Where%27s_Wally%3F. [Accessed: 18-May-2020].

[223] "Cookie Monster - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Cookie_Monster. [Accessed: 27-May-2020].

[224] "Multi-modal CPT." [Online]. Available: https://www.youtube.com/watch?v=3IWq2SJop1k. [Accessed: 07-Mar-2016].

[225] "Be first to get a Tobii EyeX Dev Kit Want to develop for Android or Linux ? Get a," vol. 595, p. 795.

[226] I. Software, "Muse : The Brain Sensing Headband Tech Spec Sheet."

[227] M. Teplan, "Fundamentals of EEG measurement," *Meas. Sci. Rev.*, vol. 2, pp. 1–11, 2002.

[228] "Tobii EyeX Controller – get your own eye tracker." 30-Jul-2015.

[229] "An introduction to Tobii EyeX | Tobii Developer Zone." [Online]. Available: http://developer.tobii.com/an-introduction-to-the-tobii-eyex-sdk/. [Accessed: 03-Mar-2016].

[230] "Specifications for EyeX – Tobii Eye Tracking Support." [Online]. Available: https://help.tobii.com/hc/en-us/articles/212818309-Specifications-for-EyeX. [Accessed: 10-Apr-2019].

[231] "Kinect hardware." [Online]. Available: https://dev.windows.com/en-us/kinect/hardware. [Accessed: 03-Mar-2016].

[232] Q. Wang, Q. Wang, and G. Kurillo, "Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect," no. November, 2015.

[233] "Welcome to our school | Oak Field School & Sports College." [Online]. Available: http://www.oakfieldsportscollege.org.uk/. [Accessed: 07-Mar-2016].

[234] National Joint Committee on Learning Disabilities, "Learning Disabilities: Issues on Definition," 1991.

[235] R. Paper, "Operationalizing the NJCLD Definition of Learning Disabilities for," vol. 40, 1998.

[236] "Teacher assessment: using P scales to report on pupils with SEN - Detailed guidance - GOV.UK." [Online]. Available: https://www.gov.uk/guidance/teacher-assessment-

using-p-scales. [Accessed: 07-Mar-2016].

[237] "P scales: attainment targets for pupils with SEN - Publications - GOV.UK," *Department for Education*. [Online]. Available: https://www.gov.uk/government/publications/p-scales-attainment-targets-for-pupils-with-sen. [Accessed: 07-Mar-2016].

[238] L. Hoste, B. Dumas, and B. Signer, "Mudra: A Unified Multimodal Interaction Framework," 2011, pp. 97–104.

[239] T. Affective and O. F. E. Emotion, "BRAIN – COMPUTER INTERFACE : FUNDAMENTALS AND ANALYSIS OF EEG-BASED EMOTION CLASSIFICATION."

[240] A. Bulling and T. O. Zander, "Cognition-Aware Computing," *IEEE Pervasive Comput.*, vol. 13, no. 3, pp. 80–83, 2014.

[241] T. Egner and M. B. Sterman, "Neurofeedback treatment of epilepsy: from basic rationale to practical application.," *Expert Rev. Neurother.*, vol. 6, no. 2, pp. 247–57, Feb. 2006.

[242] O. M. Bazanova and D. Vernon, "Interpreting EEG alpha activity.," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 94–110, 2013.

[243] M. Adjounadi, M. Cabrerizo, I. Yaylali, and P. Jayakar, "Interpreting EEG functional brain activity," *IEEE Potentials*, vol. 23, no. 1, pp. 8–13, 2004.

[244] M. Akcakaya *et al.*, "Noninvasive brain-computer interfaces for augmentative and alternative communication," *IEEE Rev. Biomed. Eng.*, vol. 7, no. c, pp. 31–49, 2014.

[245] J. G. Makin, D. A. Moses, and E. F. Chang, "Machine translation of cortical activity to text with an encoder–decoder framework," *Nat. Neurosci.*, 2020.

[246] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control.," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–91, Jun. 2002.

[247] M. Ahsan, M. Ibrahimy, and O. Khalifa, "EMG signal classification for human computer interaction: a review," *Eur. J. Sci. …*, vol. 33, no. 3, pp. 480–501, 2009.

[248] I. Daly *et al.*, "On the control of brain-computer interfaces by users with cerebral palsy," *Clin. Neurophysiol.*, vol. 124, no. 9, pp. 1787–1797, 2013.

[249] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 51, pp. 17849–54, 2004.

[250] M. Soleymani, S. Koelstra, I. Patras, and T. Pun, "Continuous Emotion Detection Using Eeg Signals and Facial Expressions," *Proc. 2011 IEEE Int. Conf. Autom. Face Gesture Recognit. Work.*, vol. 231287, no. 231287, pp. 803–808, 2011.

[251] A. Maye, D. Zhang, Y. Wang, S. Gao, and A. K. Engel, "Multimodal Brain-Computer Interfaces," *Tsinghua Sci. Technol.*, vol. 16, no. 2, pp. 133–139, Apr. 2011.

[252] W. Giang *et al.*, "Multimodal Interfaces: Literature Review of Ecological Interface

Design, Multimodal Perception and Attention, and Intelligent Adaptive Multimodal Interfaces," 2010.

[253] S. Koelstra *et al.*, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, 2012.

[254] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 211–223, 2012.

[255] T. O. Zander and C. Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general.," *J. Neural Eng.*, vol. 8, no. 2, p. 025005, Apr. 2011.

[256] E. Haselsteiner and G. Pfurtscheller, "Using time-dependent neural networks for EEG classification.," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 457–63, Dec. 2000.

[257] Y. P. Lin *et al.*, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, 2010.

[258] D. P.-O. O. Bos *et al.*, "Human-computer interaction for BCI games: Usability and user experience," *Proc. - 2010 Int. Conf. Cyberworlds, CW 2010*, pp. 277–281, Oct. 2010.

[259] C. Mühl *et al.*, "Bacteria Hunt," *J. Multimodal User Interfaces*, vol. 4, no. 1, pp. 11–25, Aug. 2010.

[260] C. Mühl *et al.*, "Bacteria Hunt: A multimodal, multiparadigm BCI game," *enterface2009*, no. 1, pp. 1–22, 2009.

[261] R. Horlings, D. Datcu, and L. J. M. Rothkrantz, "Emotion recognition using brain activity," *Proc. 9th Int. Conf. Comput. Syst. Technol. Work. PhD Students Comput. - CompSysTech '08*, no. March, p. II.1, 2008.

[262] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005.," *J. Neural Eng.*, vol. 2, no. 4, pp. L14-22, Dec. 2005.

[263] F. Lotte, M. Congedo, a Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces.," *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, Jun. 2007.

[264] C. H. Chuang, L. W. Ko, Y. P. Lin, T. P. Jung, and C. T. Lin, "Independent component ensemble of EEG for brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 2, pp. 230–238, 2014.

[265] F. Aloisea *et al.*, "Multimodal stimulation for a {P300}-based {BCI}," *Int. J. Bioelectromagn.*, vol. 9, no. July 2015, p. 128, 2007.

[266] K. Nakayama, "A Brain Computer Interface Based on FFT and Multilayer Neural Network - Feature Extraction and Generalization -," pp. 1–4, 2007.

[267] L. Čechovič and M. Hodoň, "P300 Evoked Potentials Data Classification Using Feed Forward Neural Network .," vol. 2, no. 2, pp. 2–6, 2013.

[268] Q. Dong, B. Hu, J. Zhang, X. Li, and M. Ratcliffe, "A Study on Visual Attention

Modeling--A Linear Regression method based on EEG," *2013 Int. Jt. Conf. Neural Networks*, pp. 1–6, Aug. 2011.

[269] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects.," *Clin. Neurophysiol.*, vol. 111, no. 10, pp. 1745–58, Oct. 2000.

[270] S. Bickel, E. C. Dias, M. L. Epstein, and D. C. Javitt, "Expectancy-related modulations of neural oscillations in continuous performance tasks," *Neuroimage*, vol. 62, no. 3, pp. 1867–1876, Sep. 2012.

[271] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Analysis and visualization of single-trial event-related potentials.," *Hum. Brain Mapp.*, vol. 14, no. 3, pp. 166–85, Nov. 2001.

[272] S. Makeig *et al.*, "Electroencephalographic brain dynamics following manually responded visual targets.," *PLoS Biol.*, vol. 2, no. 6, p. e176, Jun. 2004.

[273] J. D. R. Millán and J. Mouriño, "Asynchronous BCI and local neural classifiers: an overview of the Adaptive Brain Interface project.," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 159–61, Jun. 2003.

[274] V. B. Deepa and T. Nadu, "Classification of EEG data using FHT and SVM based on Bayesian Network," vol. 8, no. 5, pp. 239–243, 2011.

[275] M. Murugappan, "Classification of human emotion from EEG using discrete wavelet transform," *J. Biomed. Sci. Eng.*, vol. 03, no. April, pp. 390–396, 2010.

[276] M. Malkawi and O. Murad, "Artificial neuro fuzzy logic system for detecting human emotions," *Human-centric Comput. Inf. Sci.*, vol. 3, no. 1, p. 3, 2013.

[277] S. Sanei and J. A. Chambers, *EEG Signal Processing*. 2013.

[278] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey.," *Clin. Neurophysiol.*, vol. 118, no. 3, pp. 480–94, Mar. 2007.

[279] F. Hyder, *Dynamic Brain Imaging*, vol. 489. Totowa, NJ: Humana Press, 2009.

[280] H. H. Alwasiti, I. Aris, and A. Jantan, "Brain Computer Interface Design and Applications : Challenges and Future," *Appl. Sci.*, vol. 11, no. 7, pp. 819–825, 2010.

[281] H. S. Anupama, N. K. Cauvery, and G. M. Lingaraju, "Brain Computer Interface and Its Types - a Study," *Int. J. Adv. Eng. Technol.*, vol. 3, no. 2, pp. 739–745, 2012.

[282] N. H. Liu, C. Y. Chiang, and H. C. Chu, "Recognizing the degree of human attention using EEG signals from mobile sensors.," *Sensors (Basel).*, vol. 13, no. 8, pp. 10273–10286, Jan. 2013.

[283] P. Putman, B. Verkuil, E. Arias-Garcia, I. Pantazi, and C. van Schie, "EEG theta/beta ratio as a potential biomarker for attentional control and resilience against deleterious effects of stress on attention.," *Cogn. Affect. Behav. Neurosci.*, vol. 14, no. 2, pp. 782–91, Jun. 2014.

[284] R. B. Sangal and J. M. Sangal, "Use of EEG Beta-1 Power and Theta/Beta Ratio Over Broca's Area to confirm Diagnosis of Attention Deficit/Hyperactivity Disorder in Children," *Clin. EEG Neurosci.*, vol. 46, no. 3, pp. 177–182, Jun. 2014.

[285] P. F. Diez, A. G. Correa, L. Orosco, E. Laciar, and V. Mut, "Attention-level transitory response: a novel hybrid BCI approach.," *J. Neural Eng.*, vol. 12, no. 5, p. 56007, 2015.

[286] L. Trejo *et al.*, "EEG-based estimation of mental fatigue: Convergent evidence for a three-state model," *Found. Augment. Cogn.*, pp. 201–211, 2007.

[287] M. A. Nazari, F. Wallois, A. Aarabi, and P. Berquin, "Dynamic changes in quantitative electroencephalogram during continuous performance test in children with attention-deficit/hyperactivity disorder.," *Int. J. Psychophysiol.*, vol. 81, no. 3, pp. 230–6, Sep. 2011.

[288] B. Bracken, N. Palmon, L. Kellogg, S. Elkin-Frankston, and M. Farry, "A cross-domain approach to designing an unobtrusive system to assess human state and predict upcoming performance deficits," *Proc. Hum. Factors Ergon. Soc.*, pp. 706–710, 2016.

[289] C. Mühl, D. Heylen, and A. Nijholt, "Affective brain-computer interfaces: neuroscientific approaches to affect detection," *Handb. Affect. Comput.*, pp. 1–40, 2014.

[290] V. P. Oikonomou, a T. Tzallas, and D. I. Fotiadis, "A Kalman filter based methodology for EEG spike enhancement.," *Comput. Methods Programs Biomed.*, vol. 85, no. 2, pp. 101–8, Feb. 2007.

[291] V. P., A. T., S. Konitsiotis, D. G., and D. I., "The Use of Kalman Filter in Biomedical Signal Processing," in *Kalman Filter Recent Advances and Applications*, no. April, 2009.

[292] M. Arnold, W. H. Miltner, H. Witte, R. Bauer, and C. Braun, "Adaptive AR modeling of nonstationary time series by means of Kalman filtering.," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 5, pp. 553–62, May 1998.

[293] G. Pfurtscheller, C. Neuper, a Schlögl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters.," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 316–25, Sep. 1998.

[294] A. Kubler, E. Holz, T. Kaufmann, and C. Zickler, "A User Centred Approach for Bringing BCI Controlled Applications to End-Users," in *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, 2013.

[295] F. Morbidi, S. Member, A. Garulli, S. Member, C. Rizzo, and S. Rossi, "Application of Kalman filter to remove TMS-induced artifacts from EEG recordings," pp. 1–7.

[296] "Muse the Brain Sensing Headband User Guide."

[297] Trans Cranial Technologies Ltd., "10 / 20 System Positioning Manual," p. 20, 2012.

[298] "Headband Configuration Presets - Muse Developers." [Online]. Available: http://developer.choosemuse.com/hardware-firmware/headband-configuration-presets. [Accessed: 04-Mar-2016].

[299] A. Gevins and M. E. Smith, "Detecting transient cognitive impairment with EEG pattern recognition methods," *Aviat. Sp. Environ. Med.*, vol. 70, no. 10, pp. 1018–1024, Oct. 1999.

[300] T. C. Hankins and G. F. Wilson, “A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight.,” *Aviat. Space. Environ. Med.*, vol. 69, no. 4, pp. 360–7, Apr. 1998.

[301] D. Ming *et al.*, “Electroencephalograph (EEG) signal processing method of motor imaginary potential for attention level classification,” *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009*, vol. 2009, pp. 4347–4351, Jan. 2009.

[302] T. McMahan, I. Parberry, and T. D. Parsons, “Evaluating Player Task Engagement and Arousal Using Electroencephalography,” *Procedia Manuf.*, vol. 3, no. Ahfe, pp. 2303–2310, 2015.

[303] A. Ameera, A. Saidatul, and Z. Ibrahim, “Analysis of EEG Spectrum Bands Using Power Spectral Density for Pleasure and Displeasure State,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 557, no. 1, 2019.

[304] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, Second. Pearson, 1999.

[305] J. Musson and J. Li, “A Comparative Survey of PSD Estimation Methods for EEG Signal Analysis,” *Student Capstone Conf. Proc.*, pp. 1–6, 2010.

[306] A. Doshi and M. M. Trivedi, “Head and gaze dynamics in visual attention and context learning,” *2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009*, pp. 77–84, 2009.

[307] Y. Matsumoto, T. Ogasawara, and A. Zelinsky, “Behavior recognition based on head pose and gaze direction measurement,” in *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000) (Cat. No.00CH37113)*, 2000, vol. 3, pp. 2127–2132.

[308] S. Brams *et al.*, “Does effective gaze behavior lead to enhanced performance in a complex error-detection cockpit task?,” *PLoS One*, vol. 13, no. 11, pp. 1–25, 2018.

[309] A. Thomas, “Communication Desvices for the Nonvocal Disabled,” 1981.

[310] P. M. Räihä and Kari-Jouko, “Twenty Years of Eye Typing: Systems and Design Issues.” .

[311] P. Simon, F. Rousseau, and J.-C. Angue, “Quantitative analysis of mental workload influence on eye scanning movements,” *Syst. Man Cybern.*, vol. 4, pp. 707–712, 1993.

[312] S. Chen, J. Epps, N. Ruiz, and F. Chen, “Eye activity as a measure of human mental effort in HCI,” *Int. Conf. Intell. User Interfaces, Proc. IUI*, no. August 2014, pp. 315–318, 2011.

[313] S. Asteriadis, K. Karpouzis, and S. Kollias, “Robust validation of Visual Focus of Attention using adaptive fusion of head and eye gaze patterns,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 414–421, 2011.

[314] A. Borji and L. Itti, “State-of-the-art in visual attention modelling,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185–207, 2013.

[315] N. Veliyath, P. De, A. A. Allen, C. B. Hodges, and A. Mitra, “Modeling students’ attention in the classroom using eyetrackers,” in *ACMSE 2019 - Proceedings of the 2019*

ACM Southeast Conference, 2019, pp. 2–9.

[316] D. Rosengrant, D. Hearrington, K. Alvarado, and D. Keeble, “Following student gaze patterns in physical science lectures,” *AIP Conf. Proc.*, vol. 1413, pp. 323–326, 2012.

[317] J. R. Anderson, *Cognitive Psychology an Its Implications*, 7th ed. New York, New York, USA: Worth Publishers, 2005.

[318] D. LaBerge, R. L. Carlson, J. K. Williams, and B. G. Bunney, “Shifting attention in visual space: Tests of moving-spotlight models versus an activity-distribution model.,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 23, no. 5, pp. 1380–1392, 1997.

[319] P. C. P. Pcpantelindianaedu and D. P. K. Dpkindianaedu, “Assessing a Bayesian account of human gaze perception,” pp. 1811–1816, 2015.

[320] G. Sperling and E. Weichselgartner, “Episodic theory of the dynamics of spatial attention.,” *Psychol. Rev.*, vol. 102, no. 3, pp. 503–532, 1995.

[321] A. C. Schütz, D. I. Braun, and K. R. Gegenfurtner, “Eye movements and perception: A selective review.,” *J. Vis.*, vol. 11, no. 5, pp. 1–30, 2011.

[322] C. Ehmke and S. Wilson, “Identifying web usability problems from eye-tracking data,” *Hci 2007*, pp. 119–128, 2007.

[323] J. A. Vigo, “Eye Gaze Tracking for Tracking Reading Progress,” p. 100, 2013.

[324] P. Fattahi, “Eyetracking Definitions: Basic Definition of Terms,” Teltow, Germany, 2011.

[325] D. C. Newton and W. R. Knecht, “Current Status of Gaze Control Research and Technology Literature Review,” *DOT Natl. Transp. Integr. Search - ROSA P*, no. January, 2017.

[326] P. Majaranta and K. J. Räihä, “Twenty years of eye typing: Systems and design issues,” *Eye Track. Res. Appl. Symp.*, no. January, pp. 15–22, 2002.

[327] M. B. Almourad, E. Bataineh, J. Stocker, and F. Marir, “Analyzing the Behavior of Autistic and Normal Developing Children Using Eye Tracking Data,” in *Advances in Intelligent Systems and Computing*, 2018, vol. 739, pp. 340–349.

[328] “EyeX controller accurary and precision test report | Tobii Developer Zone.” [Online]. Available: http://developer.tobii.com/community/forums/topic/eyex-controller-accurary-and-precision-test-report/#post-3075. [Accessed: 03-Mar-2016].

[329] “Eye tracking device integration.” 29-Jul-2015.

[330] “Specification of the Tobii EyeX Controller | Tobii Developer Zone.” [Online]. Available: http://developer.tobii.com/community/forums/topic/specification-of-the-tobii-eyex-controller/#post-1156. [Accessed: 03-Mar-2016].

[331] A. Kleinsmith and N. Bianchi-Berthouze, “Affective body expression perception and recognition: A survey,” *IEEE Trans. Affect. Comput.*, vol. 4, no. 1, pp. 15–33, Jan. 2013.

[332] A. Patwardhan and G. Knapp, “Multimodal Affect Recognition using Kinect,” 2016.

[333] N. Bianchi-Berthouze and Nadia Bianchi-Berthouze, "Understanding the role of body movement in player engagement," *Human-Computer Interact.*, vol. 28, no. 1, pp. 40–75, 2013.

[334] N. Bianchi-Berthouze, "The Affective Body Argument in Technology Design," in *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*, 2016, pp. 3–6.

[335] J. Farley, E. F. Risko, and A. Kingstone, "Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering," *Front. Psychol.*, vol. 4, no. SEP, p. 619, Sep. 2013.

[336] M. D. Smith and M. S. Barrett, "The Effect of Parent Training on Hyperactivity and Inattention in Three School-Aged Girls with Attention Deficit Hyperactivity Disorder," *Child Fam. Behav. Ther.*, vol. 24, no. 3, pp. 21–35, Oct. 2002.

[337] C. K. Conners, L. Eisenberg, and A. Barcai, "Effect of Dextroamphetamine on Children," *Arch. Gen. Psychiatry*, vol. 17, no. 4, p. 478, Oct. 1967.

[338] P. L. Holborow and P. S. Berry, "Hyperactivity and Learning Difficulties," *J. Learn. Disabil.*, vol. 19, no. 7, pp. 426–431, Aug. 1986.

[339] E. Benedetto-Nasho and R. Tannock, "Math computation, error patterns and stimulant effects in children with Attention Deficit Hyperactivity Disorder," *J. Atten. Disord.*, vol. 3, no. 3, pp. 121–134, Oct. 1999.

[340] S. L. Smith, *No easy answer : the learning disabled child at home and at school*. Bantam Books, 2012.

[341] B. Fenesi, K. Lucibello, J. A. Kim, and J. J. Heisz, "Sweat So You Don't Forget: Exercise Breaks During a University Lecture Increase On-Task Attention and Learning," *J. Appl. Res. Mem. Cogn.*, vol. 7, no. 2, pp. 261–269, 2018.

[342] S. Lis, N. Baer, C. Stein-En-Nosse, B. Gallhofer, G. Sammer, and P. Kirsch, "Objective measurement of motor activity during cognitive performance in adults with attention-deficit/hyperactivity disorder," *Acta Psychiatr. Scand.*, vol. 122, no. 4, pp. 285–294, 2010.

[343] A. M. F. Zinno, G. Douglas, S. Houghton, V. Lawrence, J. West, and K. Whiting, "Hyperactivity Disorder (ADHD) during computer video game play," *Br. J. Educ. Technol.*, vol. 32, no. 5, pp. 607–618, 2001.

[344] A. Pimenta, D. Carneiro, P. Novais, and J. Neves, "Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 8073 LNAI, pp. 222–231.

[345] M. F. Uddin, "Addressing Accuracy Paradox Using Enhanched Weighted Performance Metric in Machine Learning," pp. 319–324, 2020.

[346] F. J. Valverde-Albacete and C. Peláez-Moreno, "100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox," *PLoS One*, vol. 9, no. 1, 2014.

[347] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.

[348] P. C. Austin and E. W. Steyerberg, "Interpreting the concordance statistic of a logistic regression model: Relation to the variance and odds ratio of a continuous explanatory variable," *BMC Med. Res. Methodol.*, vol. 12, pp. 1–8, 2012.

[349] L. E. Dodd and M. S. Pepe, "Partial AUC estimation and regression," *Biometrics*, vol. 59, no. 3, pp. 614–623, 2003.

[350] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *Ann. Math. Stat.*, vol. 18, no. 1, pp. 50–60, 1947.

[351] S. J. Mason and N. E. Graham, "Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves," *Q. J. R. Meteorol. Soc.*, vol. 128, pp. 2145–2166, 2002.

[352] A. J. Viera and J. M. Garrett, "Understanding Interobserver Agreement: The Kappa Statistic," *Fam. Med.*, vol. 37, no. 5, pp. 360–3, 2005.

[353] Kampakis Stylianos, "Performance Measures: Cohen's Kappa statistic - The Data Scientist," *The Data Scientist*, 2016. [Online]. Available: https://thedatascientist.com/performance-measures-cohens-kappa-statistic/. [Accessed: 18-May-2020].

[354] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[355] J. Orlof and J. Bloom, "Maximum Likelihood Estimates Class," in *Introduction to probability and statistics*, MIT Press, 2005, p. 2.

[356] Salford-systems, "Part 10: Multinomial Classification- Log Likelihood," *YouTube*, 2014. [Online]. Available: https://www.youtube.com/watch?v=Q8EhwV3Aj7U. [Accessed: 18-May-2020].

[357] P. D.- IEEE/AIP, undefined May, and undefined 2007, "Python: Batteries Included, volume 9 of Computing in Science & Engineering."

[358] K. Milmann, M. A.- IEEE/AIP, undefined March, and undefined 2011, "Scientific Python, volume 11 of Computing in Science and Engineering."

[359] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," 2011.

[360] T. Zito, N. Wilbert, L. Wiskott, and P. Berkes, "Modular toolkit for data processing (MDP): A python data processing framework," *Front. Neuroinform.*, vol. 2, no. JAN, pp. 1–7, 2009.

[361] J. Bayer, "Pybrain PyBrain," no. May 2014, 2010.

[362] "NumPy." [Online]. Available: https://numpy.org/. [Accessed: 25-May-2020].

[363] "SciPy.org — SciPy.org." [Online]. Available: https://www.scipy.org/. [Accessed: 25-May-2020].

[364] M. Hanke, Y. O. Halchenko, P. B. Sederberg, S. J. Hanson, J. V. Haxby, and S. Pollmann, "PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data," *Neuroinformatics*, vol. 7, no. 1, pp. 37–53, Mar. 2009.

[365] "R: The R Project for Statistical Computing." [Online]. Available: https://www.r-project.org/. [Accessed: 25-May-2020].

[366] "Neuroimaging in Python — NiBabel 3.1.0+3.g1660b1a1 documentation." [Online]. Available: https://nipy.org/nibabel/. [Accessed: 25-May-2020].

[367] "Shogun Machine Learning - Home." [Online]. Available: https://www.shogun-toolbox.org/. [Accessed: 25-May-2020].

[368] C. C. Chang and C. J. Lin, "LIBSVM: A Library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011.

[369] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, 2008.

[370] T. K. Ho, "Random Decision Forests Tin Kam Ho Perceptron training," *Proc. 3rd Int. Conf. Doc. Anal. Recognit.*, pp. 278–282, 1995.

[371] L. Breiman, "Random forests," in *Random Forests*, 2001, pp. 1–122.

[372] A. Cutler, D. R. Cutler, and J. R. Stevens, "Ensemble Machine Learning," *Ensemble Mach. Learn.*, no. January, 2012.

[373] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *Neuroimage*, vol. 65, pp. 167–175, 2013.

[374] S. Hor and M. Moradi, "Scandent Tree: A Random Forest Learning Method for Incomplete Multimodal Datasets," Springer, Cham, 2015, pp. 694–701.

[375] "scikit-learn: machine learning in Python — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/index.html. [Accessed: 23-May-2020].

[376] "3.2.4.3.1. sklearn.ensemble.RandomForestClassifier — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. [Accessed: 23-May-2020].

[377] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm," *Proc. 13th Int. Conf. Mach. Learn.*, pp. 148–156, 1996.

[378] "sklearn.ensemble.AdaBoostClassifier — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html?highlight=adaboost#sklearn.ensemble.AdaBoostClassifier. [Accessed: 23-May-2020].

[379] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class AdaBoost," *Stat. Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[380] "1.17. Neural network models (supervised) — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html. [Accessed: 23-May-2020].

[381] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," Mar. 2018.

[382] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.

[383] A. A. Ahmadi, A. Olshevsky, P. A. Parrilo, and J. N. Tsitsiklis, "NP-hardness of Deciding Convexity of Quartic Polynomials and Related Problems," *Math. Program.*, vol. 137, no. 1–2, pp. 453–476, Dec. 2010.

[384] "sklearn.neighbors.KNeighborsClassifier — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html?highlight=kneighborsclassifier#sklearn.neighbors.KNeighborsClassifier. [Accessed: 23-May-2020].

[385] T. Bayes, "An essay towards solving a problem in the doctrine of chances. 1763.," *MD. Comput.*, vol. 8, no. 3, pp. 157–171, May 1991.

[386] "Orange Data Mining - Data Mining." [Online]. Available: https://orange.biolab.si/. [Accessed: 23-May-2020].

[387] "orange3/naive_bayes.py at 04b7c72740ee5ced370aa8104152c38e3245e2b1 · biolab/orange3." [Online]. Available: https://github.com/biolab/orange3/blob/04b7c72740ee5ced370aa8104152c38e3245e2b1/Orange/classification/naive_bayes.py. [Accessed: 23-May-2020].

[388] J. H. Friedman, R. Kohavi, and U. Uun, "Lazy Decision Trees," 1996.

[389] "sklearn.tree.DecisionTreeClassifier — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier. [Accessed: 23-May-2020].

[390] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2001.

[391] "sklearn.svm.LinearSVC — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC. [Accessed: 23-May-2020].

[392] "sklearn.dummy.DummyClassifier — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html. [Accessed: 22-May-2020].

[393] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, vol. 64, no. 6. 1974.

[394] P. Y. Chen, M. Smithson, and P. M. Popovich, *Correlation: Parametric and Nonparametric Measures*, vol. 1. 2002.

[395] K. F. Weaver, V. C. Morales, S. L. Dunn, K. Godde, and P. F. Weaver, *An Introduction to Statistical Analysis in Research: With Applications in the Biological and Life Sciences*. 2017.

[396] D. H. Sykes, V. I. Douglas, and G. Morgenstern, “Sustained Attention in Hyperactive Children,” *J. Child Psychol. Psychiatry*, vol. 14, no. 3, pp. 213–220, 1973.

[397] J. N. Epstein, C. K. Conners, G. Sitarenios, and D. Erhardt, “Continuous performance test results of adults with attention deficit hyperactivity disorder,” *Clin. Neuropsychol.*, vol. 12, no. 2, pp. 155–168, 1998.

[398] F. C. Bartlett and C. Burt, “Remembering: a Study in Experimental and Social Psychology,” *Br. J. Educ. Psychol.*, vol. 3, no. 2, pp. 187–192, 1933.

[399] M. W. Kibby, “The organization and teaching of things and the words that signify them,” *J. Adolesc. Adult Lit.*, vol. 39, no. 3, pp. 208–223, 1995.

Chapter 8. PHD PAPERS AS FIRST AUTHOR

IRC 2020

XIV. International Research Conference

Certificate of Best Paper Award

ICACII 2020 : XIV. International Conference on
Affective Computing and Intelligent Interaction
hereby certifies that the below mentioned paper
has been selected as the best paper

Modeling Engagement with Multimodal Multisensor Data Using the Continuous Performance Test
as an Objective Tool to Track Flow
Mohammad H. Taheri, David J. Brown, Nasser Sherkat

PARIS, FRANCE

INTERNATIONAL RESEARCH CONFERENCE
IRC

Authorized Signature

MAY, 14-15, 2020

Modeling Engagement with Multimodal Multisensor Data: The Continuous Performance Test as an Objective Tool to Track Flow

Mohammad H. Taheri, David J. Brown, Nasser Sherkat

***Abstract*—**Engagement is one of the most important factors in determining successful outcomes and deep learning in students. Existing approaches to detect student engagement involve periodic human observations that are subject to inter-rater reliability. Our solution uses real-time multimodal multisensor data labeled by objective performance outcomes to infer the engagement of students. The study involves four students with a combined diagnosis of cerebral palsy and a learning disability who took part in a 3-month trial over 59 sessions. Multimodal multisensor data were collected while they participated in a continuous performance test. Eye gaze, electroencephalogram, body pose, and interaction data were used to create a model of student engagement through objective labeling from the continuous performance test outcomes. In order to achieve this, a type of continuous performance test is introduced, the Seek-X type. Nine features were extracted including high-level handpicked compound features. Using leave-one-out cross-validation, a series of different machine learning approaches were evaluated. Overall, the random forest classification approach achieved the best classification results. Using random forest, 93.3% classification for engagement and 42.9% accuracy for disengagement were achieved. We compared these results to outcomes from different models: AdaBoost, decision tree, k-Nearest Neighbor, naïve Bayes, neural network, and support vector machine. We showed that using a multisensor approach achieved higher accuracy than using features from any reduced set of sensors. We found that using high-level handpicked features can improve the classification accuracy in every sensor mode. Our approach is robust to both sensor fallout and occlusions. The single most important sensor feature to the classification of engagement and distraction was shown to be eye gaze. It has been shown that we can accurately predict the level of engagement of students with learning disabilities in a real-time approach that is not subject to inter-rater reliability, human observation or reliant on a single mode of sensor input. This will help teachers design interventions for a heterogeneous group of students, where teachers cannot possibly attend to each of their individual needs. Our approach can be used to identify those with the greatest learning challenges so that all students are supported to reach their full potential.

***Keywords*—**Affective computing in education, affect detection, continuous performance test, engagement, flow, HCI, interaction, learning disabilities, machine learning, multimodal, multisensor, physiological sensors, Signal Detection Theory, student engagement.

M. H. Taheri is a Research Fellow at the Interactive Systems Research Group at Nottingham Trent University, Nottingham, NG14FQ, UK (e-mail: mohammad.taheri@ntu.ac.uk).

D. J. Brown is the Research Group Leader of the Interactive Systems Research Group at Nottingham Trent University, Nottingham, NG14FQ, UK (e-mail: david.brown@ntu.ac.uk).

N. Sherkat is a Professor and Head of the Engineering and Mathematics Department at Sheffield Hallam University, Sheffield, S11WB, UK (e-mail: n.sherkat@shu.ac.uk).

I. Introduction

It is often a challenge to keep children engaged in learning activities, especially if the activity requires them to retain focus and active participation for a continuous period of time. Researchers reported that students with learning disabilities do not display any significant attention deficiency compared to non-disabled students – these students can complete the same activities if given more processing time [1]. Despite this outcome, student engagement can vary greatly depending on the activity, and understanding when the student is engaged, and when they are not, is not a straightforward task.

While research has focused significantly on the ability of children with learning difficulties to recognize [2], perceive [3] and interpret [4] emotional cues, there is little to no research on the recognition of the emotional state of these students. The importance of carers being able to interpret the emotional cues and states of such students has been documented in [5]. It is found that carers made significantly more critical and 'fundamental attribution' [6] errors in the emotional expression of their clients with learning disabilities in comparison to their clients without learning disabilities. This affects the quality and quantity of their client's treatment [5] and has a negative effect on the provisional treatment [7], [8]. Currently, carers rely on their expert understanding and personal experience of the students to interpret their voices, expressions, and gestures. Dependent on the personal experience with a particular client, a carers' internal modeling of the emotional expression of that client can vary widely and demonstrate inter-rater reliability issues.

One of the main ways to measure engagement in students with special educational needs is to use the Special Schools and Academies Trust (SSAT) Engagement Scale [9]. The Engagement Profile and Scale is a classroom tool developed through SSAT's research into effective teaching and learning for children with complex learning difficulties and disabilities. It allows educators to focus on the child's engagement as a learner and create personalized learning pathways [10]. The authors describe seven components of engagement namely, awareness, curiosity, investigation, discovery, anticipation, persistence, and initiation. Teachers assign a score out of four for each component giving a total score out of 28. One potential issue with the use of this scale is that teachers assign a subjective rating to each component, which will be subject to inter-rater variability.

The scale has been used to assess the impact of new technologies in special education – especially in studies investigating the suitability of humanoid robots to support learning in students with Profound and Multiple Learning Disabilities (PMLD). The approach of using an engagement scale to create personalized learning pathways has been examined by others [11]-[13].

One way to overcome the variation in observer inter-rater reliability in tracking emotional expression is to introduce a reliable indicator of that emotion. In this research, a robust methodology for tracking engagement levels of children with PMLD or Cerebral Palsy (CP) is proposed using Signal Detection Theory (SDT) [14]. The application of this theory gives quantifiable information on the improvement of deterioration or attention in response to a Continuous Performance Test (CPT) specifically adapted to the abilities of such students [1], [15]. Performance in this test will provide objective labels to train machine learning algorithms using sensor data (e.g., on eye gaze and body pose) collected whilst the students are interacting with a PC. After obtaining a labeled dataset, machine learning models can be applied to the data so that in the future new unlabeled data can be presented to the model and engagement can be inferred.

Many traditional interactive systems use devices such as a keyboard and mouse and are constructed to emphasize the transmission of explicit messages while ignoring implicit information about user interaction. The emerging science of affective computing can only be accelerated with the abundance of sensor data [16], [17] and wearables [18]. These multimodal human cues [19]-[21] provide the multimodal multisensor data points necessary for enhanced emotional modeling. Multimodal multisensor data have been instrumental in determining user affective states [19], [22]-[28] including engagement [29]-[31]. There are a number of challenges to develop such a model including understanding the relationship between the terms used in educational contexts (e.g., 'flow' and 'engagement'), developing appropriate CPTs suitable for the abilities of students with the most profound learning disabilities, selection of appropriate sensors and features derived from these data streams from which emotional states can be inferred, finding a suitable population of end-users to collect data with to train the machine learning algorithms, and finally comparing the performance of a range of machine learning methods to infer flow and engagement. This paper addresses each of these challenges.

II. Engagement, Flow and Learning

In education, the use of the term 'engagement' is more familiar to teachers than flow. D'Mello and Graesser [32] see considerable overlap between the two terms: "we conceptualize engagement/flow as a state of engagement with a task such that concentration is intense, attention is focused, and involvement is complete" (p.146). Contrary to engagement, the concept of flow is well defined in Csíkszentmihályi's works [33], [34]. One is in flow when one is engaged [31], and steady performance has been maintained at the comfortable limits of one's skill limitations [35], [36] for the duration of time - making flow the optimal psychological state of engagement. This results in immersion, concentrated focus and deep learning [37], [38]. The relationship between flow and engagement has been illustrated in Bianchi-Berthouze's [31] engagement model, a simplified version that has been shown in (1):

$$Attention \rightarrow Flow \rightarrow Engagement \quad (1)$$

Performance trend tracking can be used as an indicator of flow [36]. This approach has been used in [39]-[41] as a model for relating affect (flow/engagement) to user performance in a pre-defined activity/task challenge.

Engagement's crucial role in learning was recognized by Carpenter [41], stating that "Sustainable learning can occur only when there is meaningful engagement". Learner engagement in the classroom is the single, most reliable indicator of deep learning [36], [42], [43] and learner satisfaction [33], [37], [38]. Its role is central to classroom performance and the achievement of learning outcomes [45]-[48]. For these reasons, flow, a sub-state of Engagement [31], [33], [48], is a more suitable measure to follow or track the quality of experience; firstly it can be objectively monitored, and secondly, through its monitoring, engagement is also established. Flow is the optimal state of engagement, where engagement meets productivity [37], [38]. Maintaining flow in learning is especially significant because it is the most reliable indicator for determining successful learning [36], [45]-[48]. In the absence of learner engagement, deep conceptual learning is also not present [47], [49], which is an essential attribute to long-term learning and new skill achievement [49].

III. Challenges in Understanding Engagement in Students with Learning Disabilities

Abrams stated [50], "The vast majority of children with learning disabilities have some emotional problem associated with the learning difficulty." Generally, however, teachers have prioritized the diagnosis and remediation of learning disabilities [51].

Studies have considered self-reported affect states as the ground truth for inter-rater agreement studies [52], [53].

These studies have looked at the level agreement and correlation between self-rated affect states and peers, clinicians, and long-term partners. The level of correlation even though significant between the 40th and 70th percentiles [52], [53], still leaves room for improvement. In addition, self-rated affect states may carry bias or not be representative of the true affect state. Therefore, an automated method that would base its ground truth on self-rated affect states would thus be impacted by such bias and unknown reliability factors. The validity of a machine learning method based on clinician, or peer-rated affect states would inherit even greater bias, reliability, and interrater reliability uncertainty, as it is one more level separated. Importantly, a machine learning method with 100% classification accuracy trained with clinician-rated affect data would at best achieve around 70% correctness of the self-rated affect states. Furthermore, the self-rated affect states may themselves have a bias or be unrepresentative. This creates a problem for both the clients and care workers as it has been shown that observation is not a reliable method of determining a person's mood and affect state [6]. This can only be more intensified with PMLD and CP users, as their behaviors, body language and voice may not have the same cues as mainstream people. Moreover, the levels of skill and experience between care workers and teachers vary widely, as does their capacity and accuracy of interpretation of others' behaviors. This uncertainty of interpretation and inaccuracy in the observation of the affect state of a person experiencing PMLD or CP (mood and emotional well-being) can be detrimental to their quality of life [5], [7], [8]. Hence, the well-being of a student with PMLD or CP can be improved if their levels of interest and engagement could be determined and tracked by more independent and repeatable means, such as using technology, and in our case sensors. This added interpretation of a student's state of affect is not meant to replace teachers' or carers' interpretation, but more to augment this judgment.

Monitoring a person's level of interest and engagement in activity allows carers, teachers, and parents to be responsive to those levels. In this study, we investigate the ability of sensor-based technology to detect and track sustained attention in a repetitive demanding activity, with a multimodal multisensor platform. This allows us to make inferences on the attention level of the student throughout the length of this activity through their responses to the challenges presented in the repetitive activity.

An objective approach to the reporting of engagement is the use of a standardized test to monitor for indicators of flow. We demonstrate the possibility of tracking and then modeling body movements, eye gaze, electroencephalogram (EEG) and interaction data from students with PMLD and CP to estimate their level of engagement, as a good indicator of what interests them and positively influences the quality of that experience.

IV. A Platform to Measure Engagement Using Multimodal Multisensor Data for PMLD

A gamified *platform* is proposed that monitors the qualities of flow, namely engagement through performance tracking using SDT [14] measures and outcomes. For the remainder of the paper, we will refer to this engagement tracking platform as *'the platform'*.

The participant is required to pay continuous *attention* to a computer screen where an *interactive* game provides them with a pre-defined signal detection *challenge*. The participant is in *control* of the response they give, and *feedback* is given to them regarding the correctness of their response to the *challenge*. This is the basis for Swanson's CPT [15]. The CPT is an integral component of *the platform,* and we have therefore created a version, the 'Seek-X' type. This test has been created to be used specifically as *an objective tool* for engagement tracking using the CPT test outcomes to label multisensor data.

We have named this CPT 'Seek-X type' because the participant is asked to *seek* the target image between other non-target images acting as a matrix of noise. 'Type-Seek-X' exercises engage eye gaze as a crucial element of answering the SDT challenge. The Seek-X type CPT is of the non-rare target type, see (2):

$$\text{CPT test types} = \begin{cases} By\ challenge \begin{cases} Type - X \longrightarrow Seek - X \\ Type - AX \longrightarrow Seek - AX \end{cases} \\ By\ target\ frequency \begin{cases} Rare \\ Non - rare \end{cases} \end{cases} \quad (2)$$

In summary, the period of sustained engagement is marked by participants' *attention* and *interest* being maintained in an interactive interaction. Maintaining sustained attention indicates the key foundation for recognizing lasting engagement. For this reason, this work explores classical methods for attention tracking using a neuropsychological test that measures a person's sustained and selective attention (the CPT) [15]. The CPT is reported to be the most popular measure of sustained attention or vigilance—the ability to sustain attentional focus and remain alert to stimuli over time [54], [55]. The first attempt to objectively evaluate the relationship between maintaining attention in students with learning disabilities using CPTs was introduced by Swanson in [15], [56] and later expanded by Eliason and Richman [1]. Using SDT [14], [15], [58]-[62], quantifiable objective data on the improvement or deterioration of attention are collected and analyzed using SDT detailed in [58], [59].

A. Data Collection

Four students were recruited for data collection (see *Participants*). They took part in an 11-week long study with up to four sessions weekly, depending on participant availability.

Each session included 48 challenges. Each test lasted between 6-32 minutes depending on participant readiness or other setting-up challenges. Every session recorded nearly 4 minutes of data. A total of 59 sessions of the CPT test were carried out (average of 15 sessions per participant). A series of 48 slides with pauses in between were displayed for each participant.

This CPT test design was based on Rosvald and Mirsky's original paper [61]. Recommended time alterations to the experiment length were made to match the shorter length activities that students with PMLD are accustomed to at school [15]. The CPT test was therefore shortened to about 4 minutes for our participants, and the whole process takes around 15 minutes. This is compared to other research, which suggests a 30-minute test for neurotypical participants [62].

The difficulty of the CPT was also adapted for each participant by making the maximum response time (slide display time + blank slide display time) shorter or longer or by adjusting the image matrix grid size. These times are initially 1.8 s and 1 ± 0.1 s, respectively, and are increased or decreased depending on participant capacity. These times (seen in Table I) were established in a series of pilot tests where the aim was to reach close to the 85% rule for learning, where the participant makes around 15% mistakes and 85% correct responses [35] when in flow. The Seek-X type CPT slide timeline is demonstrated in Fig. 1.

It is important in SDT that the participant can demonstrate they understand the difference between the target and noise, given enough time. To establish this, the game objective was re-introduced to the participants at the start of every session using a paper-based mockup to test the participants' understanding of the challenge and validate their response.

TABLE 1
CPT SETTINGS ADJUSTED PER PARTICIPANT CAPACITY

Participant alias	P scales mean	Slide display time / Stimulus duration (s)	Blank slide display time / Interstimulus interval (s)
Will	6.93	1.8	1.1 ± 0.1
Jen	19.45*	1	1.1 ± 0.1
Mark	3.7	8	2.1 ± 0.1
Rick	6.76	1.8	1.1 ± 0.1

*Jen is enrolled in the National Curriculum.

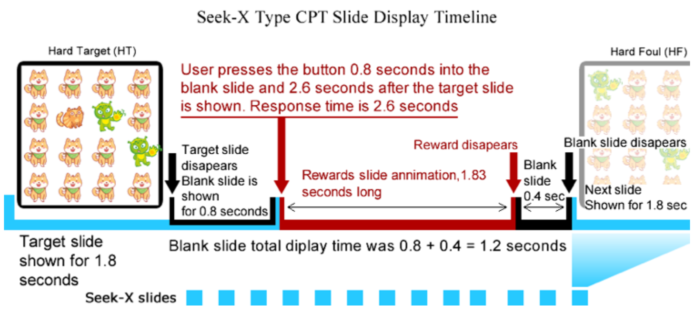


Fig. 1. Seek-X type CPT slide timeline

B. Experimental platform and the CPT

The platform tracks student performance in a repetitive game, which rewards them with exciting visual and audio feedback when they answer correctly, but ultimately fatigues the student by being exhausting over a long period. The student is required to pay attention to the game dynamic, which challenges them to pay selective and sustained attention to the elements on the screen and respond appropriately. This induces different states of affect, with lower levels of valence, as the game carries on and the students' attention capacity naturally decreases. During this game, real-time multimodal multisensor data is collected within the experimental platform, which is used later to create a machine learning model of flow. The experimental platform was developed in MATLAB to collect data from various consumer-grade sensor hardware. The experimental platform and the relative student position are visualized in Fig. 2.

The new type of CPT, of type 'Seek-X' was designed for this study. Each slide has a mixture of three images, comprising of the target image, the target imitation and the contrast image, as seen in Fig. 3. The target imitation bears a close resemblance to the target image (similar colors, general shape), however, the contrast image can easily be identified.

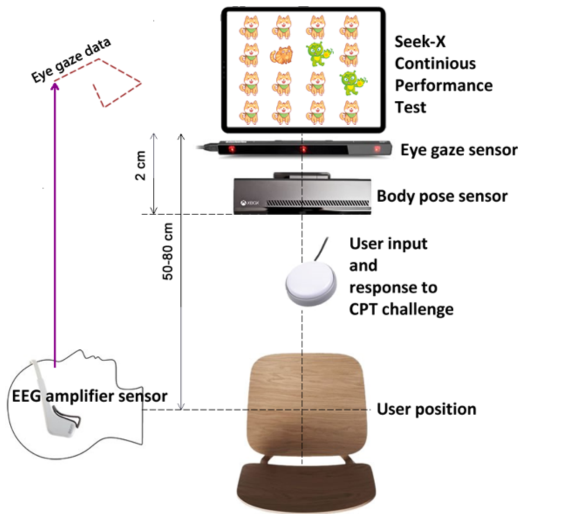


Fig. 2. The multimodal multisensor experimental platform with the eye gaze, body pose, EEG sensor and the CPT.

Target image Target imitation image Contrast image

Fig. 3. CPT image types

The ratio of the mixture of the main image to the filler image in all slide types is always 9 to 1 or as close as possible to this ratio, depending on the grid size and limited spaces available. We found that for our test user group a grid of 4 x 4 introduced enough difficulty to allow for participant responses, without being so easy that the participant would not make any mistakes when fatigued.

The distribution of the Hard Target (HT) pattern among the other random patterns has an occurrence probability of 50%. The other CPT occurrences are standardized [61] as Hard Foul (HF), Easy Target (ET) and Easy Foul (EF). These patterns and their corresponding labels are seen in Table 2.

TABLE 2
THE DISTRIBUTION OF PATTERNS IN THE SEEK-X TEST

Pattern	HT	HF	ET	EF
Distribution	50%	25%	12.5%	12.5%
CPT label	Hard Target: Target image mixed in with imitations targets with a few contrast images.	Hard Foul: Imitation target images with some contrast images.	Easy Target: Target Image mixed in contrast images with some imitation targets.	Easy Foul: Contrast images with some imitation targets.

The participants were seated in a chair in front of a 20” computer monitor, at a controlled distance of 50 cm to 80 cm from the screen. Each participant was asked to press the keyboard spacebar, or a big button if wheelchair-bound, whenever they saw the target image on the screen, and not to press the button when they did not see the target image on the screen. During this activity, participant eye gaze, body pose, EEG measurements and button interaction data were continuously recorded.

The participant was then presented with 48 instances of images displayed in a controlled random sequence on the screen. Each image was displayed for a stimulus duration (slide display time) followed by a blank slide displayed for an interstimulus interval.

Real-time eye gaze position using Tobii EyeX [63], body pose data using Kinect v2, EEG data from the Muse headband [64] and interaction data from the USB button is recorded in MATLAB [65]. The Muse EEG headband streams 16-bit voltage data in microvolt (µV) units at 500 Hz, which is equal or comparable to medical-grade EEG specifications [66]. The Tobii EyeX eye gaze tracking controller [67] uses near-infrared light to track the eye movements and gaze point of a student [68]. It works in variable light conditions and allows for student head movement while maintaining accuracy, which is crucial for our target user group. It has a frequency of 70 Hz and uses backlight assisted near-infrared (NIR 850 nm + red light (650 nm)) to achieve a 95% tracking population [69]. The Kinect 2 sensor [70] is a motion-sensing peripheral for body tracking. Using structured light and machine learning it can infer body position [70]. Kinect 2 is reported with an average depth accuracy of under 2 mm in the central viewing angle and increases to 2-4 mm in the range of up to 3.5 mm [71]. The furthest distance captured by Kinect 2 is 4.5 mm, where the average error typically increases beyond 4 mm. The experimental platform was designed to replicate the majority of the CPT test variations reported in relevant studies [1], [15], [61], [72]–[76]. The features extracted from these sensor data streams are described under feature extraction.

C. Participants

Four participants with PMLD were recruited to collect labeled sensor data whilst using the gamified platform. These four participants have a wide range of abilities, from extreme mobility restrictions to moderate learning disabilities. Our four participants are given pseudonyms, referred to in this paper as Will, Jen, Mark, and Rick.

The four participants are made up of three boys, and one girl, aged 16 to 19 years. Information leaflets were sent to the special educational needs school from which they were recruited to inform staff and parents about the project.

Students were selected based on their performance in scales, which represent a set of descriptions used to record and assess the progress of children who have special educational needs (P-scales) [77], [78] (see TABLE 1). Permission for the study was given by Nottingham Trent

University's ethics committee. The user characteristics of each participant are now described in detail.

Will is 18 years old, has a diagnosis of global development delay (GDD) and learning disability. These impact on his speech, language, and social interaction with others. This means his ability to concentrate on a single activity for an extended period is limited, which in turn, limits his sustained attention. His body mobility is not restricted, however slightly imprecise. His speech sounds imprecise and is limited in the selection of words. His capability in conducting particular tasks in quick succession is good; however, he struggles to maintain sustained attention.

Jen is 19 and has a rare form of epilepsy. She is one of the more capable students at the school; she is very cooperative and shows an interest in being involved in the study. She also talks about music and theater and has interests in fashion and celebrities.

Rick is 19 and has a global delay, a rare form of epilepsy and a severe learning difficulty. Rick has problems processing information and communication. His attention is usually committed to a single concept (an activity, a memory, a sound). He is incredibly reliant on routine, and he will try to avoid any disruptions to it. He enjoys loud motor sounds, power tools, and garden work. He often reflects on activities he has done in the past or will do in the future with single words or short phrases. His mobility is not constrained but is delayed and processing time needs to be allowed for any responses. Physical objects and sounds help him associate with new concepts.

Mark is 16 years old and has myotonic dystrophy; this makes his muscles very weak. Myotonic dystrophy is a progressive and life-limiting condition. Mark uses a wheelchair and is at risk of chest infections and sudden heart failure. He uses a specialized CP wheelchair for body support and transportation. The wheelchair supports his body frame and keeps him upright and secure with a safety belt. His head is rested against his right ear on a padded headrest. His mobility disability is extreme; however, he has some imprecise movement in his neck and arms. At the school, he uses both eye gaze technology and switches to interact with computer interfaces. Mark uses his voice to communicate; he likes sharing his sense of humor, he laughs when things go wrong, and makes the sound 'uh-oh' to signal mistakes. He enjoys making choices and can become frustrated when he is not offered choices. Mark likes interacting with computers, however, shows sensitivity to anything resting on his forehead like the EEG headband. Because of his CP, he required a member of staff to be present during the study. Mark shows a definite progression with communication and is now very accepting of and participating in a wider variety of activities, events, and opportunities in school.

D. Feature extraction

Brain-Computer Interfaces (BCIs) represent a novel mode of communication that has been used in emotional classification [79], and cognitive aware applications [80]. BCIs are also considered unique in augmentative and alternative communication (AAC) as they do not require physical movement from a user. This makes BCIs a suitable AAC method for people with Severe Speech and Physical Impairments (SSPI) [81], or CP [82]–[85] who do not have access to conventional means of communication including speech and typing [84].

The quality of a BCI — to offer a direct mode of information from the brain — makes it especially ideal as an element in potential real-time affective user state detection [86], computer interaction for rehabilitation [87] and in brain multimedia interaction [88]. A BCI can also be a complementary source of information towards multimodal interaction systems as well, used in conjunction with other modalities such as gesture, facial expressions, gaze and body posture [89]–[91].

EEG frequency has been used as a feature to determine the active brain state [92]–[94]. In this study, five channels of EEG data are recorded, TP9, AF7, FPz, AF8 and TP10 [95] at a frequency of 500 Hz. EEG Kalman filtering has been shown to be useful in removing EMG induced artifacts [96]–[102]. A robust Adaptive Autoregressive (AAR) model with an order of six detailed in [102] was used. The AAR model estimate of the EEG Kalman filter was utilized to reduce the impact of Electromyography (EMG) spikes from body movement, eye blinks and other facial muscle movements. These EMG spikes are isolated in a few samples, which makes the data ideal for AAR Kalman filtering. In Fig. 4, we see that it has removed the EMG artifact that can be seen between samples points A and B, enhanced the EEG spikes, and revealed an EEG peak between C and D.

By using an AAR Kalman filter on the data, we estimate the EEG wave during the EMG incident artifacts using surrounding neighboring EEG samples and correct those affected samples. This is done by evaluating a moving set of samples and checking for EMG contamination. The contamination is then removed by estimating a normal rate of progression for the signal to reach from point A to point B using a sliding window for the length of the recording.

Studies show [103]–[106] that the EEG beta rhythm (14–30 Hz) is activated when the brain is in a state of arousal. In other EEG studies, mental fatigue related features are associated with decreased alpha band (8-13 Hz) power at one or more parietal locations (e.g., P7 and P8). Ning-Han Liu et al. [107] connected these two factors in their study and showed that alertness can be measured by the signal power of α divided by the signal power of β. Timothy McMahan et al. [108] also demonstrated that the ratio is related to arousal.

Using the signal power of α divided by the signal power of β as the EEG feature, the EEG recordings are labeled with the CPT outcomes. A Butterworth bandpass filter was employed to extract the frequency response of the α and β

bands from the EEG signal as demonstrated in [109]. Discrete Fourier Transform (DFT) was used to calculate the Power Spectral Density (PSD) of the α and β time series.

DFT periodogram methods for estimating the spectrum power density are prone to variation [110]. Periodogram estimate variation is correlated to the square of the value of the spectrum itself. Welch's method reduces this variance by averaging independent periodogram estimates. Each Welch window covers 50% of the next, which results in the smoothed-out average of independent periodogram spectrum estimations. We use a Hamming window as it produces the least amount of overshoot $\delta_{\text{Hamming}} < \delta_{\text{Hann}} < \delta_{\text{Bartlett}}$ [110] with the most accurate results for EEG data [109], [111].

A Hamming window of M = 100 samples was chosen with a 50% overlap, and since the EEG frequency is 500 Hz, this Hamming window is equivalent to 200 ms of data. To help illustrate, an average data interval length is 2.3 seconds long and would have $2300 \div 200 \times 2 = 23$ overlapping Hamming windows. Let $\{xd(n)\}$ be the sequence, $d = 1, 2, 3 \cdots L$ signal intervals and M the interval length. Welch's method to estimate the power spectrum discrete time sequence is shown in (3). Where U is the normalization factor (4) and the Hamming window calculation is shown in (5). Using the Welch method, the ratio of the alpha band power f_α to the beta band power f_β can is simplified as (6).

Welch Method:

$$\hat{p}d(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} xd(n)w(n)e^{-j2\pi f}\right|^2 \tag{3}$$

U is the normalization factor for Welch Method:

$$U = \frac{1}{M}\sum_{n=0}^{M-1}|w(n)| \tag{4}$$

Hamming window:

$$w[n] = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{M}\right), & 0 \le n \le M, \\ 0 & otherwise. \end{cases} \tag{5}$$

The EEG Alertness feature:

$$\text{Alertness} = \frac{\hat{p}d(f_\alpha)}{\hat{p}d(f_\beta)} \tag{6}$$

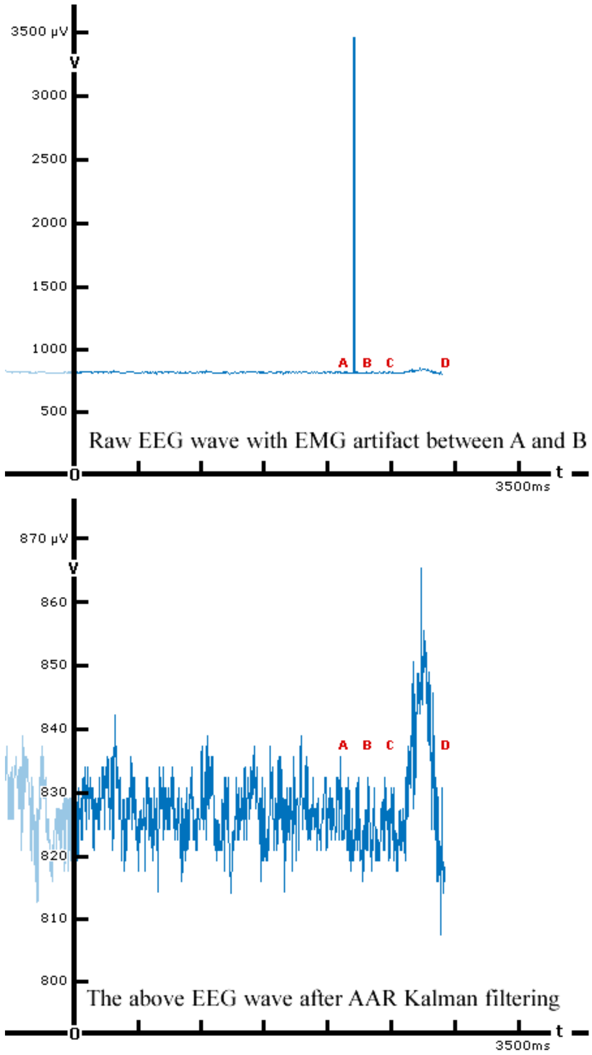


Fig. 4. AAR Kalman filtering reduces EMG noise and enhances EEG spikes

Body pose can be one of the strongest communication channels [112]. Body pose is acquired through the Kinect v2.0 SDK [70], which will provide joint tracking data at 30 Hz. Tracking of the head, neck, mid-spine, right and left shoulders and left and right hands are recorded. Lower joints are not included as occlusion from the table as part of the platform prevents such recordings. Studies have shown that body posture and gesture can communicate affective modalities and also specific emotional categories [27]. They have also been indicators of a firm or weak correlation of engagement during Human-Computer Interaction in gameplay [31]. In this study, the student is positioned in front of a computer system and is challenged to press a button when they identify the target. This type of interaction setup restricts the range of body movements and gestures a student can engage in. Numerous studies [113]–[118] have investigated the importance of body fidgeting in detecting attention for students with PMLD. Fidgeting is an indicator of the onset of attention loss, boredom and engagement deterioration [116], [122]-[125]. We calculate rapid body movement from body pose to assess fidgeting levels. The equation to extract this feature is seen in (7). Where Δd_j is the displacement vector of joint j out of N joints and Δt is the time passing between the displacement samples.

$$\text{Body fidgeting} = \frac{1}{N}\sum_{j=1}^{N}\frac{|\Delta d_j|}{\Delta t} \tag{7}$$

Eye gaze data is recorded at 70 Hz. This data includes Cartesian information regarding the eye gaze location relative to the bottom left corner of the screen. We track gaze, which is both on and off-screen. The combination of off-screen gaze tracking and eye detection provides information on when the user turns their head away from the screen. Three features were extracted from the eye gaze data: 'eye scanning', 'eye dwelling' and 'eyes off-screen'. These features are commonly used in eye gaze technologies to understand attention, interest and engagement [123], [124].

Scanning represents the eye gaze behavior of when the gaze tracks across more than one image element. The scanning feature is calculated in (8) and represents the sum of the inverse distance from the center of each element. Where r_{in} is that distance; from the eye gaze location to the center of image i out of $I = 16$ total image elements, for sample n, out of N total discrete sensor samples. This is demonstrated in Fig. 5.

$$\text{Scanning} = \sum_{n=1}^{N}\sum_{i=1}^{I}\frac{1}{r_{in}} \tag{8}$$

Dwelling represents the eye gaze behavior of when the gaze stays relatively in the same position for a duration of time. This behavior is independently calculated from the location of image elements on the screen. The dwelling feature is calculated in (9), which is the sum of the inverse distance from each eye gaze position to the next. Where n is the sample number out of N total discrete sensor samples, and Δd is the distance the eyes have moved since the previous sample, as demonstrated in Fig. 6.

$$\text{Dwelling} = \sum_{n=1}^{N}\frac{1}{\Delta d_n} \tag{9}$$

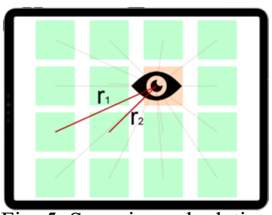


Fig. 5. Scanning calculation with respect to the active elements on the screen

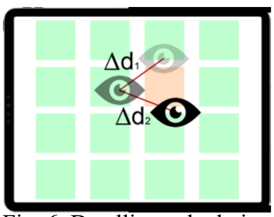


Fig. 6. Dwelling calculation independent of active elements on the screen

The third feature extracted from the eye gaze data is 'eyes off screen'. This continuous but binary feature determines if the participant is looking within the screen area, regardless of whether there was a slide or blank slide on the display. This feature is calculated as in (10).

$$\text{Eyes off screen} = \begin{cases} 1 & eyes\ off\ screen \\ 0 & eyes\ on\ screen \end{cases} \tag{10}$$

Interaction data features were extracted from the participants' behavior activating a button press. The type of pressing, including quick presses or repetitive presses, was recorded as were other sensor data with a view to *behavior*, not just input, but as an independent sensor mode. This makes our approach unique as the input device is considered not only as an objective indicator of attention but also as a separate mode of interaction. We remain impartial to which slide is displayed and only consider the interaction behavior. How the button is pressed, specifically how fast the button is pressed, and how many times it is pressed is of interest. From button presses, we extract two features: single fast button presses and repetitive button presses. Single fast button presses are calculated using the formula described in (11), with the caveat that they are only calculated if the participant presses the button once and only once during the response time duration. In other instances, the value for this feature is zero. Maximum press count is the second feature extracted from the button press data shown in (12). This value is calculated for only the allowed response time interval and is zero when the button is not pressed.

$$\text{Single fast press} = \frac{1}{response\ time} \tag{11}$$

$$\text{Max press count} = \text{total press attempts} \tag{12}$$

High-Level Compound Features (HLCF) were created to create a higher dimensionality in the feature space as described in the Mudra multimodal framework [125]. The first feature is a compound feature, which is simply a normalized mean of the features that traditionally serve indicators of attention. The High-level Attention feature is calculated as the mean of the normalized features of single fast presses, eye dwelling, eye scanning and EEG alertness, which are seen in (13).

$$\text{HLA} = \frac{1}{4}(norm.Single\ fast\ press + norm.Dwelling + norm.Scanning + norm.Alertness) \tag{13}$$

High-level Distraction feature is calculated as the mean of normalized features of body fidgeting, eyes off-screen and press count as seen in (14).

$$\text{HLD} = \frac{1}{3}(norm.Body\ fidgeting + norm.Eyes\ off\ screen\ +$$

$norm. Max\ press\ count)$ (14)

V. Experimental Results

A. Labelling and data fusion

The CPT provides an objective means of labeling the multimodal sensor data. The CPT outcome measures (correct commissions/Hits, False Alarms (FA), correct omissions and misses) are objective outcomes of the participant's attention and engagement with the game. Without these labels, there would be no objective measure or automated way of performing a supervised learning method on the data. An overview of how the data streams are collected and labeled against CPT outcome measures is shown in Fig. 7. Each slide from the moment it is displayed until the moment of the first button press, or until the moment of a new slide being shown (in case of no press), represents a sample of data. Overall, there were 2615 samples collected from the 59 sessions of data collection. The data from all four participants was collated together.

B. Machine learning results

A robust cross-validation method ensures that the results are not subject to overfitting. Leave-one-out [126] classification is a state of the art cross-validation methodology and is widely accepted not to be susceptible to overfitting. We show (regardless of the classification method), that there is a relationship between affective state and the multimodal multisensor data features. In this study, 2615 frames, over the length of 59 sessions, were collected and classified into two categories (engaged and disengaged) using nine features (7 low-level and 2 high-level compound features). The aim of classification is to determine the affective state by predicting the CPT outcome. With two classes, the random classifier classification accuracy to beat is 50%. The overall approach used to evaluate the fit of the different architectures was leave-one-out cross-validation. Impartial scoring metrics were used to competitively compare the performance of the machine learning architectures as these methods normalize across categories (and are suitable for imbalanced datasets). The evaluation parameters used for determining the comparative performance of the machine learning architectures were Area Under the ROC Curve (AUC), Negative Log-likelihood and Kappa. The software used to create this architecture is Python 3.7 and two high-performance computers, which ran in parallel over several weeks. The two PCs were both equipped with Intel i7-7700HQ 2.80 GHz CPUs, and 16 GB of DDR4 RAM. The CPU was benchmarked at 82 Gigaflops, with 15 GB/s memory transfer rate and 1 GB/s SSD disk transfer rate.

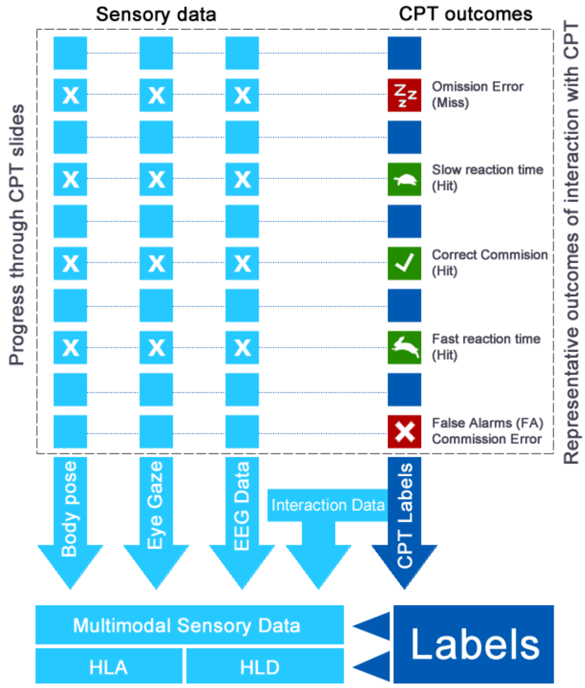


Fig. 7. Multimodal fusion diagram shows the temporal connectivity between the samples and multi-level feature fusion

The summary of results is shown in Table 3. Overall, the random forest classification approach achieved the best classification results in all modes of data. This was both when including high-level features, or when only using a sub-set of the data modes. This finding is supported by other studies [127], which suggests that random forest provides consistent pairwise similarity, crucial for multimodal data. Pairwise-similarity facilitates the combination of features, adding higher dimensionality to the feature space whilst being less sensitive to data sample size [128]. The best method, random forest used both high- and low-level features and achieved 93.3% classification for flow and a 42.9% accuracy for non-flow. The random forest method incorporated 100 trees and all nine features were included at each of the 255 nodes, with 128 leaves in total. AdaBoost, (another ensemble method), outperformed random forest for the single modality feature classification. However, in every example, using any machine learning method, multimodal data features delivered significantly better classification results than any single modality.

When compared to the second-best classification method, random forest outperforms neural network on the classification of non-flow classes with a margin of 16.5% and has an 11.7% better coverage in AUC (see Table 3). Besides neural networks, other machine learning methods were also assessed; AdaBoost, decision tree, k-Nearest Neighbor, naïve Bayes, and support vector machine, however, all had inferior performance when compared to random forest.

Including the two high-level [125] handpicked features (HLA and HLD) in the feature space, improved the classification in every sensor combination, and every machine learning methodology. In the random forest model including HLCF increased the AUC by 1.5% more coverage, and the classification of True Positives (TP) by 0.9%, and True Negatives by 2.8%. On average, if only two modes of sensor input were available, including interaction data improves the outcome of AUC coverage by 16.8%, compared to any other two modes of data, making interaction data the single most important secondary feature. The single most important mode of data on its own however is eye gaze, with 3.2% better AUC coverage compared to interaction data.

The system developed using these machine learning models would not be affected by both sensor fallout and occlusions. At best (all high- and low-level features using random forest) 80.3% AUC coverage is achieved. Using a sub-set of three sensor modes 78.1%-73% AUC coverage is achieved, whilst with a subset of two sensor modes (including interaction) 76.5%-69.2% AUC coverage is achieved. Using a subset of two sensor modes (not including interaction) 70.8%-61.0% AUC coverage is achieved, and with only a single mode of sensor data between 63.7%-48.8% AUC coverage is achieved.

VI. CONCLUSIONS

An approach to labeling multimodal sensor data to train machine-learning algorithms to infer the engagement and flow of students with profound and multiple disabilities has been presented. We posit that this approach can overcome the variation in observer inter-rater reliability when using standardized scales in tracking the emotional expression of students with such profound disabilities. The accuracy of our approach increases with multiple modes of sensor input, and our method is robust to sensor occlusion and fall-out. Multiple sources of sensor input are provided, to accommodate a wide variety of users and their needs. Our model can reliably track the flow of students with profound disabilities, regardless of the sensors available. A system incorporating this model can help teachers design personalized interventions for a very heterogeneous group of students, where teachers cannot possibly attend to each of their individual needs. This approach could be used to identify those with the greatest learning challenges, to guarantee that all students are supported to reach their full potential.

This research was conducted as part of a Ph.D. program

TABLE 3
BEST CLASSIFICATION RESULTS ACHIEVED WITH RANDOM FOREST USING MULTI-LEVEL FEATURE FUSION

Features	Best Classification Method Found	Negative log likelihood for Flow (Less is better)	Negative log likelihood for Non-Flow (Less is better)	Kappa	AUC	TP	TN	F1	Precision	Recall
All features	Random Forest	0.1377	1.0149	0.418	0.803	93.8%	42.9%	0.819	0.817	0.833
Low-Level		0.1440	0.9753	0.374	0.788	92.9%	40.1%	0.806	0.802	0.820
High-Level		0.1250	0.9547	0.237	0.686	93.1%	26.4%	0.768	0.762	0.794
All features	Neural Network	0.1237	0.8191	0.300	0.773	93.6%	31.5%	0.786	0.783	0.808
Low-Level		0.1203	0.7910	0.273	0.767	95.3%	26.4%	0.781	0.783	0.811
All features	AdaBoost	0.2775	2.1803	0.388	0.794	93.3%	40.7%	0.810	0.807	0.824
Low-Level		0.2756	2.0422	0.335	0.765	90.7%	39.7%	0.791	0.785	0.802
All features	Naïve Bays	0.1341	0.7574	0.233	0.712	91.3%	28.7%	0.764	0.755	0.784
Low-Level		0.1097	0.7463	0.095	0.728	98.0%	8.50%	0.732	0.748	0.796
All features	k-NN	0.1105	1.4139	0.207	0.746	96.4%	19.2%	0.763	0.771	0.804
Low-Level		0.1115	1.5077	0.169	0.730	96.5%	15.9%	0.752	0.760	0.799
All features	Tree	0.2545	1.6061	0.309	0.706	89.9%	38.4%	0.782	0.776	0.793
Low-Level		0.1157	1.6414	0.258	0.686	89.4%	34.0%	0.767	0.760	0.780
All features	SVM	0.1107	0.6620	0.086	0.454	76.2%	33.2%	0.686	0.701	0.673
Low-Level		0.1026	0.6750	0.059	0.467	72.7%	34.0%	0.667	0.693	0.647
Eye + EEG + Inter.	Random Forest	0.1429	1.0202	0.349	0.765	92.4%	38.4%	0.793	0.793	0.812
Eye + Body + Inter.	Random Forest	0.1433	1.0784	0.371	0.781	91.6%	41.9%	0.803	0.798	0.814
EEG + Body + Inter.	Random Forest	0.1619	1.5544	0.318	0.730	91.9%	36.0%	0.788	0.783	0.804
Eye + EEG	Random Forest	0.1335	0.7164	0.277	0.679	95.0%	27.1%	0.781	0.783	0.810
Eye + Body	Random Forest	0.1619	1.5544	0.318	0.708	93.7%	27.9%	0.776	0.772	0.801
EEG + Body	AdaBoost	0.4902	2.2224	0.122	0.610	84.2%	27.4%	0.719	0.713	0.725
Eye + Inter.	Random Forest	0.1380	1.1820	0.308	0.765	93.6%	32.2%	0.788	0.785	0.810
Body + Inter.	AdaBoost	0.2579	2.0817	0.327	0.692	83.5%	52.1%	0.776	0.783	0.770
EEG + Inter.	AdaBoost	0.3002	2.0323	0.246	0.708	85.7%	38.3%	0.756	0.753	0.759
EEG	AdaBoost	0.2821	0.8682	0.100	0.559	84.6%	24.8%	0.714	0.706	0.723
Eye gaze	AdaBoost	0.2646	1.3051	0.255	0.637	89.2%	34.0 %	0.766	0.758	0.778
Body	AdaBoost	0.3091	1.0059	0.003	0.488	93.2%	7.10%	0.702	0.674	0.754
Interaction	AdaBoost	0.2491	0.6092	0.035	0.605	95.8%	6.70%	0.713	0.694	0.774
All Features Low-Level	Constant Classifier	0.1004	0.6854	0.000	0.000	100%	0.00%	0.702	0.630	0.794

of research at Nottingham Trent University (NTU) and has been adopted as part of the Erasmus+ KA201 Pathway+ project to determine the affective state of students with mild and moderate learning disabilities (2017-1-UK01-KA201-036761) [129].

References

[1] M. J. Eliason and L. C. Richman, "The Continuous Performance Test in learning disabled and nondisabled children.," *J. Learn. Disabil.*, vol. 20, no. 10, pp. 614–9, Dec. 1987.

[2] E. Bloom and N. Heath, "Recognition, expression, and understanding facial expressions of emotion in adolescents with nonverbal and general learning disabilities," *J. Learn. Disabil.*, vol. 43, no. 2, pp. 180–192, Mar. 2010.

[3] V. L. Petti, S. L. Voelker, D. L. Shore, and S. E. Hayman-Abello, "Perception of Nonverbal Emotion Cues by Children with Nonverbal Learning Disabilities," *J. Dev. Phys. Disabil.*, vol. 15, no. 1, pp. 23–36, 2003.

[4] H. B. Holder and S. W. Kirkpatrick, "Interpretation of emotion from facial expressions in children with and without learning disabilities," *J. Learn. Disabil.*, vol. 24, no. 3, pp. 170–177, Mar. 1991.

[5] L. Weigel, P. E. Langdon, S. Collins, and Y. O'Brien, "Challenging behaviour and learning disabilities: The relationship between expressed emotion and staff attributions," *Br. J. Clin. Psychol.*, vol. 45, no. 2, pp. 205–216, 2006.

[6] F. Heider, *The psychology of interpersonal relations.* Hoboken: John Wiley & Sons Inc, 1958.

[7] D. Dagnan, P. Trower, and R. Smith, "Care staff responses to people with learning disabilities and challenging behaviour: A cognitive-emotional analysis," *Br. J. Clin. Psychol.*, vol. 37, no. 1, pp. 59–68, Feb. 1998.

[8] R. Sharrock, A. Day, F. Qazi, and C. R. Brewin, "Explanations by professional care staff, optimism and helping behaviour: an application of attribution theory.," *Psychol. Med.*, vol. 20, no. 4, pp. 849–55, Nov. 1990.

[9] "SSATrust - Engagement profile & scale."

[10] B. Carpenter, "Engaging young people with complex learning difficulties and disabilities in learning," no. Cldd.

[11] J. P. Roscoe, "'Can the use of a humanoid robot enhance the learning of children with learning disabilities?,'" University of Nottingham, 2014.

[12] P. Standen *et al.*, "Engaging Students with Profound and Multiple Disabilities Using Humanoid Robots," Springer, Cham, 2014, pp. 419–430.

[13] T. Hughes-Roberts *et al.*, "Examining engagement and achievement in learners with individual needs through robotic-based teaching sessions," *Br. J. Educ. Technol.*, vol. 50, no. 5, pp. 2736–2750, Sep. 2019.

[14] J. I. Marcum, "A statistical theory of target detection by pulsed radar," Dec. 1947.

[15] H. L. Swanson, "Vigilance Deficit in Learning Disabled Children: A Signal Detection Analysis," *J. Child Psychol. Psychiatry.*, vol. 22, no. 4, pp. 393–9, Oct. 1981.

[16] A. Lanatà, G. Valenza, and E. P. Scilingo, "Eye gaze patterns in emotional pictures," *J. Ambient Intell. Humaniz. Comput.*, vol. 4, no. 6, pp. 705–715, Dec. 2013.

[17] R. N. Khushaba, C. Wise, S. Kodagoda, J. Louviere, B. E. Kahn, and C. Townsend, "Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3803–3812, Jul. 2013.

[18] R. W. Picard, "Affective Computing," *MIT Press*, no. 321, pp. 1–16, 1995.

[19] M. Pantic, A. Nijholt, A. Pentland, and T. S. Huang, "Human-Centred Intelligent Human-Computer Interaction (HCI^2): how far are we from attaining it?," *Int. J. Auton. Adapt. Commun. Syst.*, vol. 1, no. 2, pp. 168–187, 2008.

[20] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[21] M. Pantic, A. Pentland, A. Nijholt, and T. Huang, "Human Computing and Machine Understanding of Human Behavior: A Survey," 2006.

[22] Beckman Institute, "Multimodal Human Computer Interaction Towards a Proactive Computer," 2001.

[23] R. Ishii, Y. Shinohara, I. Nakano, and T. Nishida, "Combining Multiple Types of Eye-gaze Information to Predict User's Conversational Engagement," *Hum. Factors*, pp. 1–8, 2011.

[24] A. Jaimes and N. Sebe, "Multimodal human–computer interaction: A survey," *Comput. Vis. Image Underst.*, vol. 108, no. 1–2, pp. 116–134, Oct. 2007.

[25] M. Pantic, N. Sebe, J. F. Cohn, and T. Huang, "Affective Multimodal Human-Computer Interaction," *Proc. 13th Annu. ACM Int. Conf. Multimed.*, pp. 669–676, 2005.

[26] R. Sharma, V. I. Pavlovic, and T. S. Huang, "Toward multimodal human-computer interface," *Proc. IEEE*, vol. 86, no. 5, pp. 853–869, May 1998.

[27] A. Kleinsmith and N. Bianchi-Berthouze, "Recognizing Affective Dimensions from Body Posture," in *Affective Computing and Intelligent Interaction*, 2007, pp. 48–58.

[28] N. Savva, A. Scarinzi, and N. Bianchi-Berthouze, "Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience," *IEEE Trans. Comput. Intell. AI Games*, vol. 4, no. 3, pp. 199–212, 2012.

[29] S. Asteriadis, K. Karpouzis, and S. Kollias, "The Importance of Eye Gaze and Head Pose to Estimating Levels of Attention," *2011 Third Int. Conf. Games Virtual Worlds Serious Appl.*, no. September 2015, pp. 186–191, 2011.

[30] Z. Zeng, J. Tu, B. M. Pianfetti, and T. S. Huang, "Audio-visual affective expression recognition through multistream fused HMM," *IEEE Trans. Multimed.*, vol. 10, no. 4, pp. 570–577, 2008.

[31] N. Bianchi-Berthouze, W. W. Kim, and D. Patel, "Does Body Movement Engage You More in Digital Game Play? and Why?," in *Affective Computing and Intelligent Interaction*, 2007, pp. 102–113.

[32] S. D'Mello and A. Graesser, "Dynamics of affective states during complex learning," *Learn. Instr.*, vol. 22, no. 2, pp. 145–157, 2012.

[33] M. Csikszentmihalyi and M. Csikzentmihaly, "Flow: The psychology of optimal experience," vol. 41, no. May, 1991.

[34] M. Csikszentmihalyi, "The Flow of Creativity," *Creat. Flow Psychol. Discov. Invent.*, pp. 107–126, 1996.

[35] R. C. Wilson, A. Shenhav, M. Straccia, and J. D. Cohen, "The Eighty Five Percent Rule for optimal learning," *Nat. Commun.*, vol. 10, no. 1, pp. 1–9, 2019.

[36] M. Taheri *et al.*, "State Diagram for Affective Learning in an Educational Platform," pp. 4–6, 2018.

[37] M. Csikszentmihalyi, "Creativity: Invention, Flow and the psychology of discovery and Invention," in *Personnel Psychology*, vol. 51, no. 3, 1997, p. 16.

[38] P. H. Mirvis and M. Csikszentmihalyi, "Flow: The Psychology of Optimal Experience," *Acad. Manag. Rev.*, vol. 16, no. 3, pp. 636–640, 1991.

[39] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," *MindTrek - 12th Int. MindTrek Conf. Entertain. Media Ubiquitous Era*, no. January 2008, p. 13, 2008.

[40] A. R. Basawapatna, A. Repenning, K. H. Koh, and H. Nickerson, "The zones of proximal flow," *Proc. ninth Annu. Int. ACM Conf. Int. Comput. Educ. Res. - ICER '13*, p. 67, 2013.

[41] B. Carpenter, S. Rose, H. Rawson, and J. Egerton, "The rules of engagement," *SEN Mag.*, no. 54, pp. 34–54, 2011.

[42] B. Carpenter *et al.*, *Engaging Learners with Complex Learning Difficulties and Disabilities: A resource book for teachers and teaching assistants*. Taylor & Francis, 2015.

[43] R. Iovannone, G. Dunlap, H. Huber, and D. Kincaid, "Effective Educational Practices for Students with Autism Spectrum Disorders.," *Focus Autism Other Dev. Disabil.*, vol. 18, pp. 150–165, 2003.

[44] M. J. Lee and B. Ferwerda, "Personalizing online educational tools," *Humaniz. 2017 - Proc. 2017 ACM Work. Theory-Informed User Model. Tailoring Pers. Interfaces, co-located with IUI 2017*, pp. 27–30, 2017.

[45] C. R. Henrie, L. R. Halverson, and C. R. Graham, "Measuring student engagement in technology-mediated learning: A review," *Comput. Educ.*, vol. 90, pp. 36–53, 2015.

[46] A. K. |Wiggins. J. B. |Grafsgaard. J. F. |Boyer. K. E. E. N. |Lester. J. C. Vail, "The Affective Impact of Tutor Questions: Predicting Frustration and Engagement," *Int. Educ. Data Min. Soc.*, 2016.

[47] S. Aslan *et al.*, "Investigating the Impact of a Real-time, Multimodal Student Engagement Analytics Technology in Authentic Classrooms," in *Conference on Human Factors in Computing Systems - Proceedings*, 2019.

[48] J. Hamari, D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke, and T. Edwards, "Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning," *Comput. Human Behav.*, vol. 54, no. JANUARY, pp. 170–179, 2016.

[49] "Engaging children with complex learning difficulties and disabilities in the Primary Classroom. Barry Carpenter," no. 2002, 2011.

[50] J. C. Abrams, "On learning disabilities: Affective considerations," *J. Reading, Writing, Learn. Disabil. Int.*, vol. 2, no. 3, p. 190, Jan. 1986.

[51] B. Hiebert, B. Y. Wong, and M. Hunter, "Affective influences on learning disabled adolescents.," *Learn. Disabil. Q.*, vol. 5, no. 4, pp. 334–343, 1982.

[52] R. R. McCrae, "Consensual validation of personality traits: Evidence from self-reports and ratings," *J. Pers. Soc. Psychol.*, vol. 43, no. 2, pp. 293–303, 1982.

[53] D. Watson and L. A. Clark, "Self- Versus Peer Ratings of Specific Emotional Traits: Evidence of Convergent and Discriminant Validity," *J. Pers. Soc. Psychol.*, vol. 60, no. 6, pp. 927–940, Jun. 1991.

[54] J. S. Warm, R. Parasuraman, and G. Matthews, "Vigilance requires hard mental work and is stressful.," *Hum. Factors*, vol. 50, no. 3, pp. 433–41, Jun. 2008.

[55] G. J. DuPaul, A. D. Anastopoulos, T. L. Shelton, D. C. Guevremont, and L. Metevia, "Multimethod Assessment of Attention-Deficit Hyperactivity Disorder: The Diagnostic Utility of Clinic-Based Tests," *J. Clin. Child Psychol.*, vol. 21, no. 4, pp. 394–402, Dec. 1992.

[56] H. L. Swanson, "A Developmental Study of Vigilance in Learning-disabled and Nondisabled Children," *J. Abnorm. Child Psychol.*, vol. 11, no. 3, pp. 415–29, Sep. 1983.

[57] W. Peterson, T. Birdsall, and W. Fox, "The theory of signal detectability," *Trans. IRE Prof. Gr. Inf. Theory*, vol. 4, no. 4, pp. 171–212, Sep. 1954.

[58] N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide*. Psychology Press, 2005.

[59] J. A. Swets and R. M. Pickett, *Evaluation of diagnostic systems : methods from signal detection theory*. Academic Press, 1982.

[60] W. P. Tanner and J. A. Swets, "A decision-making theory of visual detection," *Psychol. Rev.*, vol. 61, no. 6, pp. 401–409, 1954.

[61] H. E. Rosvold, A. F. Mirsky, I. Sarason, E. D. Bransome, and L. H. Beck, "A continuous performance test of brain damage.," *J. Consult. Psychol.*, vol. 20, no. 5, pp. 343–350, 1956.

[62] D. H. Sykes, V. I. Douglas, G. Weiss, and K. K. Minde, "Attention in hyperactive children and the effects methylphenidaten (Ritalin)," *J. child Psychol. psychiatry*, vol. 12, no. 2, pp. 129–139, 1971.

[63] "Be first to get a Tobii EyeX Dev Kit Want to develop for Android or Linux ? Get a," vol. 595, p. 795.

[64] I. Software, "Muse : The Brain Sensing Headband Tech Spec

Sheet."

[65] "MATLAB - The Language of Technical Computing - MathWorks United Kingdom." [Online]. Available: http://uk.mathworks.com/products/matlab/. [Accessed: 07-Mar-2016].

[66] M. Teplan, "Fundamentals of EEG measurement," *Meas. Sci. Rev.*, vol. 2, pp. 1–11, 2002.

[67] "Tobii EyeX Controller – get your own eye tracker." 30-Jul-2015.

[68] "An introduction to Tobii EyeX | Tobii Developer Zone." [Online]. Available: http://developer.tobii.com/an-introduction-to-the-tobii-eyex-sdk/. [Accessed: 03-Mar-2016].

[69] "Specifications for EyeX – Tobii Eye Tracking Support." [Online]. Available: https://help.tobii.com/hc/en-us/articles/212818309-Specifications-for-EyeX. [Accessed: 10-Apr-2019].

[70] "Kinect hardware." [Online]. Available: https://dev.windows.com/en-us/kinect/hardware. [Accessed: 03-Mar-2016].

[71] Q. Wang, Q. Wang, and G. Kurillo, "Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect," no. November, 2015.

[72] M. Rosenberg, S. Noonan, J. Degutis, and M. Esterman, "Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task.," *Atten. Percept. Psychophys.*, vol. 75, no. 3, pp. 426–39, 2013.

[73] C. A. Riccio, Cecil R. Reynolds, P. Lowe, and J. J. Moore, "The continuous performance test: a window on the neural substrates for attention?," *Arch. Clin. Neuropsychol.*, vol. 17, no. 3, pp. 235–272, Apr. 2002.

[74] S. H. Klee and B. D. Garfinkel, "The computerized continuous performance task: a new measure of inattention.," *J. Abnorm. Child Psychol.*, vol. 11, no. 4, pp. 487–495, 1983.

[75] C. K. Conners, "The computerized continuous performance test.," *Psychopharmacol. Bull.*, vol. 21, no. 4, pp. 891–2, Jan. 1985.

[76] C. A. Riccio, J. J. M. Waldrop, C. R. Reynolds, and P. Lowe, "Effects of stimulants on the continuous performance test (CPT): implications for CPT use and interpretation.," *J. Neuropsychiatry Clin. Neurosci.*, vol. 13, no. 3, pp. 326–335, Aug. 2001.

[77] "Teacher assessment: using P scales to report on pupils with SEN - Detailed guidance - GOV.UK." [Online]. Available: https://www.gov.uk/guidance/teacher-assessment-using-p-scales. [Accessed: 07-Mar-2016].

[78] "P scales: attainment targets for pupils with SEN - Publications - GOV.UK," *Department for Education*. [Online]. Available: https://www.gov.uk/government/publications/p-scales-attainment-targets-for-pupils-with-sen. [Accessed: 07-Mar-2016].

[79] T. Affective and O. F. E. Emotion, "BRAIN – COMPUTER INTERFACE : FUNDAMENTALS AND ANALYSIS OF EEG-BASED EMOTION CLASSIFICATION."

[80] A. Bulling and T. O. Zander, "Cognition-Aware Computing," *IEEE Pervasive Comput.*, vol. 13, no. 3, pp. 80–83, 2014.

[81] M. Akcakaya *et al.*, "Noninvasive brain-computer interfaces for augmentative and alternative communication," *IEEE Rev. Biomed. Eng.*, vol. 7, no. c, pp. 31–49, 2014.

[82] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control.," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–91, Jun. 2002.

[83] M. Ahsan, M. Ibrahimy, and O. Khalifa, "EMG signal classification for human computer interaction: a review," *Eur. J. Sci. …*, vol. 33, no. 3, pp. 480–501, 2009.

[84] I. Daly *et al.*, "On the control of brain-computer interfaces by users with cerebral palsy," *Clin. Neurophysiol.*, vol. 124, no. 9, pp. 1787–1797, 2013.

[85] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 51, pp. 17849–54, 2004.

[86] T. O. Zander and C. Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general.," *J. Neural Eng.*, vol. 8, no. 2, p. 025005, Apr. 2011.

[87] E. Haselsteiner and G. Pfurtscheller, "Using time-dependent neural networks for EEG classification.," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 457–63, Dec. 2000.

[88] Y. P. Lin *et al.*, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, 2010.

[89] D. P.-O. O. Bos *et al.*, "Human-computer interaction for BCI games: Usability and user experience," *Proc. - 2010 Int. Conf. Cyberworlds, CW 2010*, pp. 277–281, Oct. 2010.

[90] C. Mühl *et al.*, "Bacteria Hunt," *J. Multimodal User Interfaces*, vol. 4, no. 1, pp. 11–25, Aug. 2010.

[91] C. Mühl *et al.*, "Bacteria Hunt: A multimodal, multiparadigm BCI game," *enterface2009*, no. 1, pp. 1–22, 2009.

[92] T. Egner and M. B. Sterman, "Neurofeedback treatment of epilepsy: from basic rationale to practical application.," *Expert Rev. Neurother.*, vol. 6, no. 2, pp. 247–57, Feb. 2006.

[93] O. M. Bazanova and D. Vernon, "Interpreting EEG alpha activity.," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 94–110, 2013.

[94] M. Adjounadi, M. Cabrerizo, I. Yaylali, and P. Jayakar, "Interpreting EEG functional brain activity," *IEEE Potentials*, vol. 23, no. 1, pp. 8–13, 2004.

[95] Trans Cranial Technologies Ltd., "10 / 20 System Positioning Manual," p. 20, 2012.

[96] A. Kubler, E. Holz, T. Kaufmann, and C. Zickler, "A User Centred Approach for Bringing BCI Controlled Applications to End-Users," in *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, 2013.

[97] V. P. Oikonomou, a T. Tzallas, and D. I. Fotiadis, "A Kalman filter based methodology for EEG spike enhancement.," *Comput. Methods Programs Biomed.*, vol. 85, no. 2, pp. 101–8, Feb. 2007.

[98] F. Morbidi, S. Member, A. Garulli, S. Member, C. Rizzo, and S. Rossi, "Application of Kalman filter to remove TMS-induced artifacts from EEG recordings," pp. 1–7.

[99] M. Arnold, W. H. Miltner, H. Witte, R. Bauer, and C. Braun, "Adaptive AR modeling of nonstationary time series by means of Kalman filtering.," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 5, pp. 553–62, May 1998.

[100] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005.," *J. Neural Eng.*, vol. 2, no. 4, pp. L14-22, Dec. 2005.

[101] V. P., A. T., S. Konitsiotis, D. G., and D. I., "The Use of Kalman Filter in Biomedical Signal Processing," in *Kalman Filter Recent Advances and Applications*, no. April, 2009.

[102] G. Pfurtscheller, C. Neuper, a Schlögl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters.," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 316–25, Sep. 1998.

[103] L. Trejo *et al.*, "EEG-based estimation of mental fatigue: Convergent evidence for a three-state model," *Found. Augment. Cogn.*, pp. 201–211, 2007.

[104] A. Gevins and M. E. Smith, "Detecting transient cognitive impairment with EEG pattern recognition methods," *Aviat. Sp. Environ. Med.*, vol. 70, no. 10, pp. 1018–1024, Oct. 1999.

[105] T. C. Hankins and G. F. Wilson, "A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight.," *Aviat. Space. Environ. Med.*, vol. 69, no. 4, pp. 360–7, Apr. 1998.

[106] D. Ming *et al.*, "Electroencephalograph (EEG) signal processing method of motor imaginary potential for attention level classification," *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009*, vol. 2009, pp. 4347–4351, Jan. 2009.

[107] N. H. Liu, C. Y. Chiang, and H. C. Chu, "Recognizing the degree of human attention using EEG signals from mobile sensors.," *Sensors (Basel).*, vol. 13, no. 8, pp. 10273–10286, Jan. 2013.

[108] T. McMahan, I. Parberry, and T. D. Parsons, "Evaluating Player Task Engagement and Arousal Using Electroencephalography," *Procedia Manuf.*, vol. 3, no. Ahfe, pp. 2303–2310, 2015.

[109] A. Ameera, A. Saidatul, and Z. Ibrahim, "Analysis of EEG Spectrum Bands Using Power Spectral Density for Pleasure and Displeasure State," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 557, no. 1, 2019.

[110] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, Second. Pearson, 1999.

[111] J. Musson and J. Li, "A Comparative Survey of PSD Estimation Methods for EEG Signal Analysis," *Student Capstone Conf. Proc.*, pp. 1–6, 2010.

[112] N. Bianchi-Berthouze, "The Affective Body Argument in Technology Design," in *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*, 2016, pp. 3–6.

[113] J. Farley, E. F. Risko, and A. Kingstone, "Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering," *Front. Psychol.*, vol. 4, no. SEP, p. 619, Sep. 2013.

[114] M. D. Smith and M. S. Barrett, "The Effect of Parent Training on Hyperactivity and Inattention in Three School-Aged Girls with Attention Deficit Hyperactivity Disorder," *Child Fam. Behav. Ther.*, vol. 24, no. 3, pp. 21–35, Oct. 2002.

[115] C. K. Conners, L. Eisenberg, and A. Barcai, "Effect of Dextroamphetamine on Children," *Arch. Gen. Psychiatry*, vol. 17, no. 4, p. 478, Oct. 1967.

[116] P. L. Holborow and P. S. Berry, "Hyperactivity and Learning Difficulties," *J. Learn. Disabil.*, vol. 19, no. 7, pp. 426–431, Aug. 1986.

[117] E. Benedetto-Nasho and R. Tannock, "Math computation, error patterns and stimulant effects in children with Attention Deficit Hyperactivity Disorder," *J. Atten. Disord.*, vol. 3, no. 3, pp. 121–134, Oct. 1999.

[118] S. L. Smith, *No easy answer : the learning disabled child at home and at school*. Bantam Books, 2012.

[119] N. Bianchi-Berthouze and Nadia Bianchi-Berthouze, "Understanding the role of body movement in player engagement," *Human-Computer Interact.*, vol. 28, no. 1, pp. 40–75, 2013.

[120] B. Fenesi, K. Lucibello, J. A. Kim, and J. J. Heisz, "Sweat So You Don't Forget: Exercise Breaks During a University Lecture Increase On-Task Attention and Learning," *J. Appl. Res. Mem. Cogn.*, vol. 7, no. 2, pp. 261–269, 2018.

[121] S. Lis, N. Baer, C. Stein-En-Nosse, B. Gallhofer, G. Sammer, and P. Kirsch, "Objective measurement of motor activity during cognitive performance in adults with attention-deficit/hyperactivity disorder," *Acta Psychiatr. Scand.*, vol. 122, no. 4, pp. 285–294, 2010.

[122] A. M. F. Zinno, G. Douglas, S. Houghton, V. Lawrence, J. West, and K. Whiting, "Hyperactivity Disorder (ADHD) during computer video game play," *Br. J. Educ. Technol.*, vol. 32, no. 5, pp. 607–618, 2001.

[123] J. A. Vigo, "Eye Gaze Tracking for Tracking Reading Progress," p. 100, 2013.

[124] P. Fattahi, "Eyetracking Definitions: Basic Definition of Terms," Teltow, Germany, 2011.

[125] L. Hoste, B. Dumas, and B. Signer, "Mudra: A Unified Multimodal Interaction Framework," 2011, pp. 97–104.

[126] H. Zhang, Y. Zhu, J. Maniyeri, and C. Guan, "Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine.," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, pp. 2985–8, Jan. 2014.

[127] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *Neuroimage*, vol. 65, pp. 167–175, 2013.

[128] S. Hor and M. Moradi, “Scandent Tree: A Random Forest Learning Method for Incomplete Multimodal Datasets,” Springer, Cham, 2015, pp. 694–701.

[129] “Mobile Pedagogical Assistant to develop meaningful pathways to personalised learning.” [Online]. Available: https://pedagogics-pathway.eu/. [Accessed: 09-Apr-2020].

Multimodal Multisensor Data Relationship to Learner Potential

Physiological Data Relationships with Academic Achievement

Mohammad H. Taheri
Science and Technology
Nottingham Trent University
Nottingham UK
mohammad.taheri@ntu.ac.uk.com

David J. Brown
Science and Technology Nottingham
Trent University
Nottingham UK
david.brown@ntu.ac.uk

Nasser Sherkat
Science and Technology
Sheffield Hallam University
Sheffield UK
n.sherkat@shu.ac.uk

ABSTRACT

There are different types of the Continuous Performance Test (CPT) that are used to track vigilance and inhibition. In this work, we introduce a new type of CPT, the Seek-X, coupled with a novel Multimodal Multisensor tracking platform to investigate the relationships between the academic progress of students with a range of learning disabilities, and the outcomes of these types of tests (such as the number of errors they make, and reaction times in spotting the 'target' signal). Four students with moderate and severe intellectual disabilities participated in a 'Seek-X' Type CPT in which their eye gaze, EEG, body pose, and interaction data were collected using a range of sensors whilst performing the test. The study took place over 13 weeks, including 59 sessions from which 2615 data samples were collected. Overall classification results using a random forest model, with both high- and low-level features using person-specific data, achieved 84.8% classification for 'attention' and 65.4% classification for 'inattention'. In order to investigate the suitability of the CPT to label this data to infer selective and sustained attention, we investigated the cross-correlation between 55 data points and 2475 separate correlation assessments. Correlations between the Multimodal Multisensor data, outcomes from the CPT, and participant academic performance using 'P scale' were assessed longitudinally, and significant correlations were found. Importantly, a strong positive correlation was found between participant ability to sustain attention (higher d') and their academic performance ($p < .01$). Participants who were less impulsive and displayed greater inhibition had greater academic performance ($p < .01$). Participants that displayed eye gaze scanning or eye dwelling behavior more often progressed further in their 'Reading' and 'Listening' scales ($p <= .05$). Even though previous studies present conflicting evidence regarding the independence of bias and sensitivity, we conclude that they are independent. A high-level compound feature 'Attention' was shown to have a direct correlation with participant attention. Our results show that the 'Seek-X' Type CPT can be utilized to assess and monitor participant performance, and also to help understand the specific challenges that students with severe and complex disabilities face, including physiological and cognitive impairments that may hinder their academic performance. Given the range of the correlations found, we propose that the 'Seek-X' Type CPT is a novel and accurate method of labeling Multimodal Multisensor data to automatically infer the attention state of students with moderate and severe intellectual disabilities that can subsequently be used to personalize their learning experiences.

CCS CONCEPTS

• Human-centered computing → Interaction design; • Applied computing → Education → Interactive learning environments; • Theory of computation → Models of computation → Interactive computation; • Human-centered computing → Interaction design → Interaction design process and methods → User centered design; • Human-centered computing → Interaction design → Empirical studies in interaction design;

KEYWORDS

Academic Achievement, ADHD, Affective Computing in Education, Affect Detection, Continuous Performance Test, Attention, Engagement, Flow, HCI, Interaction, Learning Disabilities, Machine Learning, Multimodal, Multisensor, Physiological Sensors, Signal Detection Theory, Student Engagement.

ICMI'22, October 2020, Utrecht, The Netherlands

978-1-4503-0000-0/18/06...$15.00
https://doi.org/10.1145/

ACM Reference format:

Multimodal Multisensor Data Relationship to Learner Potential: Physiological Data Relationships with Learner Capacity. In *Proceedings of ACM ICMI conference (ICMI'20). ACM, Utrecht, The Netherlands, 12 pages.* https://doi.org/10.1145/1234567890

1 Introduction

Numerous research studies use the objective outcomes of the Continuous Performance Test (CPT) to understand the characteristics of participants with learning disabilities [1], [2], [11], [12], [3]–[10]. Little work has been carried out to establish the relationships between CPT outcomes and participant physiological data or academic performance demonstrated by students in academic subjects.

The CPT was first introduced by Swanson as a standardization of the Signal Detection Theory (SDT) in the '80s [4], [7] to objectively study vigilance and sustained and selective attention in children with learning disabilities. Swanson's CPT [7] is reported to be the most popular measure of sustained attention or vigilance—the ability to sustain attentional focus and remain alert to stimuli over time [13], [14].

CPTs provide objective [15]–[17] outcomes of sustained and selective attention [18]–[23] and have been positively related to academic performance [15], [24]. Others have used the concept of continuous performance monitoring of game outcomes as a method to infer affect state [25]–[28], in combination with the Zone of Proximal Flow (ZPF) theory [27]. The CPT has also been used in combination with body pose and head tracking data to track attention [16], and used to label multimodal sensory data with objective attention labels [17].

Many studies relate flow, a closely related affect state of engagement, to greater learning outcomes [29]. One is in flow when one is engaged [30] and steady performance has been maintained at the comfortable limits of one's skill limitations [28]. Maintaining flow in learning is especially significant because it is the most reliable indicator for determining successful learning [36], [45]-[48]. This results in immersion, concentrated focus, and deep learning [35], [36]. Flow is central to classroom performance and the achievement of learning outcomes [45]-[48], which is closely linked with attention [28], [30], [37], [38]. During flow, attention is completely absorbed in the task at hand, and the person's performance is maximized [29]. The ability to sustain attention often coincides with inhibition [29], [35], [36], [39], [40], which increases performance [29], [40]. Therefore, one can infer flow by tracking peak-performance in a pre-defined activity/task with a known challenge. This has been validated in a number of studies [25], [27], [28], [31], [33], [41]–[44]. For example, flow is positively related to artistic and scientific creativity [45], [46], effective teaching [35], learning [47] and peak performance in sports [48], [49]. This positive relationship between flow and academic performance has been demonstrated in a number of studies in a school setting where students achieved a higher grade point average (GPA) [29], [47], [49], [50].

Flow tends to occur when there is a balance between perceived challenges [29], [51] and perceived skills, otherwise known as the concept of "optimal arousal" [26]–[28], [52], [53]. When perceived challenges and skills are balanced, attention is completely invested [29]. This balance is inherently fragile; if challenge gradually exceeds skill, one typically becomes anxious or frustrated [54]; if skill begins to exceed challenge, one relaxes and then becomes bored [27]–[29]. This represents a push-pull dynamic. The equilibrium of skill and challenge is also represented in the Zone of Proximal Flow (ZPF) theory [27], [28]. We adapt and modify the engagement diagram from Chen *et al.* [55] and Bianchi-Berthouze *et al.* [30] in Figure 1, with a focus on learning challenge and learner skill, using the definitions of performance, attention and flow from the works of Csíkszentmihályi [29], [56], [57] and the ZPF theory [27], [28].

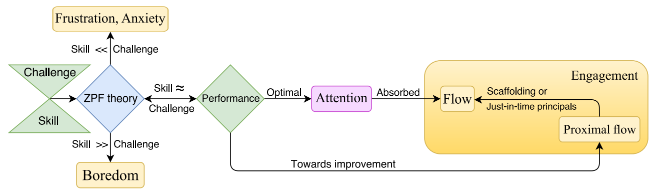


Figure 1. Visualization of the relationship between performance, attention and flow using the Zone of Proximal Flow theory [27], and definitions from Csíkszentmihályi theory of flow [29]

To understand the relationships between Multimodal Multisensor (MM) data (e.g., *eye gaze, body pose, interaction data*, etc.) and SDT/CPT outcomes and participant physiological characteristics (e.g., *fidgeting, eye scanning, eye dwelling,* etc.) we propose a platform that facilitates the tracking of these measures. An MM platform incorporating a new type of CPT the Seek-X type is proposed. MM data is collected and labeled using SDT from a CPT. Features from the data are used in a machine learning model to classify participant moment-to-moment attention state.

In this study, we explore the relationships between participant physiological data and their performance in a range of academic subjects. We also assess prior assumptions regarding the attentional capacity of students with learning disabilities, their inhibition, and bias. We investigate the potential of using a new type of CPT, adapted to suit the needs of students with moderate and severe intellectual disabilities, in conjunction with MM data to help understand the capacity of these students to demonstrate performance against a common range of descriptors in academic subjects (including English, Maths, and Computing). The correlations found in our data help present a case for using our new form of the CPT as an outcome measure in research, and as a valid method for labeling MM data to automatically infer attention, and its relationship to performance and flow.

2 Methodology

Four students were recruited for data collection, *see Participants (section 3).* They took part in an 11-week study with up to four sessions weekly. Each session included 48 challenges. Each test lasted between 6-32 minutes depending on participant readiness or other setting-up challenges. Every session recorded nearly 4 minutes of data. A total of 59 sessions of the CPT were carried out (an average of 15 sessions per participant). A series of 48 slides with pauses in between were displayed for each participant. A total of 2615 SDT signal challenges were presented to the participants. Despite the participant count being low, and recognizing the challenges setting up experiments of this nature (collecting data from multiple data sensors from participants with moderate and severe intellectual disabilities), we still achieved statistical significance using repetition. Swanson's study collected 2240 SDT signal challenges from participants with learning disabilities [7], Goldberg *et al.* collected a total of 240 SDT signal samples [58]. Our study with 2615 SDT samples falls within the norms of these prior studies.

2.1 Experimental platform

This CPT design was based on Rosvald and Mirsky's original research [59]. Time alterations to the experiment length were made to match the shorter length activities that students with moderate and severe intellectual disabilities are accustomed to at school [7]. An image of the platform in its early developmental stages can be seen in Figure 2. A new type of CPT, the 'Seek-X Type' is developed. Data points from the CPT are extracted, such as the qualities of the participant interaction behavior (e.g., *Response Time*, *Press Count, and Single presses*) and include mean, maximum, and minimums of these CPT attributes.

Each participant was asked to press a button, whenever they saw the target image on the screen, and not to press the button when they did not see the target image on the screen (see Figure 2). During this activity, participant eye gaze, body pose, electroencephalogram (EEG) measurements, and button interaction data were continuously recorded in real-time. The MM data is later used to create a machine learning model of flow using a random forest classifier. The experimental platform and the relative student position are visualized in Figure 3.

Real-time eye gaze position using Tobii EyeX [60], body pose data using Kinect v2, EEG data from the Muse headband [61], and interaction data from the USB button is recorded in MATLAB [62]. The Muse EEG headband streams 16-bit voltage data in microvolt (μV) units at 500 Hz, which is equal or comparable to medical-grade EEG specifications [63]. The Tobii EyeX eye gaze tracking controller [64] uses near-infrared light to track the eye movements and gaze point of each participant [65]. It has a frequency of 70 Hz and uses backlight assisted near-infrared (NIR 850 nm + red light (650 nm)) to achieve a 95% tracking population [66]. The Kinect 2 sensor [67] is a motion-sensing peripheral for body tracking. Using structured light and machine learning it can infer body position [67]. Kinect 2 is reported to have an average depth accuracy of under 2 mm in the central viewing angle and increases to 2-4 mm in the range of up to 3.5 mm [68]. The features extracted from these sensor data streams are described under *Features and data points (section 4).*

Figure 2. The platform being tested during the developmental stage

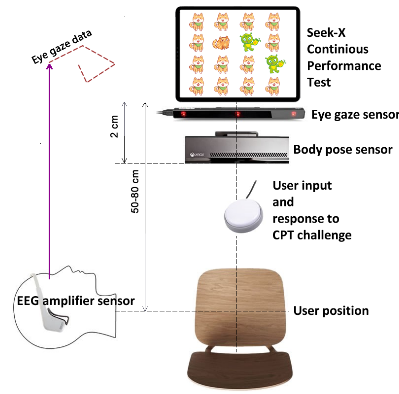


Figure 3. The MM experimental platform with the eye gaze, body pose, EEG sensor, and the CPT

3 Participants

Four participants were recruited from a local school catering for students with physical difficulties, severe learning disabilities and profound multiple learning difficulties aged 3-19 years (Nottingham, UK), to collect labeled sensor data while using the gamified platform. Ethical permission for the study was granted from Nottingham Trent University's non-invasive ethics committee. These four participants have a wide range of skills, from extreme mobility restrictions to moderate and severe intellectual disabilities. The four participants (pseudonyms: Will, Jen, Mark, and Rick) include three boys, and one girl, aged 16 to 19 years. They were selected based on their attainment in performance scales, representing a set of descriptions used to record and assess the progress of children who have special educational needs (P scales) [69], [70] (see Table 1). The characteristics of each participant are now described in detail.

Will is 18 years old, has a diagnosis of global development delay (GDD) and Learning Disability (LD). These impact on his speech, language, and social interaction with others. This means his ability to concentrate on a single activity for an extended period is limited, which in turn limits his sustained attention. His body mobility is not restricted but is slightly imprecise. His speech can be difficult to understand, and he is limited in the selection of words. His capability in conducting particular tasks in quick succession is good; however, he struggles to maintain sustained attention.

Table 1. Participant characteristics and CPT settings

Participant alias	Age	P scales mean	Display time (s)	Interstimulus interval (s)
Will	18	6.93	1.8	1.1 ± 0.1
Jen	19	19.45*	1	1.1 ± 0.1
Mark	16	3.7	8	2.1 ± 0.1
Rick	19	6.76	1.8	1.1 ± 0.1

*Jen is enrolled in the National Curriculum.

Jen is 19 and has a rare form of epilepsy. She is one of the more capable learners at the school. She is very cooperative and shows an interest in being involved in the study. She also talks about music and theater and has interests in fashion and celebrities.

Rick is 19 and has global delay, a rare form of epilepsy, and a severe learning disability. Rick has problems processing information and communicating. His attention is usually limited to a single concept (an activity, a memory, a sound). He is incredibly reliant on routine, and he will try to avoid any disruptions to it. He often reflects on activities he has done in the past or will do in the future with single words or short phrases. His mobility is not constrained but is delayed and processing time needs to be allowed for any responses. Physical objects and sounds help him make associations with new concepts.

Mark is 16 years old and has myotonic dystrophy; this makes his muscles very weak. Myotonic dystrophy is a progressive and life-limiting condition. Mark is at risk of chest infections and sudden heart failure. He uses a specialized wheelchair for body support and transportation. The wheelchair supports his body frame and keeps him upright and secure with a safety belt. His head is rested against his right ear on a padded headrest. His mobility disability is extreme; however, he has some imprecise movement in his neck and arms. Mark likes interacting with computers; at school, he uses both eye gaze technology and switches to interact with computer interfaces. Mark uses his voice to communicate; he likes sharing his sense of humor, he laughs when things go wrong, and makes the sound 'uh-oh' to signal mistakes. He enjoys making choices and can become frustrated when he is not offered choices. Mark shows a definite progression with communication and is now very accepting of and participating in a wider variety of activities, events, and opportunities in school. He responds well to unforeseen changes in his daily routine, even with little notice.

4 Features and data points

While the participants were participating in the CPT, MM data, from eye gaze, EEG, body pose, and interaction data were collected. Data collected from the four participants is averaged by session, and by week. The data can be categorized by source, as seen in (1). These categories include SDT outcomes, CPT outcomes, MM data features, and participant characteristics. A summary of the MM features used, and other data points of interest are shown in Table 2. The data was used to develop a machine learning model of flow using a random forest classifier. An analysis of correlations between the data points is discussed in *Results (section 5).*

$$\text{Data point source} = \begin{cases} Platform \begin{cases} SDT\ outcomes \\ CPT\ outcomes \\ MM\ features \end{cases} \\ Participant \begin{cases} Age \\ P\ scales \end{cases} \end{cases} \tag{1}$$

Using SDT [4], [7], [71]–[75], quantifiable objective data on the improvement or deterioration of attention is collected and analyzed using SDT analysis as detailed in [73], [74]. SDT measures and outcomes [71] are calculated from the Seek-X CPT. These outcomes are detectability or sensitivity d' and bias B''_D which will be explained further in the paper.

The signal detection trial in the form of the Seek-X Type CPT was controlled to be aligned to the capabilities of our participants (this was rigorously checked before each session). This means that it is not a test of cognitive processing power but one of sustained attention, and capacity to maintain this only. The hit rate represents the probability of responding *yes* on signal slides, and the false-alarm rate is the probability of responding *yes* on the imitation target or distractor slides [76]. Hit rate (H) is the probability of a 'yes' response given the target is present. H can vary between 0 and 1. H is calculated in (2).

$$H = P\ (yes|present) = \frac{\text{Correct Commissions}}{Total\ Target\ slides} \tag{2}$$

False alarm rate (FAR) [also shown as F or FA] is the probability of a 'yes' response given the target is absent. FAR can vary between 0 and 1. FAR is calculated in (3).

$$FAR = P\ (yes|absent) = \frac{\text{Wrong Commissions}}{Total\ non\ Targets\ slides} \tag{3}$$

Detectability [7], otherwise known as sensitivity [77], is a measure of the quality of participant performance in a CPT or, in other words, the capacity of a participant to sustain attention and is measured by d'. Measured in standard deviation units [73], the formula for d' can be seen in (4), where z is the z-score, or the z transformation. A perfect score would be a Hit rate (H) of 1, a False Alarm Rate (FAR) of 0 and a d' of $+\infty$.

$$d' = z(H) - z(FAR) \tag{4}$$

Bias represented by B''_D [78] is a way of measuring if a participant is liberal or conservative in their commissions during trials. It is a good way of contrasting the participant outcome result against how much risk they are willing to take. Neutral bias (no bias) is indicated by $B''_D = 0$. Bias (B''_D) is neutral when $H + FAR = 1$. Negative bias numbers represent liberal bias, positive numbers represent conservative bias, and the maximum in either direction is 1. The formula for B''_D is represented in (5).

$$B''_D = \frac{[(1-H)(1-FAR) - H\times FAR]}{[(1-H)(1-FAR) + H\times FAR]} \tag{5}$$

Interaction data, such as keyboard presses, keypress duration, and speed have been used in studies to detect positive and negative affect states using self-reported assessment [79]–[86]. Alen *et al.* also used keystroke behavior to detect flow and boredom in students in an essay writing context [87]. Bixler and D'Mello used keystroke behavior to detect the affective states of 'engagement', 'boredom' and 'neutral' [86]. Valdez *et al.* used keystroke and mouse tracking features to detect the affective states including 'flow', 'boredom', and 'frustration' [88]. Fast response times are an indicator of participant attention [38], and greater button presses, a sign of higher arousal states [87].

In this study, we use similar features from the button presses of the participant. We extract two features; single fast button presses and repetitive button presses. Single fast button presses are calculated using the formula described in (6), which is only calculated if the participant presses the button only once during the allowed response time window. In other instances, the value for this feature is zero. The maximum press count is the second feature extracted from the button press data shown in (7). This value is calculated for only the allowed response time interval and is zero when the button is not pressed.

$$\text{Single fast press} = \frac{1}{response\ time} \tag{6}$$

$$\text{Max press count} = \text{total press attempts} \tag{7}$$

Studies show [89]–[92] that the EEG beta rhythm (14–30 Hz) is activated when the brain is in a state of arousal. In other EEG studies, mental fatigue related features are associated with decreased alpha band (8-13 Hz) power at one or more parietal locations (e.g., P7 and P8). Ning-Han Liu *et al.* [93] linked these two factors in their study and showed that alertness can be measured by the signal power of α divided by the signal power of β. Timothy McMahan *et al.* [94] also demonstrated that the ratio is related to arousal. We adopt the feature from [93] in our study.

We use Discrete Fourier Transform (DFT) to calculate the Power Spectral Density (PSD) ratio of the α and β time series. Using the Welch method (8), the ratio of the alpha band power f_α to the beta band power f_β can be simplified as in (10).

Welch Method:

$$\hat{p}d(f) = \frac{1}{MU} \left|\sum_{n=0}^{M-1} xd(n)w(n)e^{-j2\pi f}\right|^2 \tag{8}$$

U is the normalization factor for the Welch Method:

$$U = \frac{1}{M}\sum_{n=0}^{M-1}|w(n)| \tag{9}$$

The EEG Alertness feature:

$$\text{Alertness} = \frac{\hat{p}d(f_\alpha)}{\hat{p}d(f_\beta)} \tag{10}$$

Eye gaze serves as a marker of attention, the sustained emphasis of cognitive processing power on targeted information while ignoring distracting information [95]. From the eye gaze data, three features were extracted; 'eye scanning', 'eye dwelling', and 'eyes off-screen'. These features are commonly used in eye gaze technologies to understand attention, interest, and engagement [96], [97], [106], [107], [98]–[105]. We introduce formulas to calculate eye gaze scanning and dwelling, which are based on the definitions provided in [97]. All eye gaze features are impartial to the type of elements on the screen (target or not) and are calculated equally for signal and noise trials. The eye-scanning feature is calculated as in (11) and represents the sum of the inverse distance from the center of each element (where r_{in} is that distance; from the eye gaze location to the center of image i out of $I = 16$ total image elements, for sample n, out of N total discrete sensor samples). The relevant diagram is shown in Figure 4.

$$\text{Eye scanning} = \sum_{n=1}^{N}\sum_{i=1}^{I}\frac{1}{r_{in}} \tag{11}$$

Dwelling represents the eye gaze behavior of when the gaze stays relatively in the same position for a duration of time. This behavior is independently calculated from the location of the image elements on the screen. The dwelling feature is calculated as in (12), which is the sum of the inverse distance from each eye gaze position to the next (where n is the sample number out of N total discrete sensor samples, and Δd is the distance the eyes have moved since the previous sample), as demonstrated in Figure 5.

$$\text{Eye dwelling} = \sum_{n=1}^{N}\frac{1}{\Delta d_n} \tag{12}$$

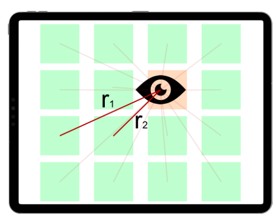


Figure 4. Scanning calculation with respect to the active elements on the screen

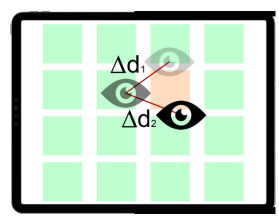


Figure 5. Dwelling calculation independent of active elements on the screen

Studies have shown that body posture and gesture can communicate affective modalities and also specific emotional categories [30], [108]–[115]. They have also been indicators of a firm or weak correlation of engagement during Human-Computer Interaction in gameplay [30], [109], [110], [113]. Numerous studies [116]–[121] have investigated the importance of body fidgeting in detecting attention for students with PMLD. Fidgeting is an indicator of the onset of attention loss, boredom, and engagement deterioration [113], [116], [122]–[124]. We calculate rapid body movement from body pose to assess fidgeting levels. The equation to extract this feature is seen in (13) (where Δd_j is the displacement vector of joint j out of N joints and Δt is the time passing between the displacement samples). Tracking of non-occluded joints; the head, neck, mid-spine, right and left shoulders and right and left hands are recorded.

$$\text{Body fidgeting} = \frac{1}{N}\sum_{j=1}^{N}\frac{|\Delta d_j|}{\Delta t} \quad (13)$$

To create an MM model of flow using a random forest classifier, High-Level Compound Features (HLCF) were developed in [44] using the Mudra methodology [125]. Compound features are constructed from high-level features and allow the representation of complex actions [126]. We propose two HLCF in this work and later assess their effectiveness. The first feature is a normalized mean of the features that traditionally serve as indicators of attention. The High-Level Attention feature is calculated as the mean of the normalized features of single fast presses, eye dwelling, eye scanning, and EEG alertness, which is seen in (14).

$$\text{HLA} = \frac{1}{4}(norm.Single\,fast\,press + norm.Dwelling + norm.Scanning + norm.Alertness) \quad (14)$$

High-Level Distraction feature is calculated as the mean of normalized features of body fidgeting, eyes off-screen, and press count as seen in (15).

$$\text{HLD} = \frac{1}{3}(norm.Body\,fidgeting + norm.Eyes\,off\,screen + norm.Max\,press\,count) \quad (15)$$

Table 2. Summary of platform datapoint sources and Bonferroni familywise correlation family

Data point	Source	Bonferroni Corr. family
d'(Sensitivity)	SDT	SDT
$B''_D(Bias)$	SDT	SDT
Hit rate	SDT	SDT
False Alarm Rate	SDT	SDT
Correct Omissions	SDT	SDT
Wrong Omissions	SDT	SDT
Response time	CPT	CPT
Press count	CPT	CPT
Single press	CPT	CPT
Eye scanning	Eye gaze	Physiological
Eye dwelling	Eye gaze	Physiological
Eyes off screen	Eye gaze	Physiological
EEG Alertness	EGG	Physiological
Body fidgeting	Body pose	Physiological
Body joint entropy	Body Pose	Physiological
Age	Participant	Academic
P scale Speaking	Participant	Academic
P scale Listening	Participant	Academic
P scale Reading	Participant	Academic
P scale Writing	Participant	Academic
P scale Maths using and applying	Participant	Academic

5. Results

We found that the best classification model was achieved with a random forest classifier using participant-specific data. The random forest model [127] using both high- and low-level features achieved an 84.8% classification for 'flow' and a 65.4% accuracy for 'non-flow'. This model's Area Under the Curve (AUC) has an 80.7% coverage. In a second approach, all data was put in a single group, and all participant data was considered as a whole. Random forest using both high- and low-level features achieved 93.3% classification for 'flow' and 42.9% accuracy for 'non-flow' with 80.3% AUC. HLCF was shown to improve the classification accuracy in every mode of data and every subset of sensors. The complete classification results are reported in an earlier study [44]. For the rest of this paper, we discuss the correlations discovered in the data.

Overall, 55 data points from the categories in (1) were assessed for cross-correlation significance. The data collected from the platform is then correlated against the participant characteristics. The normalized means of each data point per session were evaluated for correlations between other data points, and also against the passing of time, in session and week durations. In total 2475 cross-correlation assessments were evaluated using Pearson's correlation coefficient r [128]–[130]. To control for multiple comparison false positives we adjust using the Bonferroni correction [131] for the independent variables. The Bonferroni familywise groups can be seen in Table 2. The groups and their corrected critical values are as follows; 'SDT' (α = .0083), 'CPT' (α = .017), 'Physiological' (α = .0083) and 'Academic' (α = .0083).

We investigate the correlations between the data points in the category of 'platform' against themselves and the data points from 'platform' with 'participant'. Participant data points included characteristics such as age and their academic performance in 'P scales' [70]. The data points regarding learner P scales are 'P scales mean', 'Speaking', 'Listening', 'Reading', 'Writing', 'Maths using and applying' and 'P scale Computing'. The results of these tests showed a significant correlation between some of the data points (see Table 3), and a summary of the most interesting ones follows.

Participant progress in their academic curriculum (P scales mean) had a strong positive correlation with participant ability to maintain attention in the CPT (d'), ($r(57)$ = .986, p = .0068). This agrees with other studies that found that sustained attention related to greater academic outcomes [29], [39], [47], [49], [50]. Participant progress in their academic curriculum however, has a strong negative correlation with their impulsivity (FAR), ($r(57)$ = -.991, p = .0040). Participants who can maintain selective and sustained attention for longer, and do not give in to impulsive responses when fatigued, progressed further in their academic attainment targets. Participants that demonstrated greater inhibition (through higher Correct Omissions) progressed further in their P scales ($r(57)$ = .995, p = .0025). This selectivity also has a strong positive correlation with the participants' ability to maintain attention and engagement for longer (d'), ($r(57)$ = .994, p = .0029). This echoes other longitudinal studies that found that self-control and inhibition, increase academic performance [29], [40].

We found that students with moderate and severe intellectual disabilities did not have a significant change in their sensitivity or bias over time. Sensitivity (d') showed a negative relationship of ($r(57)$ = -.1406, p = .6175) to session number, which shows decrease of detectability, but not significantly. This confirms Swanson, Eliason and Richman's findings from three separate studies, which reported that students with learning disabilities retained their sensitivity (d') over time without significant change as did the mainstream control group of students in these studies [4], [7], [18]. We found that bias (B''_D) did not change significantly over time (after each session; $r(57)$ = .0005, p = .9985). In contrast, Swanson showed that students, regardless of having a learning disability, were more liberal on CPT sessions that were twice as long (p < .05) [7].

Studies [1], [4] have reported that children with Learning Disabilities are traditionally more conservative when compared to mainstream students [1], [4], [132], [133]. For example, Swanson's study found $B''_D = -.285$ (LD) > -.390 (mainstream) [7]; Eliason and Richman found $B''_D = .908$ (LD) > .865 (mainstream) [1]. In our study we found that out of the 59 sessions, 24 sessions had a positive bias, 33 had negative bias and 2 were neutral. Participant bias was overall liberal (negative) $\overline{B''_D}$ = -.1105, $\sigma(B''_D) = .6323$, as seen in Figure 6, however we did not have a mainstream population to compare to. Swanson argued that rather than considering students with learning disabilities as being developmentally delayed in inhibition, it is more appropriate to consider them as less "risk taking" [4]. Swanson reported that this decreases with age (p < .01), regardless of having a learning disability [4]. Even though we did find a negative correlation between bias (B''_D) and age, (r = -.680), it was not significant (p = .3204).

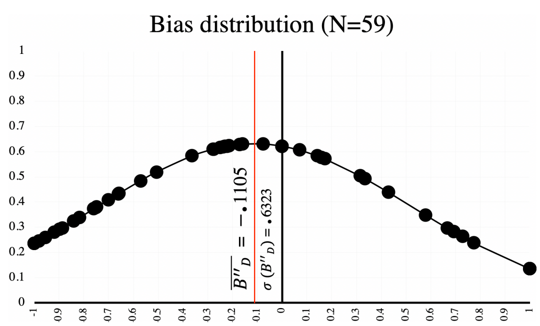


Figure 6. Bias distribution for participants with learning disability shows a slight negative outcome

While there are varied reports on bias (B''_D) and sensitivity (d') being independent variables in SDT [132]; Sostek *et al.* found a moderate positive relationship between the two (p < .01) [132], while Epstein *et al.* found a strong negative relationship (p < .0001) [134], Eliason and Richman reported the two as independent variables [18], as did Swanson (p > .05) [4], [7]. We found no significant correlation ($r(57)$ = -.138, p = .8615), and consider bias and sensitivity independent variables in our study. This confirms that data samples collected in this study meet the underlying assumptions of signal detection analysis [135]; that response bias changes are independent of participant sensitivity [136].

When the participants were more alert (higher EEG Alertness values) they made more Hits ($r(57)$ = .991, p = .0047) and had faster response times ($r(57)$ = -.975, p = .0125). Participants that had higher 'EEG Alertness' during their CPT, had progressed further in both P scales 'Maths using and applying' and 'Computing' ($r(57)$ = .988, p = .0060; $r(57)$ = .985, p = .0073 respectively). These two P scales had a direct correlation with greater Hit rate ($r(57)$ = .978, p = .0108; $r(57)$ = .976, p = .0117).

Participants that looked away from the screen less, had progressed further in their 'Listening' scales rate ($r(57)$ = .991, p = .0047). Swanson reported that there were no significant differences between the length of time students with learning disabilities looked away from the screen, compared to mainstream students [4].

Despite participants having quicker response times after every session ($r(57)$ = .908, $p = 4 \times 10^{-8}$), their responses did not become any more precise in the SDT challenge. Participant press count was significantly increased for all participants after each session ($r(57)$ = .644, p = .0015), which could indicate a higher arousal state [87].

Participants that had higher HLA feature values had more control and inhibition in the CPT activity (lower FAR) and were able to hold selective and sustained attention (d') for longer periods of time ($r(57)$ = -.989, p = .0054).

Some physiological correlations were discovered; however, they are significant only when we consider the traditional critical value of (α = .05). We include them here, as the relative size of the r compared to other correlations indicates an area worthy of further investigation. Participants that have greater body movement range progressed further in their 'Reading' scales ($r(57)$ = .916, p = .0420) and participants who had showed use of their eye movements in structured form (eye scanning and eye dwelling) had progressed further in their 'Listening' P scales ($r(57)$ = .915, p = .0426; $r(57)$ = .9400, p = .0300).

6 Conclusions

From the Multimodal Multisensor data collected, 2615 data samples were assessed for cross-correlation significance. Over 2475 separate correlation tests were carried out using Pearson's correlation coefficient r, N = 59. These results show significant correlations between SDT, CPT, and participant characteristics. Importantly, a strong positive correlation was found between participant ability to maintain sustained and selective attention and engagement in the CPT to their academic performance against attainment targets in school (d'), ($r(57)$ = .986, p = .0068). Participants who were less impulsive and more selective in the test also did better in their academic performance ($r(57)$ = .995, p = .0025).

The Seek-X Type CPT also showed that specific participant physiological characteristics (including body movement range and eye gaze behavior), were significant in their performance against academic attainment targets, such as 'Reading' and 'Listening' ($p < .05$, $p < .0083$). It seems possible that participants' ability to sustain attention has enabled them to demonstrate a wider range of performance that pupils with such disabilities who cannot access the national curriculum might characteristically demonstrate. At the same time, a higher level of cognitive functioning may enable better sustained attention, as well as the ability to progress at school. Given these correlations, the new type of CPT would appear an appropriate method by which to label Multimodal Multisensor data of students with moderate and severe intellectual disabilities to assess sustained attention. The CPT could be utilized in the future to help to assess and monitor participant performance, and also to help understand the specific challenges (such as physiological or cognitive impairments and limitations) that may hinder their academic performance. We posit that it could even be used as an outcome measure in other studies with students with moderate to severe intellectual disabilities. A system incorporating these models can help teachers track attention in students using the most appropriate set of sensors for that individual student, and use this information to offer a degree of personalization in the classroom. Every student would have the benefits of personalized tuition for at least part of every lesson, which would ensure that their own needs were individually addressed [137].

This research was conducted as part of a Ph.D. program of research at Nottingham Trent University (NTU) and has been adopted as part of the Erasmus+ KA201 Pathway+ project to determine the affective state of students with mild and moderate learning disabilities (2017-1-UK01-KA201-036761) [138].

Table 3. Pearson correlation r values between data points, (N = 59). Blue indicates positive correlation and red indicates negative correlation

	P scales	Speaking	Listening	Reading	Writing	Maths Space, Shape, Measure	Maths using and applying	Computing	Age	AHLF	DHLF	d'	B''_D	H	FAR	Miss %	Omission %	Gaze Inv-Entropy	Fidgeting	Joint entropy	EEG Attention	Eyes Off Screen	Eye Scanning	Eye Dwelling	Single Press	Press Count	Response Time
P scales	1.00	0.37	0.76	0.91	-0.31	0.55	0.88	0.88	0.66	1.00	0.67	0.99	-0.17	0.80	-0.99	-0.80	0.99	0.24	0.53	0.72	0.79	-0.71	0.49	0.61	0.52	0.11	0.46
Speaking	0.37	1.00	-0.24	0.30	-0.70	0.76	0.04	0.05	-0.45	0.38	-0.44	0.26	0.52	-0.17	-0.40	0.17	0.36	-0.79	-0.45	0.50	-0.05	0.23	-0.35	-0.22	0.76	-0.78	0.97
Listening	0.76	-0.24	1.00	0.56	-0.12	-0.12	0.71	0.72	0.87	0.74	0.93	0.85	-0.22	0.77	-0.77	-0.77	0.79	0.63	0.61	0.18	0.64	-0.99	0.91	0.94	-0.15	0.74	-0.06
Reading	0.91	0.30	0.56	1.00	0.03	0.72	0.96	0.96	0.68	0.92	0.63	0.84	-0.49	0.88	-0.85	-0.88	0.85	0.35	0.71	0.92	0.94	-0.47	0.19	0.30	0.70	-0.06	0.29
Writing	-0.31	-0.70	-0.12	0.03	1.00	-0.14	0.18	0.16	0.36	-0.31	0.25	-0.33	-0.88	0.30	0.42	-0.30	-0.39	0.68	0.62	0.07	0.33	0.21	-0.25	-0.34	-0.13	0.20	-0.85
Maths Space, Shape, Measure	0.55	0.76	-0.12	0.72	-0.14	1.00	0.51	0.51	-0.01	0.57	-0.08	0.40	-0.16	0.33	-0.50	-0.33	0.48	-0.30	0.16	0.92	0.49	0.20	-0.45	-0.32	1.00	-0.73	0.64
Maths using and applying	0.88	0.04	0.71	0.96	0.18	0.51	1.00	1.00	0.85	0.88	0.81	0.85	-0.62	0.98	-0.81	-0.98	0.83	0.59	0.86	0.79	0.99	-0.62	0.37	0.44	0.49	0.20	0.05
Computing	0.88	0.05	0.72	0.96	0.16	0.51	1.00	1.00	0.85	0.89	0.81	0.86	-0.61	0.98	-0.82	-0.98	0.84	0.58	0.85	0.79	0.99	-0.62	0.37	0.45	0.49	0.20	0.07
Age	0.66	-0.45	0.87	0.68	0.36	-0.01	0.85	0.85	1.00	0.65	0.99	0.71	-0.68	0.94	-0.61	-0.94	0.64	0.89	0.92	0.37	0.86	-0.80	0.66	0.66	-0.03	0.67	-0.37
AHLF	1.00	0.38	0.74	0.92	-0.31	0.57	0.88	0.89	0.65	1.00	0.66	0.98	-0.18	0.80	-0.99	-0.80	0.99	0.23	0.52	0.74	0.80	-0.69	0.47	0.59	0.55	0.09	0.47
IHLF	0.67	-0.44	0.93	0.63	0.25	-0.08	0.81	0.81	0.99	0.66	1.00	0.74	-0.57	0.90	-0.64	-0.90	0.67	0.86	0.86	0.29	0.80	-0.87	0.76	0.76	-0.11	0.73	-0.33
d'	0.99	0.26	0.85	0.84	-0.33	0.40	0.85	0.86	0.71	0.98	0.74	1.00	-0.14	0.80	-0.99	-0.80	0.99	0.31	0.53	0.60	0.76	-0.81	0.63	0.73	0.37	0.27	0.39
B''_D	-0.17	0.52	-0.22	-0.49	-0.88	-0.16	-0.62	-0.61	-0.68	-0.18	-0.57	-0.14	1.00	-0.70	0.05	0.70	-0.09	-0.81	-0.90	-0.45	-0.73	0.10	0.05	0.09	-0.15	-0.22	0.63
H	0.80	-0.17	0.77	0.88	0.30	0.33	0.98	0.98	0.94	0.80	0.90	0.80	-0.70	1.00	-0.73	-1.00	0.76	0.74	0.93	0.67	0.98	-0.68	0.46	0.51	0.31	0.38	-0.14
FAR	-0.99	-0.40	-0.77	-0.85	0.42	-0.50	-0.81	-0.82	-0.61	-0.99	-0.64	-0.99	0.05	-0.73	1.00	0.73	-1.00	-0.18	-0.43	-0.65	-0.71	0.74	-0.55	-0.66	-0.47	-0.14	-0.51
Miss %	-0.80	0.17	-0.77	-0.88	-0.30	-0.33	-0.98	-0.98	-0.94	-0.80	-0.90	-0.80	0.70	-1.00	0.73	1.00	-0.76	-0.74	-0.93	-0.67	-0.98	0.68	-0.46	-0.51	-0.31	-0.38	0.14
Omission %	0.99	0.36	0.79	0.85	-0.39	0.48	0.83	0.84	0.64	0.99	0.67	0.99	-0.09	0.76	-1.00	-0.76	1.00	0.22	0.47	0.64	0.74	-0.76	0.56	0.68	0.45	0.17	0.48
Gaze Inv-Entropy	0.24	-0.79	0.63	0.35	0.68	-0.30	0.59	0.58	0.89	0.23	0.86	0.31	-0.81	0.74	-0.18	-0.74	0.22	1.00	0.89	0.08	0.65	-0.57	0.51	0.44	-0.31	0.75	-0.75
Fidgeting	0.53	-0.45	0.61	0.71	0.62	0.16	0.86	0.85	0.92	0.52	0.86	0.53	-0.90	0.93	-0.43	-0.93	0.47	0.89	1.00	0.53	0.91	-0.51	0.33	0.33	0.15	0.44	-0.47
Joint entropy	0.72	0.50	0.18	0.92	0.07	0.92	0.79	0.79	0.37	0.74	0.29	0.60	-0.45	0.67	-0.65	-0.67	0.64	0.08	0.53	1.00	0.79	-0.08	-0.21	-0.09	0.91	-0.44	0.40
EEG Attention	0.79	-0.05	0.64	0.94	0.33	0.49	0.99	0.99	0.86	0.80	0.80	0.76	-0.73	0.98	-0.71	-0.98	0.74	0.65	0.91	0.79	1.00	-0.54	0.29	0.35	0.47	0.19	-0.06
Eyes Off Screen	-0.71	0.23	-0.99	-0.47	0.21	0.20	-0.62	-0.62	-0.80	-0.69	-0.87	-0.81	0.10	-0.68	0.74	0.68	-0.76	-0.57	-0.51	-0.08	-0.54	1.00	-0.96	-0.98	0.23	-0.76	0.03
Eye Scanning	0.49	-0.35	0.91	0.19	-0.25	-0.45	0.37	0.37	0.66	0.47	0.76	0.63	0.05	0.46	-0.55	-0.46	0.56	0.51	0.33	-0.21	0.29	-0.96	1.00	0.99	-0.48	0.86	-0.12
Eye Dwelling	0.61	-0.22	0.94	0.30	-0.34	-0.32	0.44	0.45	0.66	0.59	0.76	0.73	0.09	0.51	-0.66	-0.51	0.68	0.44	0.33	-0.09	0.35	-0.98	0.99	1.00	-0.35	0.78	0.02
Single Press	0.52	0.76	-0.15	0.70	-0.13	1.00	0.49	0.49	-0.03	0.55	-0.11	0.37	-0.15	0.31	-0.47	-0.31	0.45	-0.31	0.15	0.91	0.47	0.23	-0.48	-0.35	1.00	-0.75	0.63
Press Count	0.11	-0.78	0.74	-0.06	0.20	-0.73	0.20	0.20	0.67	0.09	0.73	0.27	-0.22	0.38	-0.14	-0.38	0.17	0.75	0.44	-0.44	0.19	-0.76	0.86	0.78	-0.75	1.00	-0.60
Response Time	0.46	0.97	-0.06	0.29	-0.85	0.64	0.05	0.07	-0.37	0.47	-0.33	0.39	0.63	-0.14	-0.51	0.14	0.48	-0.75	-0.47	0.40	-0.06	0.03	-0.12	0.02	0.63	-0.60	1.00

Colour scale: 1.0, 0.5, 0, -0.5, -1.0

REFERENCES

[1] M. J. Eliason and L. C. Richman, "The Continuous Performance Test in learning disabled and nondisabled children.," *J. Learn. Disabil.*, vol. 20, no. 10, pp. 614–9, Dec. 1987.

[2] F. Vrana and R. O. O. Pihl, "Selective attention deficit in learning disabled children: a cognitive interpretation.," *J. Learn. Disabil.*, vol. 13, no. 7, pp. 387–391, Aug. 1980.

[3] R. A. Dykman, P. T. Ackerman, S. D. Clements, and J. E. Peters, "Specific learning disabilities: An attentional deficit syndrome," in *Progress in Learning Disabilities*, pp. 56–93.

[4] H. L. Swanson, "A Developmental Study of Vigilance in Learning-disabled

and Nondisabled Children," *J. Abnorm. Child Psychol.*, vol. 11, no. 3, pp. 415–29, Sep. 1983.

[5] L. Swanson, "Modification of Comprehension Deficits in Learning Disabled Children," *Learn. Disabil. Q.*, vol. 4, no. 2, pp. 189–202, May 1981.

[6] J. K. Torgesen, "Studies of children with learning disabilities who perform poorly on memory span tasks.," *Journal of Learning Disabilities*, vol. 21, no. 10. SAGE PublicationsSage UK: London, England, pp. 605–612, 18-Dec-1988.

[7] H. L. Swanson, "Vigilance Deficit in Learning Disabled Children: A Signal Detection Analysis," *J. Child Psychol. Psychiatry.*, vol. 22, no. 4, pp. 393–9, Oct. 1981.

[8] S. G. Tarver and D. P. Hallahan, "Attention Deficits In Children With Learning Disabilities," *J. Learn. Disabil.*, vol. 7, no. 9, pp. 560–569, Nov. 1974.

[9] R. P. Anderson, C. G. Halcomb, and R. B. Doyle, "The Measurement of Attentional Deficits," *Except. Child.*, vol. 39, no. 7, pp. 534–539, Apr. 1973.

[10] R. P. Anderson, C. G. Halcomb, R. B. Doyle, C. Holcomb, R. D.-E. Children, and U. 1973, "Measurement of attention in hyperactive children," *Except. Child.*, vol. 39, no. 7, pp. 534–539, Apr. 1973.

[11] E. C. Noland and W. J. Schuldt, "Sustained Attention and Reading Retardation," *J. Exp. Educ.*, vol. 40, no. 2, pp. 73–76, Dec. 1971.

[12] E. Malkovsky, C. Merrifield, Y. Goldberg, and J. Danckert, "Exploring the Relationship Between Boredom and Sustained Attention," *Exp. Brain Res.*, vol. 221, no. 1, pp. 59–67, Aug. 2012.

[13] J. S. Warm, R. Parasuraman, and G. Matthews, "Vigilance requires hard mental work and is stressful.," *Hum. Factors*, vol. 50, no. 3, pp. 433–41, Jun. 2008.

[14] G. J. DuPaul, A. D. Anastopoulos, T. L. Shelton, D. C. Guevremont, and L. Metevia, "Multimethod Assessment of Attention-Deficit Hyperactivity Disorder: The Diagnostic Utility of Clinic-Based Tests," *J. Clin. Child Psychol.*, vol. 21, no. 4, pp. 394–402, Dec. 1992.

[15] R. L. Lindsay, T. Tomazic, M. D. Levine, and P. J. Accardo, "Attentional function as measured by a Continuous Performance Task in children with dyscalculia.," *J. Dev. Behav. Pediatr.*, vol. 22, no. 5, pp. 287–92, 2001.

[16] D. Delgado-Gomez *et al.*, "Microsoft kinect-based continuous performance test: An objective attention deficit hyperactivity disorder assessment," *J. Med. Internet Res.*, vol. 19, no. 3, pp. 1–8, 2017.

[17] X. Li and Q. Ji, "Active affective state detection and user assistance with dynamic Bayesian networks," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans.*, vol. 35, no. 1, pp. 93–105, 2005.

[18] C. A. Riccio, Cecil R. Reynolds, P. Lowe, and J. J. Moore, "The continuous performance test: a window on the neural substrates for attention?," *Arch. Clin. Neuropsychol.*, vol. 17, no. 3, pp. 235–272, Apr. 2002.

[19] P. V. Corkum and L. S. Siegel, "Is the Continuous Performance Task a Valuable Research Tool for use with Children with Attention-Deficit-Hyperactivity Disorder?," *J. Child Psychol. Psychiatry*, vol. 34, no. 7, pp. 1217–1239, 1993.

[20] F. Levy, "The Development of Sustained Attention (Vigilance) and Inhibition in Children: Some Normative Data," *J. Child Psychol. Psychiatry*, vol. 21, no. 1, pp. 77–84, 1980.

[21] M. Rosenberg, S. Noonan, J. Degutis, and M. Esterman, "Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task.," *Atten. Percept. Psychophys.*, vol. 75, no. 3, pp. 426–39, 2013.

[22] S. Raz, Y. Bar-Haim, A. Sadeh, and O. Dan, "Reliability and validity of the online continuous performance test among young adults.," *Assessment*, vol. 21, no. 1, pp. 108–18, 2014.

[23] G. P. Richards, "Sustained and Selective Attention in Children with Learning Disabilities," pp. 1–2, 2015.

[24] K. S. Lassiter, J. C. Whitten, A. N. Bardos, R. C. D'Amato, and D. J. Raggio, "The construct specificity of the continuous performance test: Does inattention relate to behavior and achievement?," *Dev. Neuropsychol.*, vol. 10, no. 2, pp. 179–188, 1994.

[25] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," *MindTrek - 12th Int. MindTrek Conf. Entertain. Media Ubiquitous Era*, no. January 2008, p. 13, 2008.

[26] A. Basawapatna, A. Repenning, and M. Savignano, "The Zones of Proximal Flow Tutorial Designing Computational Thinking Cliffhangers," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 2019, no. February, pp. 428–434.

[27] A. R. Basawapatna, A. Repenning, K. H. Koh, and H. Nickerson, "The zones of proximal flow," *Proc. ninth Annu. Int. ACM Conf. Int. Comput. Educ. Res. - ICER '13*, p. 67, 2013.

[28] M. Taheri *et al.*, "State Diagram for Affective Learning in an Educational Platform," pp. 4–6, 2018.

[29] M. Csikszentmihalyi, J. Nakamura, and S. Abuhamdeh, "Flow A General Context," *Handb. Competence Motiv. (pp. 598-608)*, pp. 598–608, 2005.

[30] N. Bianchi-Berthouze, W. W. Kim, and D. Patel, "Does Body Movement Engage You More in Digital Game Play? and Why?," in *Affective Computing and Intelligent Interaction*, 2007, pp. 102–113.

[31] S. Aslan *et al.*, "Investigating the Impact of a Real-time, Multimodal Student Engagement Analytics Technology in Authentic Classrooms," in *Conference on Human Factors in Computing Systems - Proceedings*, 2019.

[32] C. R. Henrie, L. R. Halverson, and C. R. Graham, "Measuring student engagement in technology-mediated learning: A review," *Comput. Educ.*, vol. 90, pp. 36–53, 2015.

[33] M. J. Lee and B. Ferwerda, "Personalizing online educational tools," *Humaniz. 2017 - Proc. 2017 ACM Work. Theory-Informed User Model. Tailoring Pers. Interfaces, co-located with IUI 2017*, pp. 27–30, 2017.

[34] A. K. |Wiggins. J. B. |Grafsgaard. J. F. |Boyer. K. E. E. N. |Lester. J. C. Vail, "The Affective Impact of Tutor Questions: Predicting Frustration and Engagement," *Int. Educ. Data Min. Soc.*, 2016.

[35] M. Csikszentmihalyi, "Creativity: Invention, Flow and the psychology of discovery and Invention," in *Personnel Psychology*, vol. 51, no. 3, 1997, p. 16.

[36] P. H. Mirvis and M. Csikszentmihalyi, "Flow: The Psychology of Optimal Experience," *Acad. Manag. Rev.*, vol. 16, no. 3, pp. 636–640, 1991.

[37] S. Craig, A. Graesser, J. Sullins, and B. Gholson, "Affect and learning: An exploratory look into the role of affect in learning with AutoTutor," *J. Educ. Media*, vol. 29, no. 3, pp. 241–250, 2004.

[38] X. Li and Q. Ji, "Active affective state detection and user assistance with dynamic Bayesian networks," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans.*, vol. 35, no. 1, pp. 93–105, 2005.

[39] E. A. Locke, K. N. Shaw, L. M. Saari, and G. P. Latham, "Goal setting and task performance: 1969-1980," *Psychol. Bull.*, vol. 90, no. 1, pp. 125–152, 1981.

[40] A. Possamai and A. J. Blasi, "The Protestant Ethic and the Spirit of Capitalism, by Max Weber," *The SAGE Encyclopedia of the Sociology of Religion.* 2020.

[41] K. Karpouzis, "Affective and gameful interfaces," in *International Conference on Intelligent User Interfaces, Proceedings IUI*, 2017, pp. 33–34.

[42] G. Hookham and K. Nesbitt, “A Systematic Review of the Definition and Measurement of Engagement in Serious Games,” in *ACM International Conference Proceeding Series*, 2019.

[43] E. Gottardo and A. R. Pimentel, “Affective human-computer interaction in educational software: A hybrid model for emotion inference,” in *ACM International Conference Proceeding Series*, 2017.

[44] M. Taheri, D. Brown, and N. Sherkat, “Modeling Engagement with Multimodal Multisensor Data: The Continuous Performance Test as an Objective Tool to Track Flow,” in *ICACII 2020: 14. International Conference on Affective Computing and Intelligent Interaction*, 2020, pp. 1–13.

[45] S. K. Perry, *Writing in flow: Keys to enhance creativity*, 1st ed. Cincinnati Ohio: Writer's Digest Books, 1999.

[46] K. Sawyer, “Improvisational Creativity: An Analysis of Jazz Performance,” *Creat. Res. J.*, vol. 5, no. 3, pp. 253–263, 1992.

[47] M. Csikszentmihalyi, K. R. Rathunde, S. Whalen, and M. S. M. Wong, “Talented teenagers : the roots of success and failure,” p. 307, 1993.

[48] S. A. Jackson, P. R. Thomas, H. W. Marsh, and C. J. Smethurst, “Relationships between Flow, Self-Concept, Psychological Skills, and Performance,” *J. Appl. Sport Psychol.*, vol. 13, no. 2, pp. 129–153, 2001.

[49] G. L. Stein, J. C. Kimiecik, J. Daniels, and S. A. Jackson, “Psychological antecedents of flow in recreational sport. Personality and Social Psychology Bulletin,” *Socirty Personal. Soc. Pscycology, Inc.*, vol. 21, no. 2, pp. 125–135, 1995.

[50] C. Heine, “Flow and Achievement in Mathematics,” University of Chicago, Chicago, IL, 1996.

[51] E. A. Linnenbrink and P. R. Pintrich, “Achievement goal theory and affect: An asymmetrical bidirectional model,” *Educational Psychologist*, vol. 37, no. 2. Lawrence Erlbaum Associates Inc., pp. 69–78, 2002.

[52] D. E. Berlyne, *Conflict, arousal, and curiosity*. McGraw-Hill Book Company, 2006.

[53] J. M. Hunt, “Intrinsic motivation and its role in psychological development,” in *Nebraska symposium on motivation*, 1965, vol. 13, pp. 189–282.

[54] J. Nakamura and M. Csikszentmihalyi, “The Concept of Flow Optimal Experience and Its Role in Development,” *Handb. Posit. Psychol.*, pp. 89–105, 2002.

[55] M. Chen, B. Kolko, E. Cuddihy, and E. Medina, “Modelling and measuring engagement in computer games,” in *The annual conference for the Digital Games Research Association (DiGRA)*, 2005.

[56] M. Csikszentmihalyi, “Applications of flow in human development and education: The collected works of Mihaly Csikszentmihalyi,” *Appl. Flow Hum. Dev. Educ. Collect. Work. Mihaly Csikszentmihalyi*, pp. 1–494, 2014.

[57] M. Csikszentmihalyi and I. Csikszentmihalyi, *Optimal experience: Psychological studies of flow in consciousness*. 1992.

[58] J. O. Goldberg and M. M. Konstantareas, “Vigilance in Hyperactive and Normal Children on a Self-Paced Operant Task,” *J. Child Psychol. Psychiatry*, vol. 22, no. 1, pp. 55–63, Jan. 1981.

[59] H. E. Rosvold, A. F. Mirsky, I. Sarason, E. D. Bransome, and L. H. Beck, “A continuous performance test of brain damage.,” *J. Consult. Psychol.*, vol. 20, no. 5, pp. 343–350, 1956.

[60] “Be first to get a Tobii EyeX Dev Kit Want to develop for Android or Linux ? Get a,” vol. 595, p. 795.

[61] I. Software, “Muse : The Brain Sensing Headband Tech Spec Sheet.”

[62] “MATLAB - The Language of Technical Computing - MathWorks United Kingdom.” [Online]. Available: http://uk.mathworks.com/products/matlab/. [Accessed: 07-Mar-2016].

[63] M. Teplan, “Fundamentals of EEG measurement,” *Meas. Sci. Rev.*, vol. 2, pp. 1–11, 2002.

[64] “Tobii EyeX Controller – get your own eye tracker.” 30-Jul-2015.

[65] “An introduction to Tobii EyeX | Tobii Developer Zone.” [Online]. Available: http://developer.tobii.com/an-introduction-to-the-tobii-eyex-sdk/. [Accessed: 03-Mar-2016].

[66] “Specifications for EyeX – Tobii Eye Tracking Support.” [Online]. Available: https://help.tobii.com/hc/en-us/articles/212818309-Specifications-for-EyeX. [Accessed: 10-Apr-2019].

[67] “Kinect hardware.” [Online]. Available: https://dev.windows.com/en-us/kinect/hardware. [Accessed: 03-Mar-2016].

[68] Q. Wang, Q. Wang, and G. Kurillo, “Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect,” no. November, 2015.

[69] “Teacher assessment: using P scales to report on pupils with SEN - Detailed guidance - GOV.UK.” [Online]. Available: https://www.gov.uk/guidance/teacher-assessment-using-p-scales. [Accessed: 07-Mar-2016].

[70] “P scales: attainment targets for pupils with SEN - Publications - GOV.UK,” *Department for Education.* [Online]. Available: https://www.gov.uk/government/publications/p-scales-attainment-targets-for-pupils-with-sen. [Accessed: 07-Mar-2016].

[71] J. I. Marcum, “A statistical theory of target detection by pulsed radar,” Dec. 1947.

[72] W. Peterson, T. Birdsall, and W. Fox, “The theory of signal detectability,” *Trans. IRE Prof. Gr. Inf. Theory*, vol. 4, no. 4, pp. 171–212, Sep. 1954.

[73] N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide*. Psychology Press, 2005.

[74] J. A. Swets and R. M. Pickett, *Evaluation of diagnostic systems : methods from signal detection theory*. Academic Press, 1982.

[75] W. P. Tanner and J. A. Swets, “A decision-making theory of visual detection,” *Psychol. Rev.*, vol. 61, no. 6, pp. 401–409, 1954.

[76] H. Stanislaw and N. Todorov, “Calculation of signal detection theory measures,” *Behav. Res. Methods, Instruments, Comput.*, vol. 31, no. 1, pp. 137–149, Mar. 1999.

[77] J. A. Swets and D. M. Green, *Signal detection theory and psychophysics*. Los Altos, Calif. : Peninsula Publishing, 1988.

[78] W. Donaldson, “Measuring Recognition Memory,” *J. Exp. Psychol. Gen.*, vol. 121, no. 3, pp. 275–277, 1992.

[79] S. Salmeron-Majadas, O. C. Santos, and J. G. Boticario, “An evaluation of mouse and keyboard interaction indicators towards non-intrusive and low cost affective modeling in an educational context,” in *Procedia Computer Science*, 2014, vol. 35, no. C, pp. 691–700.

[80] A. Pimenta, D. Carneiro, P. Novais, and J. Neves, “Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 8073 LNAI, pp. 222–231.

[81] J. Estrada, J. Buhia, A. Guevarra, and M. R. Forcado, *Keyboard and Mouse: Tools in Identifying Emotions During Computer Activities*. 2016.

[82] L. Vea and M. M. Rodrigo, “Modeling negative affect detector of novice

programming students using keyboard dynamics and mouse behavior," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10004 LNAI, pp. 127–138, 2017.

[83] T. S. Saini and M. Bedekar, "Inferring user emotions from keyboard and mouse," in *Advances in Intelligent Systems and Computing*, 2018, vol. 673, pp. 591–601.

[84] K. Bakhtiyari, M. Taghavi, and H. Husain, "Hybrid affective computing—keyboard, mouse and touch screen: from review to experiment," *Neural Comput. Appl.*, vol. 26, no. 6, pp. 1277–1296, 2015.

[85] H. Liu, O. N. Newton Fernando, and J. C. Rajapakse, "Predicting Affective States of Programming Using Keyboard Data and Mouse Behaviors," *2018 15th Int. Conf. Control. Autom. Robot. Vision, ICARCV 2018*, pp. 1408–1413, Nov. 2018.

[86] R. Bixler and S. D'Mello, "Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits," in *International Conference on Intelligent User Interfaces, Proceedings IUI*, 2013, pp. 225–233.

[87] L. K. Allen, C. Mills, M. E. Jacovina, S. Crossley, S. D'mello, and D. S. Mcnamara, "Investigating boredom and engagement during writing using multiple sources of information: The essay, the writer, and keystrokes," in *ACM International Conference Proceeding Series*, 2016, vol. 25-29-Apri, pp. 114–123.

[88] M. Garcia Valdez, J.-J. Merelo, A. Hernandez Aguila, and A. Mancilla Soto, *Mining of Keystroke and Mouse Dynamics to Increase the Engagement of Students with Programming Assignments*, vol. 56, no. 7. 2019.

[89] L. Trejo *et al.*, "EEG-based estimation of mental fatigue: Convergent evidence for a three-state model," *Found. Augment. Cogn.*, pp. 201–211, 2007.

[90] A. Gevins and M. E. Smith, "Detecting transient cognitive impairment with EEG pattern recognition methods," *Aviat. Sp. Environ. Med.*, vol. 70, no. 10, pp. 1018–1024, Oct. 1999.

[91] T. C. Hankins and G. F. Wilson, "A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight.," *Aviat. Space. Environ. Med.*, vol. 69, no. 4, pp. 360–7, Apr. 1998.

[92] D. Ming *et al.*, "Electroencephalograph (EEG) signal processing method of motor imaginary potential for attention level classification," *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009*, vol. 2009, pp. 4347–4351, Jan. 2009.

[93] N. H. Liu, C. Y. Chiang, and H. C. Chu, "Recognizing the degree of human attention using EEG signals from mobile sensors.," *Sensors (Basel).*, vol. 13, no. 8, pp. 10273–10286, Jan. 2013.

[94] T. McMahan, I. Parberry, and T. D. Parsons, "Evaluating Player Task Engagement and Arousal Using Electroencephalography," *Procedia Manuf.*, vol. 3, no. Ahfe, pp. 2303–2310, 2015.

[95] J. R. Anderson, *Cognitive Psychology an Its Implications*, 7th ed. New York, New York, USA: Worth Publishers, 2005.

[96] J. A. Vigo, "Eye Gaze Tracking for Tracking Reading Progress," p. 100, 2013.

[97] P. Fattahi, "Eyetracking Definitions: Basic Definition of Terms," Teltow, Germany, 2011.

[98] D. C. Newton and W. R. Knecht, "Current Status of Gaze Control Research and Technology Literature Review," *DOT Natl. Transp. Integr. Search - ROSA P*, no. January, 2017.

[99] P. Majaranta and K. J. Räihä, "Twenty years of eye typing: Systems and design issues," *Eye Track. Res. Appl. Symp.*, no. January, pp. 15–22, 2002.

[100] S. Brams *et al.*, "Does effective gaze behavior lead to enhanced performance in a complex error-detection cockpit task?," *PLoS One*, vol. 13, no. 11, pp. 1–25, 2018.

[101] A. Bulling and T. O. Zander, "Cognition-Aware Computing," *IEEE Pervasive Comput.*, vol. 13, no. 3, pp. 80–83, 2014.

[102] M. B. Almourad, E. Bataineh, J. Stocker, and F. Marir, "Analyzing the Behavior of Autistic and Normal Developing Children Using Eye Tracking Data," in *Advances in Intelligent Systems and Computing*, 2018, vol. 739, pp. 340–349.

[103] A. C. Schütz, D. I. Braun, and K. R. Gegenfurtner, "Eye movements and perception: A selective review.," *J. Vis.*, vol. 11, no. 5, pp. 1–30, 2011.

[104] A. Borji and L. Itti, "State-of-the-art in visual attention modelling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185–207, 2013.

[105] M. Csíkszentmihályi, *Flow*. 2012.

[106] C. Ehmke and S. Wilson, "Identifying web usability problems from eye-tracking data," *Hci 2007*, pp. 119–128, 2007.

[107] P. Simon, F. Rousseau, and J.-C. Angue, "Quantitative analysis of mental workload influence on eye scanning movements," *Syst. Man Cybern.*, vol. 4, pp. 707–712, 1993.

[108] N. Savva and N. Bianchi-Berthouze, "Automatic recognition of affective body movement in a video game scenario," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng.*, vol. 78 LNICST, pp. 149–159, 2012.

[109] N. Savva, A. Scarinzi, and N. Bianchi-Berthouze, "Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience," *IEEE Trans. Comput. Intell. AI Games*, vol. 4, no. 3, pp. 199–212, 2012.

[110] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Trans. Affect. Comput.*, vol. 4, no. 1, pp. 15–33, Jan. 2013.

[111] A. Patwardhan and G. Knapp, "Multimodal Affect Recognition using Kinect," 2016.

[112] H. Gunes and M. Piccardi, "Affect recognition from face and body: early fusion vs. late fusion," *Syst. Man Cybern. 2005 IEEE Int. Conf.*, vol. 4, pp. 3437–3443, 2005.

[113] N. Bianchi-Berthouze and Nadia Bianchi-Berthouze, "Understanding the role of body movement in player engagement," *Human-Computer Interact.*, vol. 28, no. 1, pp. 40–75, 2013.

[114] N. Bianchi-Berthouze, "The Affective Body Argument in Technology Design," in *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*, 2016, pp. 3–6.

[115] A. Kleinsmith and N. Bianchi-Berthouze, "Recognizing Affective Dimensions from Body Posture," in *Affective Computing and Intelligent Interaction*, 2007, pp. 48–58.

[116] J. Farley, E. F. Risko, and A. Kingstone, "Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering," *Front. Psychol.*, vol. 4, no. SEP, p. 619, Sep. 2013.

[117] M. D. Smith and M. S. Barrett, "The Effect of Parent Training on Hyperactivity and Inattention in Three School-Aged Girls with Attention Deficit Hyperactivity Disorder," *Child Fam. Behav. Ther.*, vol. 24, no. 3, pp. 21–35, Oct. 2002.

[118] C. K. Conners, L. Eisenberg, and A. Barcai, "Effect of Dextroamphetamine on Children," *Arch. Gen. Psychiatry*, vol. 17, no. 4, p. 478, Oct. 1967.

[119] P. L. Holborow and P. S. Berry, "Hyperactivity and Learning Difficulties," *J. Learn. Disabil.*, vol. 19, no. 7, pp. 426–431, Aug. 1986.

[120] E. Benedetto-Nasho and R. Tannock, "Math computation, error patterns and stimulant effects in children with Attention Deficit Hyperactivity

Disorder," *J. Atten. Disord.*, vol. 3, no. 3, pp. 121–134, Oct. 1999.

[121] S. L. Smith, *No easy answer : the learning disabled child at home and at school.* Bantam Books, 2012.

[122] B. Fenesi, K. Lucibello, J. A. Kim, and J. J. Heisz, "Sweat So You Don't Forget: Exercise Breaks During a University Lecture Increase On-Task Attention and Learning," *J. Appl. Res. Mem. Cogn.*, vol. 7, no. 2, pp. 261–269, 2018.

[123] S. Lis, N. Baer, C. Stein-En-Nosse, B. Gallhofer, G. Sammer, and P. Kirsch, "Objective measurement of motor activity during cognitive performance in adults with attention-deficit/hyperactivity disorder," *Acta Psychiatr. Scand.*, vol. 122, no. 4, pp. 285–294, 2010.

[124] A. M. F. Zinno, G. Douglas, S. Houghton, V. Lawrence, J. West, and K. Whiting, "Hyperactivity Disorder (ADHD) during computer video game play," *Br. J. Educ. Technol.*, vol. 32, no. 5, pp. 607–618, 2001.

[125] L. Hoste, B. Dumas, and B. Signer, "Mudra: A Unified Multimodal Interaction Framework," 2011, pp. 97–104.

[126] A. Gilbert, J. Illingworth, and R. Bowden, "Action Recognition Using Mined Hierarchical Compound Features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 883–897, May 2011.

[127] "3.2.4.3.1. sklearn.ensemble.RandomForestClassifier — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. [Accessed: 23-May-2020].

[128] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, vol. 64, no. 6. 1974.

[129] P. Y. Chen, M. Smithson, and P. M. Popovich, *Correlation: Parametric and Nonparametric Measures*, vol. 1. 2002.

[130] K. F. Weaver, V. C. Morales, S. L. Dunn, K. Godde, and P. F. Weaver, *An Introduction to Statistical Analysis in Research: With Applications in the Biological and Life Sciences.* 2017.

[131] J. H. McDonald, *Handbook of Biological Statistics*, 3rd. Baltimore, Maryland, U.S.A.: Sparky House Publishing, 2014.

[132] A. J. Sostek, M. S. Buchsbaum, and J. L. Rapoport, "Effects of amphetamine on vigilance performance in normal and hyperactive children," *J. Abnorm. Child Psychol.*, vol. 8, no. 4, pp. 491–500, 1980.

[133] C. A. Riccio, J. J. M. Waldrop, C. R. Reynolds, and P. Lowe, "Effects of stimulants on the continuous performance test (CPT): implications for CPT use and interpretation.," *J. Neuropsychiatry Clin. Neurosci.*, vol. 13, no. 3, pp. 326–335, Aug. 2001.

[134] J. N. Epstein, A. Erkanli, C. K. Conners, J. Klaric, J. E. Costello, and A. Angold, "Relations between Continuous Performance Test performance measures and ADHD behaviors.," *J. Abnorm. Child Psychol.*, vol. 31, no. 5, pp. 543–54, Oct. 2003.

[135] D. Broadbent, *Decision and stress.* London ;New York: Academic P, 1971.

[136] R. E. Pastore and C. J. Scheirer, "Signal detection theory: Considerations for general application," *Psychol. Bull.*, vol. 81, no. 12, pp. 945–958, 1974.

[137] N. Dong and Z. Chen, "Seldon, A., and Abidoye, O.: The fourth education revolution: will artificial intelligence liberate or infantilise humanity," *High. Educ.*, Jan. 2020.

[138] "Mobile Pedagogical Assistant to develop meaningful pathways to personalised learning." [Online]. Available: https://pedagogics-pathway.eu/. [Accessed: 09-Apr-2020].

State Diagram for Affective Learning in an Educational Platform

Mohammad H. Taheri[1], David J. Brown[2], Penny Standen[3], Nasser Sherkat[4], Caroline Langensiepen[5], Georgina Cosma[6]

[1,2,5,6]Science and Technology, Nottingham Trent University,

50 Shakespeare Street, Nottingham, UK

[3]School of Medicine, Nottingham University,

Clifton Boulevard, Nottingham, UK

[4]Engineering and Mathematics, Sheffield Hallam University,

Howard Street, Sheffield, UK

[1]*mohammad.taheri@ntu.ac.uk*, [2]*david.brown@ntu.ac.uk*, [3]*p.standen@nottingham.ac.uk*, [4]*n.sherkat@shu.ac.uk*, [5]*caroline.langensiepen@ntu.ac.uk*, [6]*georgina.cosma@ntu.ac.uk*

[1,2,5,6]*www.ntu.ac.uk*, [3]*www.nottingham.ac.uk*, [4]*www.shu.ac.uk*

ABSTRACT

The impact of learner affect state in goal achievement and educational performance is discussed. Different learning theory models are explored and compared, a new combined state diagram for modelling learning challenge and skill level is adopted and expanded for the purpose of an adaptive learning platform. To accomplish this, a thorough understanding of affect state diagrams for learning and the relationship of affect, learner skill and learning material challenge is established. Optimal learning paths are explored and objective ways of tracking learner skill are and increasing learning challenge linearly is explained and demonstrated. To conclude, a proposal for an experimental study is presented to explore the impact that an adaptive affect conscious learning platform has on learning and learner engagement.

1. INTRODUCTION

There is now an accumulation of evidence to indicate the link between affect and cognitive performance and decision-making (Eysenck, Derakshan, Santos, & Calvo, 2007). The goal to learn and understand is associated with an increase in positive emotions like enjoyment of learning as well as a decrease in negative emotions like boredom. Affect can direct attention and influence the level of that attention. According to Thompson and McGill (Thompson & McGill, 2017), affect also functions as a motivator, influencing the tendency to approach or avoid a situation as well as how information is processed.

Student engagement (participation in learning) was found to be the most reliable feature for determining successful learning (Barry Carpenter et al., 2015; Iovannone, Dunlap, Huber, & Kincaid, 2003). Without engagement, deep learning is not possible (Hargreaves, 2006). Effective personalized learning was shown to encourage participation and engagement not only in the classroom but in extra-curricular activities and work related learning in the local community (Sebba, Brown, Steward, Galton, & James, 2007). As the tutor or the technological learning facilitator forms a better understanding of the learners' strengths and challenges, they are in a better position to go through scaffolding objectives, involving choice of skill to train at a given moment and choice of learning activities, while preserving the learners' interest and engagement (Dolan & Hall, 2001).

According to Carpenter ("Engaging children with complex learning difficulties and disabilities in the Primary Classroom. Barry Carpenter," 2011) the process of engagement is a journey that connects learners and their environment (including people, ideas, materials and concepts) and enables learning and achievement. Students who are disengaged can become frustrated or bored, which can have a negative effect on achievement and lead to disruption of learning, for the individual learner, as well as for other learners when learning takes place in a collective/collaborative environment like a classroom.

The existence of the link between affect state and achievement suggests that a learning session may be improved if the teacher is sensitive and responsive to the emotional state of the learner (Goleman, 1995). However, the success of this strategy depends on the skill and experience of the human tutor and there is evidence to suggest that, especially with learners with special needs, teachers may find it particularly challenging to determine affect state (Vos et al., 2012). This was the motivation behind the MaTHiSiS (Managing Affective-learning THrough Intelligent atoms and Smart InteractionS) ("MaTHiSiS Project Website,"

©2018 ICDVRAT; ISBN 978-0-7049-1548-0

n.d.), an educational platform developed as part of a H2020 project. MaTHiSiS aims to deliver personalized and adaptive learning to range of user groups. The MaTHiSiS system introduces two novel elements into the teaching situation: Smart Learning Atoms and an affective responsive delivery of learning materials.

Herein, we focus on the affect states identified as important by D'Mello, Picard (Mello, Picard, & Graesser, n.d.) who identified frustration, boredom and flow to be the most relevant emotions to skill acquisition. The concept of the Zone of Proximal Flow proposed by Basawapatna et al (Basawapatna, Repenning, Koh, & Nickerson, 2013) reflects these affect states in a two dimensional diagram of learner skill and learning challenge. Referencing independent learning limit and scaffolding from Vygotsky's Zone of Proximal Development (ZPD) with Csikszentmihalyi et. al.'s (Csíkszentmihályi, 2012) theory of flow. Carpenter (B. Carpenter, 2010) and Iovannone et al. (Blackburn, 2012) see engagement as the single best predictor of successful learning for children with intellectual disabilities. Capturing a range of raw data (e.g. eye gaze, body pose and movement, vocalisation), multimodal fusion, labelling, and inference related to learners' affect state allows us to model affect states from a wide range of types of learner from those with Profound and Multiple Learning Disabilities (PMLD) and autism through to adult learners on a career guidance course. In a separate paper (Boulton et al., 2018) the use of multimodalities in tracking learner engagement has been explored. These concepts underlie the learning vision in providing an engaging learning environment in which learners with diverse needs and varying levels of ability are supported with assisted learning.

The relationship between affect state and learning achievement is crucial for the development for the affect based learning platform. This is the subject of this paper based on an examination of the literature and extension of Vygotsky's Zone of Proximal Development (ZPD) and Csikszentmihalyi's theory of flow. To this end we extend the work of other theorists who have already combined these two theories (Basawapatna et al., 2013).

The paper structure is as follows: Section 2 described the affect states for learning. Section 3, describes a linear approach to increasing learning challenge and delivers a formula for the objective calculation of learner skill in a linear learning task. Lastly, Section 4 explores the optimal pathway through an affective state learning diagram and proposes an experimental study to demonstrate the impact of the MaTHiSiS system, which incorporates an affective state diagram for learning.

2. AFFECTIVE STATES OF LEARNING

Affect states are defined as neurophysiological states best described as moods and emotion (Shernoff et al., 2003). Some affect states relevant to learning include frustration, boredom, flow, eureka (Craig, Graesser, Sullins, & Gholson, 2004). Classroom-related affective states are linked to the students' goal structure and their adoption of specific achievement goal orientations. The goal to learn and understand is associated with an increase in positive emotions like enjoyment of learning as well as a decrease in negative emotions like boredom. Adopting a performance approach goal—that is, the goal to be better than others—was found to be associated with the positive emotions. In contrast, the adoption of a performance avoidance goal—that is, a goal not to appear incompetent, stupid, or uninformed in comparison to others—was related to a negative emotion like anxiety and hopelessness. However, the relation between goals and affect might not be a unidirectional but a reciprocal one as proposed in Linnenbrink and Pintrich's bidirectional model.

Figure 1: Linnenbrink and Pintrich's asymmetrical bidirectional model of achievement goals and affect.

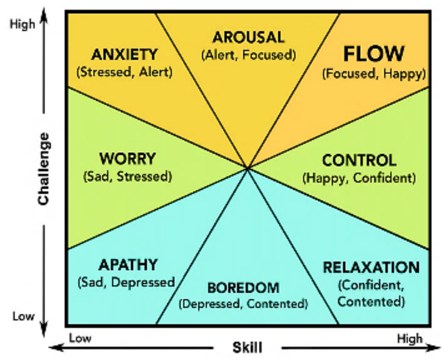


Figure 2: Csikszentmihalyi's theory of flow states[1]

In 2002, Linnenbrink and Pintrich described a model of affect in which goal achievement is reciprocally related to the learner's emotional state (Pekrun & Linnenbrink-Garcia, 2010). In this model (see **Figure 1**) the learners' personal goals are highly influenced by their perception of the learning activity challenge. This perception in turn has a direct influence on their affect state. Based on the wider literature, positive moods predict goal endorsement while negative moods predict avoidance goal endorsement.

[1] Frank Vandeven, accessed http://frankvandeven.com/getting-into-the-flow-what-does-that-mean/ on 7/3/2018

This relationship between skill and affect states has been more specifically described in Csikszentmihalyi's Theory of Flow (Mihalyi Csikszentmihalyi, 1996), where learner skill and their perception of the task challenge leads the learner to a variety of affect states, which he presented in **Figure 2**. Importantly, not all emotions are relevant to learning and parts of the theory of flow are less relevant to the scaffolding process in identifying optimal learning experience and moment where the learner requires scaffolding intervention. Sidney D'Mello and Rosalind Picard (Mello et al., n.d.) conducted a study on the relevance of emotions to learning and found Frustration, Boredom and Flow to be the most relevant emotions to skill acquisition. This has reduced the focus of the theory of flow to the most relevant and influential states of affect for learning.

In 1978, Vygotsky investigated the advancement of cognitive understanding by becoming interested in the process (L. Vygotsky, 1978). The boundaries of learner skill were broken into segments, where learners have the capacity to learn independently, and assisted learning (instructional scaffolding) from a tutor or a more knowledgeable peer, the later called the 'Zone of Proximal Development'.

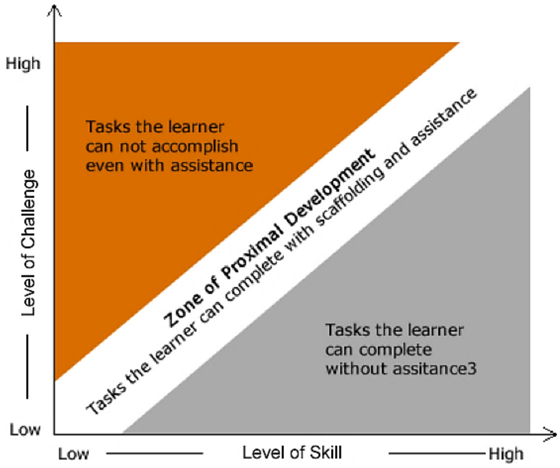


Figure 3: Morsink's graphical representation of The Zone of Proximal Development[2]

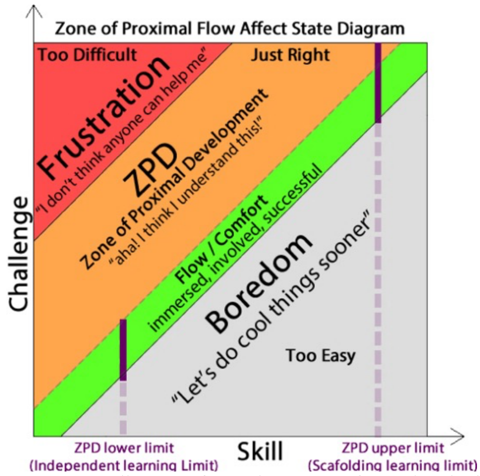


Figure 4: State diagram from the Zone of Proximal Flow theory adapted from Basawapatna et al (Basawapatna et al., 2013) to include both the ZPD lower independent and upper scaffolding learning limits.

Only later, in 2013, was it that Basawapatna et al. (Basawapatna et al., 2013) combined learner skill, independent learning limit and scaffolding in the 'Zones of Proximal Flow' (ZPF) state change diagram. Critically this work provided the first state change diagram to reference both Vygotsky's ZPD and the affect states from Csikszentmihalyi's theory of flow. Moreover, to adapt this diagram to facilitate an educational platform, knowing the imitations of the individual learning is important in designating when the platform should mediate and deliver scaffolding intervention. To this aim, the ZPD limit to independent learning from Vygotsky's theory has been applied to the ZPF diagram and we introduce **Figure 3** as the more complete affect state diagram for learning. The learner's skill level is displayed as the X-Axis and the task challenge is displayed as the Y- Axis. Unlike Csikszentmihalyi's flow diagram, or Vygotsky's ZPD, a single ZPF graph can be used to track the learner's progress in a learning activity and any permutations of level of skill or task difficulty.

A learning experience with a learning platform comprises interactions with the learning material, the 'challenge' of an activity, as depicted in the diagram of **Figure 3** and will consist of learning material with different levels of difficulty. The maximum level of difficulty observed by an external expert judge is use as a baseline for the highest level of challenge in the graph. In this way, the graph can plot more than one learner. Two ballet students learning the same ballet move could be plotted on the same ZPF graph - but importantly setting a global not relative ground truth allows the system to influence the user's movements in the graph with only one independent variable, 'challenge'. The ground truth is set against tangible measures that can be tested (by the expert or the indicators the expert sets the system to monitor) and in this monitoring is achieved through performance analytics (correct and incorrect responses and response time measures) and affect state tracking. To evaluate the learner accuracy and success, completion time (e.g. learning achievement completion time, time taken to answer a question) is tracked alongside the learning material challenge level in order to determine learner performance in relationship with activity challenge, affect state and learner skill level.

2.1 Frustration

According to Zone of Proximal Development theory, Frustration is where the learner cannot achieve new learning even with assistance. Studies have found that actors who perceive that they lack the skills to take on effectively the challenges presented by the activity in which they are participating experience frustration. Simply put, if a learner feels incompetent in a given

[2] Paul Morsink, *TILE-SIG Feature: The "Digitally Enhanced" Zone of Proximal Development*, http://literacyworldwide.org/blog/literacy-daily/2013/09/20/tile-sig-feature-the-digitally-enhanced-zone-of-proximal-development, accessed on 7/3/2018

situation, he or she will tend not be motivated (Mihaly Csikszentmihalyi, Nakamura, & Abuhamdeh, 2005). This is a negative experience and its gravity pulls the learner further into frustration, in a deteriorating cycle that hampers the learning process. In this state, the learner is exposed to a hopeless feeling; his or her emotional state could be represented by the statement "I do not think anyone can help me".

2.2 Vygotsky Zone of Proximal Development

The ZPD refers to 'the state of arousal where the learner can perform an action or skill provided the aid of a skilled or knowledgeable tutor or in collaboration with more capable peers'(L. L. S. Vygotsky, 1978). This achievement is limited by the ZPD upper limit, however this limit is dependent on the skill of the 'more knowledgeable peer' or scaffolding tools, better tools achieve better results as do more knowledgeable peers induce and encourage higher levels of skill achievement on others due to their access to higher levels of knowledge. This zone limit has been illustrated in **Figure 4**, with the vertical line. While in this zone, the student with assistance can acquire higher skill (Chaiklin, 2003; Radford, Bosanquet, Webster, & Blatchford, 2015; Read, 2006; Verenikina, 2003). In this zone, the level of challenge provides the optimal arousal and engaging experience for the learner to obtain new skills. In this state, the most engaging learning experiences for the learner can happen; it is where optimal and deep learning opportunities manifest themselves. According to (Hermida, 2015) "deep learning is a committed approach to learning. It is a process of constructing and interpreting new knowledge in light of prior cognitive structures and experiences, which can be applied in new, unfamiliar contexts". Deep learning results in better quality learning and profound understanding. While in this zone the student with the assistance of the tutor (instructional scaffolding or assisted learning) acquires higher skill and is encouraged to learn and mentally develop (Chaiklin, 2003; Radford et al., 2015; Read, 2006; Verenikina, 2003).

2.3 Mihaly Csikszentmihalyi Flow

Csikszentmihalyi first described flow in 1997 (Shernoff et al., 2003) as the state where the learners are fully immersed, feeling involved and successful. Flow is a delicate state where the skill level and task challenge levels are balanced. This state represents the learner state where the learner is functioning within their independent capacity, i.e. where the learners find themselves in their comfort zone, both in terms of the learning challenge or learning styles. Flow is also the state where new learning materializes as a new skill in the mind of the learner, which provides the learner an opportunity for reinforcement learning, that carries a successful emotional feeling.

Skill advancement in flow however is limited by the learner's lower limit of ZPD (the maximum a learner can achieve independently) which has been shown with a vertical line in **Figure 4**. Therefore, in order for the learner to achieve new learning outside their independent capacity, the learner must eventually leave flow and be lead to ZPD, to pursue new learning opportunities (i.e. acquire a new skill or to complete competence of partially acquired skill). In either case, in flow or while in ZPD the learner is limited to the upper level of ZPD, which is dependent on the scaffolding tools and scaffolder.

2.4 Boredom

Boredom is the state where the learner is not challenged sufficiently. This state can manifest through the addition of a dry skill base through lecture style teaching, or by providing interactive activities that do not challenge the learner outside what they have already learned. Boredom is a negative feeling and its gravity pulls the learner further into this state, leading to learner disengagement and stifling the learning progress. In this state, the learner's emotional state could be represented by the statement: "let's do interesting things sooner".

In boredom, the low level of challenge relative to skill allows attention to drift. Particularly in contexts of extrinsic motivation, attention shifts to the self and its shortcomings, creating a self-consciousness that impedes engagement of the challenges. Goetz and Hall review development of learners' boredom, and call it an emotion that is frequently experienced by students and can undermine their learning and performance (Robertson, 2015).

3. LEARNING CHALLENGE AND LEARNER SKILL

Learning challenge is a measure for the skill required to complete an activity. Consequently, the unit of learning challenge is the same as skill. When visualized in a graph (see **Figure 5**), any challenge above the diagonal of the graph would represent an increased difficulty and anything underneath the diagonal would represent a lower difficulty. Difficulty does not necessarily match one for one with the user's perceived difficulty of the task, but it is a simple way of representing a tasks challenge relative to others with variable difficulty levels.

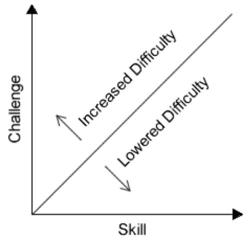


Figure 5:Learning material difficulty vs learner Skill

3.1 A linear approach calculating learning challenge

Learning action difficulty can be changed by adding more steps to the workload (see **Equation 1** which demonstrates this for a mathematics addition problem) or by adding more complexity (see **Equation 2** which demonstrates more difficulty achieved by replacing single digit addition with double digit addition)

$$x + y \textbf{ can become } x + y + z \qquad \textbf{Equation 1}$$

$$x + y \textbf{ can become } xy + wz \qquad \textbf{Equation 2}$$

Adding more steps or complexity also translates to different subjects. For example, in Geography, difficulty increase is made possible by enumerating the question with more answer combinations. This requires more processing before the correct answer is discovered.

Which country is in Africa? *[Difficulty 1]*

a) Lesotho b) Ecuador c) Guyana d) Trinidad and Tobago

Can become:

Which three countries are in Africa? *[Difficulty 3]*

a) Lesotho, Togo, Djibouti

b) Ecuador, Surinam, Eritrea

c) Guyana, Burundi, Mauritania

d) Trinidad and Tobago, Morocco, Togo

These are simple examples on how to achieve a simple linear progression in question and answer difficulty–however, in a more complex procedural skill acquisition task, the number of steps required to complete the task could represent the level of difficulty and this can be translated to almost any type of activity. For example, the number of steps in solving a physics problem, playing a musical score, performing a specific ballet move.

3.2 Calculating Learner skill

To quantify learner skill, a learner's response accuracy percentage (correct to total available attempts) for each level of difficulty towards the greater encompassing skill will be recorded not as a percentage but a value towards the completion of the learning activity. In this way, 100% accuracy of a learning activity difficulty is awarded 100 points, and 90% accuracy is looked on as 90 points, and so on. The sum of these points is divided by the total levels of overall skill difficulty (in this example 10 levels) becomes the average skill of the learner for that skill (see **Figure 6**).

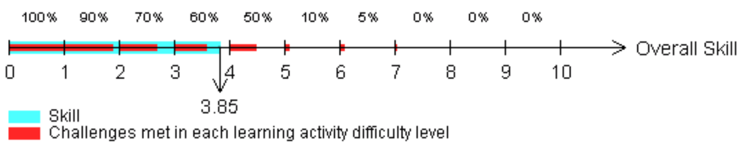


Figure 6: Learner skill calculation in a learning subject with 10 different levels of difficulty

4. LEARNER ENGAGEMENT THROUGH ADAPTIVE LEARNING

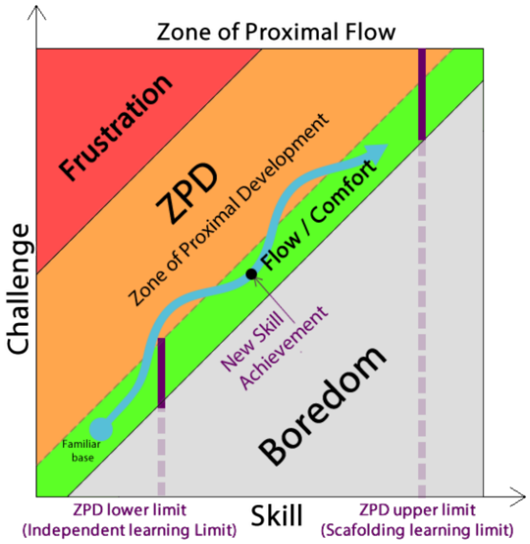


Figure 7: Optimal learning experience loop in ZPF diagram adapted from Basawapatna et al (Basawapatna et al., 2013).

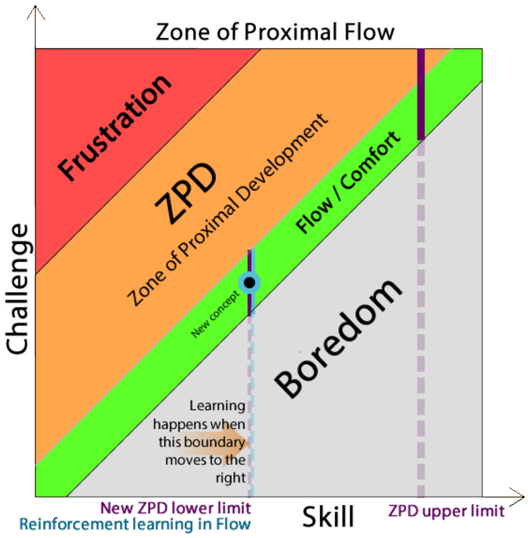


Figure 8: State change is paused while reinforcement learning takes place adapted from Basawapatna et al (Basawapatna et al., 2013). The ZPD lower limit is moved to reflect the new skill achievement.

Supporting learner engagement is important for deep learning and skill achievement. Students who are not engaged can become frustrated or bored which can have a negative effect on achievement and lead to disruption in the classroom, which influences the learning of others. Importantly, many learning processes depend on a simple 'text' for the transfer of knowledge and evaluation. A single mode of learning can have limitations. For example, for a dyslexic student that has a reading-related learning disability, the single source of information transfer therefore becomes a problem (Dolan & Hall, 2001). Universal Design for Learning recognizes this problem (Rose & Meyer, 2001) by embracing the pupil learning diversity by offering multiple means of learning accessibility. Multi-media learning platforms can use audio, audio text, video and tangible objects in a smart learning environment to offer the student a choice of the most accessible formats. This allows for multiple means of recognition, expression and engagement (Dolan & Hall, 2001; Rose & Meyer, 2001). Multi-media approaches in learning resources are best demonstrated in the use of computer-mediated learning where educational games are developed around the learning outcomes and aims. This approach, which is not new, was shown in a systematic review of 129 papers by Connelly et al. (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012) that playing educational games impacts across a range of areas including engagement, cognitive ability and, most commonly, knowledge acquisition and content understanding.

4.1 Optimal learner experience reflected through Zone of Proximal Flow state transitions

How do we apply affect knowledge to a learning platform? Guided by the affect state and the ZPF state diagram of a learner the appropriate level of learning material challenge in order to maintain the learner in an optimal condition where both engagement, as well as skill achievement is maximised can be determined. 'Flow' and 'ZPD arousal' learning states are the active learning states of the learner. New skill acquisition and skill uptake maximisation happens in ZPD, while maintaining the learner in the state of flow and provides the opportunity for reinforcement learning (as visualized in **Figure 7**) which can solidify skills acquired during the learning process and enhance the learning experience itself. Although new skill is not acquired in flow, a slow parallel growth over the long term, with the increase of the level of challenge, introduces an increase in learner skill. This is however limited to the to the lower ZPD independent learning limit, and to increase skill further beyond that, the learner must enter the ZPD.

Any learning material difficulty adaptation processes needs to maintain the learner in the optimal path, as portrayed in **Figure 7**. The optimal path is the one with the shortest forecasted achievement time and one that facilitates the most positive affect states. This path must take the learner through arousal and avoid boredom or frustration. It should start at the lower limit of the ZPD (familiar base) to avoid unnecessary repetition while allowing the learner to remain in the state of flow to enable reinforcement of acquired knowledge (reinforcement learning) visualized in visualized in **Figure 8.** This loop of leaving flow and entering the ZPD (shown as a snakelike pattern in **Figure 7**) should continue until the maximum possible skill achievement is obtained (highlighted by the upper ZPD limit). The caveat being that the 'familiar base' (starting concept) should be challenged for specific learners with disability to re-evaluate previously established learning outcomes.

This process is a delicate one, when the learner is in flow, a continuous effort to push the learner out of their comfort zone and into the ZPD zone by challenging them to greater levels of difficulty will stimulate the learner. However, if the learner is projected too far into arousal, the learner becomes frustrated. By monitoring the learner affect state and learner skill level, a learner can will oscillate between the ZPD and the state of flow with new skill materialization. As a result, a new concept materializes as the learner's skill with just-in-time principles as displayed in **Figure 8** as a circle in the green area. Adaptive learning requires continuous monitoring of learner affective state and Learning Activity progress. The learning path is far more engaging and optimal and the learner is always in a positive affective state. Maintaining the learner in the state of flow provides the opportunity for reinforcement learning.

4.2 Proposed experimental *study*

In this paper, we have proposed how affect state model can be employed to maximize engagement and therefore learning outcome. We plan to raise two research questions; first, what is the relationship between affect state and learning? Second, what impact does an active affect state guided learning platform have on learning and engagement?

The first research question we will investigate by correlating indicators of progress through learning materials with affect-based sensor data (e.g. eye gaze, body pose and movement, vocalisation) extracted from MaTHiSiS educational platform.

The second research question we will use a within subjects ABAB design. Each participant acts as their own control undergoing a series of sessions, some of which are the intervention (A–with the MaTHiSiS system where affect information drives progression through learning materials) and some of which are the control condition (B–the MaTHiSiS system where affect information does not drive the progression through the learning materials). This approach is taken because while reusable learning objects have been widely investigated, MaTHiSiS is the first to introduce an affect state driven response to the learning material presentation. With that in mind, the evaluation is designed to compare the addition of the affective element.

5. CONCLUSION

A literature review has been carried out to form the theoretical background for coupling learning to the emotional state of the learner. The wider literature shows that there is a strong relationship between learning goal outcome achievement and a learner's affect state. Positive affect state have been shown to encourage greater learner outcomes and sustained engagement leads to deep learning and long-term skill retention. The zone of proximal development and the theory of flow's usefulness in explaining the learning process of an educational platform has been described. The Zone of Proximal Flow (ZPF) state diagram is expanded to include the lower ZPD limit and its usefulness as a state diagram for the platform is explored. ZPF affective states in the educational platform have been defined and each affect states importance and impact on individualised learning through the platform has been compared.

A methodology for measuring relative task challenge for a learning activity has been proposed. Methods for linear progression in learning challenge have been developed and examples have been provided. An objective learner skill progression calculation methodology for a wide range of activities is proposed and examples demonstrated. Optimal learning pathways in an educational platform have been described using the ZPD state diagram, and methods for sustaining learner engagement are proposed. To conclude, using the ZPD affect state diagram, methods have been proposed to calculate learner skill and learning activity challenge—in order to locate the learner's affect state and location on the ZPD graph and in turn offer the best intervention for sustained learner engagement. Sustained learner engagement has been shown to be the most reliable identifier for deep learning and retained knowledge.

An experimental design to investigate the impact of this approach system which incorporates the affect state diagram for learning is proposed. This will help us to understand the relationship between learning and affect state and the impact of the system on learning and engagement.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of the EU funded project "Managing Affective-learning THrough Intelligent atoms and Smart InteractionS — MATHISIS': Project Nº 687772".

6. REFERENCES

Basawapatna, A. R., Repenning, A., Koh, K. H., & Nickerson, H. (2013). The zones of proximal flow. *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research - ICER '13*, 67. https://doi.org/10.1145/2493394.2493404

Blackburn, C. (2012). Engaging young children with complex learning dificulties and disabilities, *14*(2), 38–44.

Boulton, H., Brown, D., Standen, P., Belmonte, M., Kwiatkowska, G., Hughes-Roberts, T., & Taheri, M. (2018). Multi-modalities in Classroom Learning Environments. In *INTED2018 Proceedings* (Vol. 1, pp. 1542–1547). IATED. https://doi.org/10.21125/inted.2018.0264

Carpenter, B. (2010). A vision for the 21st century special school. *London: Specialist Schools and Academies Trust*.

Carpenter, Barry, Egerton, J., Cockbill, B., Bloom, T., Fotheringham, J., Rawson, H., & Thistlethwaite, J. (2015). *Engaging Learners with Complex Learning Difficulties and Disabilities: A resource book for teachers and teaching assistants*. Taylor & Francis. Retrieved from https://books.google.com/books?hl=en&lr=&id=sEShBgAAQBAJ&pgis=1

Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. *Vygotsky's Educational Theory and Practice in Cultural Context*, 1–21. https://doi.org/10.1002/bit.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers and Education*, *59*(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, *29*(3), 241–250. https://doi.org/10.1080/1358165042000283101

Csíkszentmihályi, M. (2012). *Flow. 19 Wikipedia Articles*. Retrieved from http://www.reference.com/browse/Mihály Csíkszentmihályi

Csikszentmihalyi, Mihaly, Nakamura, J., & Abuhamdeh, S. (2005). Flow A General Context. *Handbook of Competence and Motivation (Pp. 598-608)*, 598–608. https://doi.org/10.1007/978-94-017-9088

Csikszentmihalyi, Mihalyi. (1996). The Flow of Creativity. *Creativity: Flow and the Psychology of Discovery and Invention*, 107–126.

Dolan, R., & Hall, T. (2001). Universal Design for Learning: Implications for large-scale assessment. *IDA Perspectives*, *27*, 22–25.

Engaging children with complex learning difficulties and disabilities in the Primary Classroom. Barry Carpenter. (2011), (2002).

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007, May). Anxiety and cognitive performance: Attentional control theory. *Emotion*. https://doi.org/10.1037/1528-3542.7.2.336

Goleman, D. (1995). *Emotional intelligence*. Bantam Books. Retrieved from https://books.google.co.uk/books/about/Emotional_Intelligence.html?id=TQQlAQAAIAAJ&redir_esc=y

Hargreaves, D. H. (2006). *A new shape for schooling? World Development*. https://doi.org/10.1016/0305-750X(73)90218-0

Hermida, J. (2015). *Facilitating Deep Learning: Pathways to Success for University and College Teachers*. https://doi.org/9781771880053

Iovannone, R., Dunlap, G., Huber, H., & Kincaid, D. (2003). Effective Educational Practices for Students with Autism Spectrum Disorders. *Focus on Autism & Other Developmental Disabilities*, *18*, 150–165. https://doi.org/Article

MaTHiSiS Project Website. (n.d.). Retrieved March 11, 2018, from http://mathisis-project.eu/

Mello, S. D., Picard, R., & Graesser, A. (n.d.). Towards An Affect-Sensitive AutoTutor.

Pekrun, R., & Linnenbrink-Garcia, L. (2010). *International Handbook of Psychology in Education*. https://doi.org/10.4324/9780203148211

Radford, J., Bosanquet, P., Webster, R., & Blatchford, P. (2015). Scaffolding learning for independence: Clarifying teacher and teaching assistant roles for children with special educational needs. *Learning and Instruction*, *36*, 1–10. https://doi.org/10.1016/j.learninstruc.2014.10.005

Read, C. (2006). Scaffolding children's talk and learning. *Child Psychology and Psychiatry*, 1–18. Retrieved from http://carolread.com/articles/s talk and learning.pdf

Robertson, L. (2015). *International handbook of emotions in education*. *Educational Psychology in Practice* (Vol. 31). https://doi.org/10.1080/02667363.2014.994350

Rose, D. H., & Meyer, A. (2001). Teaching Every Student in the Digital Age: Universal Design for Learning. Retrieved from http://eric.ed.gov/?id=ED466086

Sebba, J., Brown, N., Steward, S., Galton, M., & James, M. (2007). *An Investigation of Personalised Learning Approaches Used by Schools*. Retrieved from http://sro.sussex.ac.uk/22137/

Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., Shernoff, E. S., Schneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, *18*(2), 158–176. https://doi.org/10.1521/scpq.18.2.158.21860

Thompson, N., & McGill, T. J. (2017). Genetics with Jean: the design, development and evaluation of an affective tutoring system. *Educational Technology Research and Development*, *65*(2), 279–299. https://doi.org/10.1007/s11423-016-9470-5

Verenikina, I. (2003). Understanding Scaffolding and the ZPD in Educational Research. *Development*, 8. Retrieved from www.aare.edu.au/03pap/ver03682.pdf

Vos, P., De Cock, P., Munde, V., Petry, K., Van Den Noortgate, W., & Maes, B. (2012). The tell-tale: What do heart rate; skin temperature and skin conductance reveal about emotions of people with severe and profound intellectual disabilities? *Research in Developmental Disabilities*, *33*(4), 1117–1127. https://doi.org/10.1016/j.ridd.2012.02.006

Vygotsky, L. (1978). The interaction between learning and development. Retrieved from http://www.oerafrica.org/sites/default/files/L %26 L reader_section one-reading_4.pdf

Vygotsky, L. L. S. (1978). Mind in society: The development of higher psychological processes. *Mind in Society The Development of Higher Psychological Processes*, *Mind in So*, 159. https://doi.org/10.1007/978-3-540-92784-6